American Journal of Epidemiology

Vol. 190, No. 8
https://doi.org/10.1093/aje/kwab093
Advance Access publication:
March 31, 2021

Practice of Epidemiology

Diagnostic Accuracy Estimates for COVID-19 Real-Time Polymerase Chain Reaction and Lateral Flow Immunoassay Tests With Bayesian Latent-Class Models

Polychronis Kostoulas*, Paolo Eusebi, and Sonja Hartnack

* Correspondence to Dr. Polychronis Kostoulas, Laboratory of Epidemiology and Artificial Intelligence, Faculty of Public and One Health, School of Health Sciences, University of Thessaly, Karditsa, Terma Mavromichali, GR-43100, Greece (e-mail: pkost@uth.gr).

Initially submitted July 7, 2020; accepted for publication March 29, 2021.

Our objective was to estimate the diagnostic accuracy of real-time polymerase chain reaction (RT-PCR) and lateral flow immunoassay (LFIA) tests for coronavirus disease 2019 (COVID-19), depending on the time after symptom onset. Based on the cross-classified results of RT-PCR and LFIA, we used Bayesian latent-class models, which do not require a gold standard for the evaluation of diagnostics. Data were extracted from studies that evaluated LFIA (immunoglobulin G (IgG) and/or immunoglobulin M (IgM)) assays using RT-PCR as the reference method. The sensitivity of RT-PCR was 0.68 (95% probability interval (PrI): 0.63, 0.73). IgG/M sensitivity was 0.32 (95% PrI :0.23; 0.41) for the first week and increased steadily. It was 0.75 (95% PrI: 0.67; 0.83) and 0.93 (95% PrI: 0.88; 0.97) for the second and third weeks after symptom onset, respectively. Both tests had a high to absolute specificity, with higher point median estimates for RT-PCR specificity and narrower probability intervals. The specificity of RT-PCR was 0.99 (95% PrI: 0.98; 1.00). and the specificity of IgG/IgM was 0.97 (95% PrI: 0.92, 1.00), 0.98 (95% PrI: 0.95, 1.00) and 0.98 (95% PrI: 0.94, 1.00) for the first, second, and third weeks after symptom onset. The diagnostic accuracy of LFIA varies with time after symptom onset. Bayesian latent-class models provide a valid and efficient alternative for evaluating the rapidly evolving diagnostics for COVID-19, under various clinical settings and different risk profiles.

Bayesian latent-class models; COVID-19; LFIA; RT-PCR; sensitivity; specificity

Abbreviations: BLCMs, Bayesian latent-class models; *cdn*, conditional covariance between the specificities; *cdp*, conditional covariance between the sensitivities; COVID-19, coronavirus disease 2019; Ig G, immunoglobulin G; Ig M, immunoglobulin M; LFIA, lateral flow immunoassays; RT-PCR, real-time reverse-transcriptase polymerase chain reaction; SARS-CoV-2, serious acute respiratory syndrome coronavirus-2; $Se_{\mathrm{IgG/M}}$, sensitivity of IgG/IgM; $Se_{\mathrm{RT\text{-}PCR}}$, sensitivity of RT-PCR; $Sp_{\mathrm{IgG/M}}$, specificity of IgG/IgM; $Sp_{\mathrm{RT\text{-}PCR}}$, specificity of RT-PCR.

Over the past few months, there has been a need for rapid development of diagnostic tests that will efficiently detect serious acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection. Real-time reverse-transcriptase polymerase chain reaction (RT-PCR) tests, which detect the RNA of SARS-CoV-2, are considered to be the reference standard (1) for a coronavirus disease 2019 (COVID-19) diagnosis. In addition, the development of serological assays detecting SARS-CoV-2–specific immunoglobulin M (IgM) and/or immunoglobulin G (IgG) started immediately and is ongoing (2), with a large portion of them being lateral flow immunoassays (LFIA). These immunoassays are evaluated using RT-PCR as a gold standard (3–5). However, it is known that RT-PCR is less than 100% sensitive (6), while false-positive results can also occur (7). 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 might 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 sensitivity and specificity for each test is estimated from the analysis of the cross-classified results of 2 or more tests in 1 or more populations. Latent models for diagnostic accuracy studies were introduced with the 2-test, 2-population model, which is often referred to as the Hui and Walter paradigm (8). The first thorough discussion on the applicability of these methods in diagnostic accuracy studies was given by Walter and Irwig (9), and their implementation within a Bayesian framework has been evolving for over 20 years (10–12). A meta-analytical alternative for the evaluation of diagnostics from multiple studies in the absence of a reference test has been proposed and can be used, if a sufficiently large number of studies is available (13). Recently, guidelines for the application and sound reporting of BLCMs in diagnostic accuracy studies, the STARD-BLCM statement, have been proposed (14, 15). STARD-BLCM is an adaptation of the STARD statement (16) 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 (17).

To the best of our knowledge, BLCMs (or latent-class models) 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 given that 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 after symptom onset with the use of BLCMs. We followed the STARD-BLCM guidelines (Web Table 1, available at https://doi.org/10.1093/aje/kwab093) (15).

METHODS

Literature search and selection of studies—data sets

A flow chart for the selection process is in Web Figure 1. We conducted the literature search using PubMed, medRxiv, and bioRxiv without any language restrictions. The search strategy and results for each database are presented in Web Table 2.

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 2 researchers independently screened articles. Disagreements in the initial evaluation were resolved by consensus.

Eligible articles were required to meet the following criteria: 1) inclusion of COVID-19 cases (noncases) confirmed (ruled out) by RT-PCR or by a combination of RT-PCR and clinical findings; 2) results concerning IgM and/or IgG antibodies using lateral flow immunoassay; 3) availability of clinical information, in particular with respect to days from onset of symptoms; 4) RT-PCR preceding IgG/M testing by at least 7 days.

In order to construct the 2-by-2 contingency table and obtain estimates for the sensitivity and specificity, we obtained the numbers for antibody (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, 448 nonduplicated records were screened, and 28 full-text resources were scrutinized. Finally, 4 studies (18–21) were identified that fulfilled criteria 1–4 and had cross-classified results that could be extracted (patient characteristics, study design, and diagnostic tests of these studies are summarized in Web Table 3).

Bayesian latent-class model for sensitivity and specificity 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 sensitivity and specificity of the tests is based on the cross-classified results. With 2 tests in 2 populations, the model is fully identifiable because there are 6 degrees of freedom (i.e., 3 from each population) and 6 parameters to be estimated: the sensitivity and specificity of each test and the true prevalence of disease/infection in each population. Here, we extend this model in a 2-test (i.e., RT-PCR and IgG/IgM), 4-population model (i.e., each study is considered a different population) and analyzed weekly with RT-PCR test sensitivity ($Se_{\text{RT-PCR}}$) and specificity ($Sp_{\text{RT-PCR}}$) being constant across all weeks, while IgG/IgM test sensitivity ($Se_{\text{IgG/M}}$) and specificity ($Sp_{\text{IgG/M}}$) were allowed to vary between weeks after symptom onset.

Briefly, we assume that for each of the i populations—in our case the 4 different studies—the cross-classified results of the 2 tests follow an independent multinomial sampling distribution:

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

with the multinomial cell probabilities being expressed as:

$$\begin{aligned}
p_{i_{11}} &= p_i Se_{\text{RT-PCR}} Se_{\text{IgG/M}} + (1-p_i)(1-Sp_{\text{RT-PCR}})\\
&\quad (1-Sp_{\text{IgG/M}})\\
p_{i_{12}} &= p_i Se_{\text{RT-PCR}}(1-Se_{\text{IgG/M}}) + (1-p_i)(1-Sp_{\text{RT-PCR}})\\
&\quad Sp_{\text{IgG/M}}\\
p_{i_{21}} &= p_i(1-Se_{\text{RT-PCR}})Se_{\text{IgG/M}} + (1-p_i)Sp_{\text{RT-PCR}}\\
&\quad (1-Sp_{\text{IgG/M}})\\
p_{i_{22}} &= p_i(1-Se_{\text{RT-PCR}})(1-Se_{\text{IgG/M}}) + (1-p_i)\\
&\quad Sp_{\text{RT-PCR}} Sp_{\text{IgG/M}}.
\end{aligned}$$

Within a fully Bayesian estimation framework, beta distributions $Be\ (a,\ b)$, are used as priors for the parameters

of interest: $Se_{\text{RT-PCR}}$, $Sp_{\text{RT-PCR}}$ $Se_{\text{IgG/M}}$, $Sp_{\text{IgG/M}}$ and the prevalence p_i in each population.

Assessing conditional dependence

Our model assumed that RT-PCR and LFIA are conditionally independent, an assumption that is expected to be valid because the 2 tests are based on a different biological principle (10). 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:

$$
\begin{aligned}
p_{i_{11}} &= p_i(Se_{\text{RT-PCR}}Se_{\text{IgG/M}} + cdp) + (1 - p_i) \\
&\quad ((1 - Sp_{\text{RT-PCR}})(1 - Sp_{\text{IgG/M}}) + cdn) \\
p_{i_{12}} &= p_i(Se_{\text{RT-PCR}}(1 - Se_{\text{IgG/M}}) - cdp) + (1 - p_i) \\
&\quad ((1 - Sp_{\text{RT-PCR}})Sp_{\text{IgG/M}} - cdn) \\
p_{i_{21}} &= p_i((1 - Se_{\text{RT-PCR}})Se_{\text{IgG/M}} - cdp) + (1 - p_i) \\
&\quad (Sp_{\text{RT-PCR}}(1 - Sp_{\text{IgG/M}}) - cdn) \\
p_{i_{22}} &= p_i((1 - Se_{\text{RT-PCR}})(1 - Se_{\text{IgG/M}}) + cdp) + (1 - p_i) \\
&\quad (Sp_{\text{RT-PCR}}Sp_{\text{IgG/M}} + cdn),
\end{aligned}
$$

where cdp is the conditional covariance between the sensitivities and cdn is the conditional covariance between the specificities. Uniform priors were specified for cdp and cdn with their limits being directly affected by the magnitude of the sensitivity and specificity values (22):

$$
\begin{aligned}
cdp &\sim \text{Uniform}((Se_{\text{RT-PCR}} - 1)(1 - Se_{\text{IgG/M}}), \\
&\quad (\min(Se_{\text{RT-PCR}}, Se_{\text{IgG/M}}) - Se_{\text{RT-PCR}}Se_{\text{IgG/M}})) \\
cdn &\sim \text{Uniform}((Sp_{\text{RT-PCR}} - 1\ )(1 - Sp_{\text{IgG/M}}), \\
&\quad (\min(Sp_{\text{RT-PCR}}, Sp_{\text{IgG/M}}) - Sp_{\text{RT-PCR}}Sp_{\text{IgG/M}})).
\end{aligned}
$$

Priors and sensitivity analysis

We have a 2-test, 4-subpopulation model, which is fully identifiable because the numbers of parameters to be estimated are 8 (i.e., the sensitivity and specificity of each test and the prevalence of SARS-CoV-2 infection in each population) for the independence model, and the degrees of freedom available from the data are 12. In all alternative prior combinations, a noninformative, uniform beta prior distribution, Be (1), over the range from 0 to 1, was adopted for the $Se_{\text{RT-PCR}}$, $Se_{\text{IgG/M}}$ and the prevalence of SARS-CoV-2 infection in each population p_i.

For our primary analysis (prior set I) $Sp_{\text{RT-PCR}}$ 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 $Sp_{\text{IgG/M}}$ 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).

Table 1. Cross-Classified Results of the Real-Time Reverse-Transcriptase Polymerase Chain Reaction and the Lateral Flow Immunoassay Tests Detecting Either Immunoglobulin G or Immunoglobulin M Antibodies Against Coronavirus Disease 2019, Using Data From Multiple Studies

Study	PCR(+) IgG/IgM(+)	PCR(+) IgG/IgM(−)	PCR(−) IgG/IgM (+)	PCR (−) IgG/IgM (−)
Week 1				
A (18)	1	7	0	0
B (19)	3	13	2	7
C (20)	3	24	4	5
D (21)	12	15	14	38
Week 2				
A (18)	8	16	15	3
B (19)	6	0	1	1
C (20)	26	2	5	1
D (21)	28	8	14	38
Week 3				
A (18)	17	6	41	4
B (19)	68	0	5	9
C (20)	30	1	5	2
D (21)	17	4	14	38

Abbreviations: IgG, immunoglobulin G; IgM, immunoglobulin M; PCR, polymerase chain reaction.

Table 2. Medians and 95% Probability Intervals for the Sensitivity and Specificity of the Real-Time Reverse-Transcriptase Polymerase Chain Reaction and the Lateral Flow Immunoassay Tests Detecting Immunoglobulin G or Immunoglobulin M Antibodies Against Coronavirus Disease 2019, Using Bayesian Latent-Class Models

Model	Median	PrI	Week 1		Week 2		Week 3	
			Median	PrI	Median	PrI	Median	PrI
A[a]								
$Se_{RT\text{-}PCR}$[b]	0.68	0.63, 0.73						
$Sp_{RT\text{-}PCR}$[b]	0.99	0.98, 1.00						
$Se_{IgG/M}$			0.32	0.23, 0.41	0.75	0.67, 0.83	0.93	0.88, 0.97
$Sp_{IgG/M}$			0.97	0.92, 1.00	0.98	0.95, 1.00	0.98	0.94, 1.00
B[a]								
$Se_{RT\text{-}PCR}$[b]	0.70	0.65, 0.75						
$Sp_{RT\text{-}PCR}$[b]	0.99	0.98, 1.00						
$Se_{IgG/M}$			0.38	0.27, 0.48	0.78	0.70, 0.86	0.93	0.88, 0.97
$Sp_{IgG/M}$			0.98	0.94, 1.00	0.98	0.95, 1.00	0.98	0.94, 1.00
cdp			−0.09	−0.15, −0.03	−0.04	−0.07, 0.00	0.00	−0.02, 0.02
cdn			0.00	0.00, 0.01	0.00	0.00, 0.01	0.00	0.00, 0.01

Abbreviations: *cdn*, conditional covariance between the specificities; *cdp*, conditional covariance between the sensitivities; $Se_{IgG/M}$, sensitivity of immunoglobulin G/immunoglobulin M; $Se_{RT\text{-}PCR}$, sensitivity of RT-PCR; $Sp_{IgG/M}$, specificity of immunoglobulin G/immunoglobulin M; $Sp_{RT\text{-}PCR}$, specificity of real-time reverse-transcriptase polymerase chain reaction.

[a] Model A assumes conditional independence while model B adjusts for the potential dependencies between the sensitivities and specificities of the tests. Data are from multiple studies (18–21).

[b] $Se_{RT\text{-}PCR}$ and $Sp_{RT\text{-}PCR}$ estimates are not week-specific.

Alternative prior combinations were: 1) fixing $Sp_{RT\text{-}PCR}$ equal to 1 and using the same prior for $Sp_{IgG/M}$ (prior set II) and 2) assuming for both $Sp_{RT\text{-}PCR}$ and $Sp_{IgG/M}$ an a priori median of 0.95 and a lower value of 0.90 with 95% certainty. This is a *Be* (76.63, 4.35). The latter prior 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.

Convergence diagnostics and software

We used a combination of checks because convergence diagnostics of the Markov chain Monte Carlo methods are not foolproof. Specifically, the Raftery and Lewis method (23), the Gelman-Rubin diagnostic (24), autocorrelation checks and visual inspection of the trace plots, and summary statistics were used as recommended (25). Parameter estimates were based on analytical summaries of 60,000 iterations of 3 chains after a burn-in adaptation phase of 10,000 iterations. All checks suggested that convergence occurred and autocorrelations dropped off quickly (Web Figure 2). Models were fitted using the freeware program JAGS (26) through the rjags package (26) for R (R Foundation for Statistical Computing, Vienna, Austria) (27). Priors were generated with the PriorGen package (28). The code is available at https://github.com/paoloeusebi/BLCM-Covid19.

RESULTS

A total of 448 studies were initially identified as studies of 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 RT-PCR and LFIA results for each week after symptom onset (Web Figure 1).

Cross classified results of the RT-PCR and the LFIA tests for each week from the onset of COVID-19 symptoms are presented in Table 1. Sensitivity and specificity estimates for each week are in Table 2. $Se_{RT\text{-}PCR}$ was 0.68 (95% probability interval: 0.63; 0.73), while $Se_{IgG/M}$ increased week by week with nonoverlapping probability intervals. Both tests were of high to absolute specificity that did not differ, with point estimates for $Sp_{RT\text{-}PCR}$ being consistently higher. Further, $Sp_{IgG/M}$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 evidence of conditional dependence; covariance parameters, *cdp* and *cdn*, had probability intervals that included zero. Finally, alternative prior specifications—prior set II and III—gave similar results (Web Table 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 sensitivity and specificity estimation of both tests. They provide a valid and efficient alternative to classical test evaluation (8, 15). Importantly, the degrees of freedom provided by the data (i.e., 12) exceeded the

number of parameters that had to be estimated (i.e., 8 for the conditional independence model), satisfying a necessary condition for identifiability. Further, sensitivity analysis revealed that under alternative prior specifications our results were similar (Web Table 4) without differences in the estimates between the 2 model structures and alternative prior sets. The assumption of constant sensitivity/specificity across populations (i.e., the different studies) was assessed by re-analyzing all possible dyads of populations. Sensitivity and specificity for all dyads had overlapping probability intervals, and point estimates were close except for 2 dyads (Web Table 5). These results indicate no evidence against the validity of this assumption. Even in cases of deviations from the assumption of constant accuracy, sensitivity and specificity estimates under BLCMs can be seen as average sensitivity/specificity estimates across all populations, and it is only in the case of distinct differences that such differences must be accounted for (15). Finally, there was no evidence that the assumption of conditional independence was violated, 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 (10).

This is, to the best of our knowledge, the first study using BLCMs for the evaluation of COVID-19 diagnostics. A reason for this could be the absence of suitable data: Despite the vast literature on the evaluation diagnostic tests for SARS-CoV-2 (Web Figure 1), only 4 studies were identified with adequate information. Diagnostic accuracy studies are generally resource-intensive with limited funding and are, most often, based on preexisting samples. However, we scrutinized the literature to identify studies with tests that, although not identical, are based on the same biological principle, target the same biomarker (i.e., detection of IgG/IgM), and follow the same technique (i.e., LFIA). Proper study design with the aim to evaluate diagnostics by the use of BLCMs will provide an efficient and valid framework for sensitivity and specificity estimation under various clinical settings and different target populations and will also allow for a thorough and robust assessment of the validity of the assumptions underlying these models.

BLCMs allowed for the derivation of diagnostic accuracy estimates that were specific to the week after symptom onset. Our $Se_{\text{RT-PCR}}$ estimate was moderate to high—0.68 (95% probability interval: 0.63; 0.73)—and in line with current evidence (29). The $Se_{\text{IgG/M}}$ estimates were low for the first week and showed a steep increase to moderate in the second week that further continued, resulting in high sensitivity values for the third week (Table 2). Importantly, weekly $Se_{\text{IgG/M}}$ estimates had nonoverlapping probability intervals. At the early stages of SARS-CoV-2 infection, IgG/IgM 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 10 days after the onset of symptoms (30). The subsequent increase is in line with published findings (31). Further, an increase in IgG and/or IgM during the first 3 weeks is also recorded (21, 32–34). The median seroconversion time is expected to occur 10 and 12 days after symptom onset for IgG and IgM, with a rapid increase after day 6 that can be followed by a decline in viral load (35). 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 (36, 37). $Se_{\text{IgG/M}}$ is higher than $Se_{\text{RT-PCR}}$ 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 (38). Further, other authors also found a steep increase for antibodies, particularly in the second week, that was accompanied by a slight decrease in the probability of detection with nasopharyngeal swabs/bronchoalveolar/sputum PCR over the first 3 weeks after symptom onset (39).

The $Sp_{\text{RT-PCR}}$ estimate was close to unity but false-positive results can occur (7). There is a scarcity of specificity estimates for RT-PCR methods because they are considered to be the reference standard for the evaluation of diagnostic tests for SARS-CoV-2 infection. False-positive RT-PCR results are assumed to occur only as a result of sample contamination or the high cycle threshold (Ct) values (40). Nevertheless, we do not believe that the estimated false-positive rate could only be due to contamination issues. In studies comparing RT-PCR results with chest computed tomography, a substantial number of samples were found to be chest–computed tomography negative but RT-PCR positive (41, 42). Given that chest-computed tomography has emerged as a valid test for early diagnosis of SARS-CoV-2 infection and its combination with RT-PCR is suggested (40), the perfect specificity of RT-PCR is at best in question. Undoubtedly, $Sp_{\text{RT-PCR}}$ 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 (i.e., if a disease-free population of 10,000 is screened, up to 100 false-positive RT-PCR results should be expected).

Finally, $Sp_{\text{IgG/M}}$ was also close to perfect, but with median estimates consistently lower than those for $Sp_{\text{RT-PCR}}$, although 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 (43). Cross-reactivity between SARS-CoV-2 IgM assays and rheumatoid factor IgM (RF-IgM) has also been observed (44).

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 (10, 15). 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) coexist. This might 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 might 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 sensitivity/specificity estimates without the need for 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 epidemiologic data, BLCMs can provide a framework for sensitivity/specificity estimates that will be specific to different risk profiles and will allow for the interpretation of test outcomes according to the relevant epidemiologic situation in each case.

ACKNOWLEDGMENTS

Author affiliations: Laboratory of Epidemiology and Artificial Intelligence, Faculty of Public and One Health, School of Health Sciences, University of Thessaly, Karditsa, Greece (Polychronis Kostoulas); Health Planning Service, Regional Health Authority of Umbria, Perugia, Italy (Paolo Eusebi); and Section of Epidemiology, University of Zurich, Zurich, Switzerland (Sonja Hartnack).

This work was funded by COST Action CA18208: HARMONY—Novel tools for test evaluation and disease prevalence estimation (https://harmony-net.eu/). The open-access publication of the article was financially supported by HEAL-Link (https://www.heal-link.gr/).

Data availability: All data, models, codes, and priors necessary to reproduce the results of this article are available in the main text or as supplementary files.

Conflict of interest: none declared.

REFERENCES

1. Cheng MP, Papenburg J, Desjardins M, et al. Diagnostic testing for severe acute respiratory syndrome–related coronavirus-2: a narrative review. *Ann Intern Med*. 2020; 172(11):726–734.
2. Petherick A. Developing antibody tests for SARS-CoV-2. *Lancet*. 2020;395(10230):1101–1102.
3. Jia X, Zhang P, Tian Y, et al. Clinical significance of IgM and IgG test for diagnosis of highly suspected COVID-19 infection. *medRxiv*. 2020. (https://doi.org/10.1101/2020.02.28.20029025). Accessed November 2, 2020.
4. Paradiso AV, De Summa S, Loconsole D, et al. Clinical meanings of rapid serological assay in patients tested for SARS-Co2 RT-PCR. *medRxiv*. 2020. (https://doi.org/10.1101/2020.04.03.20052183). Accessed November 2, 2020.
5. Li Z, Yi Y, Luo X, et al. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol.* 2020;92(9):1518–1524.
6. Xiao AT, Tong YX, Zhang S. False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *J Med Virol*. 2020;92(10):1755–1756.
7. Lan L, Xu D, Ye G, et al. Positive RT-PCR test results in patients recovered from COVID-19. *JAMA*. 2020;323(15): 1502–1503.
8. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics*. 1980;36(1):167–171.
9. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *J Clin Epidemiol*. 1988;41(9):923–937.
10. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med*. 2005;68(2–4): 145–163.
11. Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: A review of latent class methods. *Stat Med.* 2014;33(24):4141–4169.
12. Kostoulas P, Nielsen SS, Browne WJ, et al. Sample size estimation to substantiate freedom from disease for clustered binary data with a specific risk profile. *Epidemiol Infect*. 2013;141(6):1318–1327.
13. Dendukuri N, Schiller I, Joseph L, et al. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics*. 2012;68(4):1285–1293.
14. Kostoulas P, Nielsen SS, Branscum AJ, et al. STARD-BLCM: standards for the reporting of diagnostic accuracy studies that use Bayesian latent class models. *Prev Vet Med.* 2017;138:37–47.
15. Kostoulas P, Nielsen SS, Branscum AJ, et al. Reporting guidelines for diagnostic accuracy studies that use Bayesian latent class models (STARD-BLCM). *Stat Med.* 2017;36(23): 3603–3604.
16. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med*. 2003; 138(1):W1–W12.
17. The HARMONY Consortium. HARMONY:novel tools for test evaluation and disease prevalence estimation. https://harmony-net.eu/. Accessed November 8, 2020.
18. Garcia FP, Tanoira RP, Cabrera JPR, 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. (https://doi.org/10.1101/2020.04.11.20062158). Accessed July 4, 2020.
19. Liu Y, Liu Y, Diao B, et al. Diagnostic indexes of a rapid IgG/IgM combined antibody test for SARS-CoV-2. *medRxiv*. 2020. (https://doi.org/10.1101/2020.03.26.20044883). Accessed July 4, 2020.
20. Pan Y, Li X, Yang G, et al. Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. *J Infect*. 2020; 81(1):e28–e32.
21. Whitman JD, Hiatt J, Mowery CT, et al. Test performance evaluation of SARS-CoV-2 serological assays. *medRxiv*. 2020. (https://doi.org/10.1101/2020.04.25.20074856). Accessed July 4, 2020.
22. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics.* 2001;57(1):158–167.
23. Raftery AE, Lewis SM. [Practical Markov Chain Monte Carlo]: comment: one long run with diagnostics: implementation strategies for Markov Chain Monte Carlo. *Stat Sci*. 1992;7(4):493–497.
24. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457–472.
25. Best NG, Cowles MK. CODA: Convergence diagnosis and output analysis software for Gibbs sampling output. Version

0.3. MRC Biostatistics Unit, Cambridge University, Cambridge, UK. 1996.
26. Plummer M. *rjags: Bayesian graphical models using MCMC. R package version 3–13*. https://cran.r-project.org/web/packages/rjags/. Accessed May 7, 2021.
27. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016.
28. Kostoulas P. *PriorGen: Generates Prior Distributions for Proportions. R package version 1*. https://cran.r-project.org/web/packages/PriorGen/index.html. Accessed May 7, 2021.
29. Al-Tawfiq JA, Memish ZA. Diagnosis of SARS-CoV-2 infection based on CT scan vs. RT-PCR: reflecting on experience from MERS-CoV. *J Hosp Infect.* 2020;105(2):154–155.
30. To KK-W, Tsang OT-Y, Leung W-S, 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;20(5):565–574.
31. Adams E, Ainsworth M, Anand R, et al. Antibody testing for COVID-19: A report from the National COVID Scientific Advisory Panel. *Wellcome Open Res*. 2020;5:139.
32. Long Q, Deng H, Chen J, et al. Antibody responses to SARS-CoV-2 in COVID-19 patients: the perspective application of serological tests in clinical practice. *medRxiv*. 2020. (https://doi.org/10.1101/2020.03.18.20038018). Accessed November 8, 2020.
33. Zhang J, Zhang X, Liu J, et al. Serological detection of 2019-nCoV respond to the epidemic: A useful complement to nucleic acid testing. *Int Immunopharmacol*. 2020;88:106861.
34. Guo L, Ren L, Yang S, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis*. 2020;71(15):778–785.
35. Lou B, Li T-D, Zheng S-F, et al. Serology characteristics of SARS-CoV-2 infection since exposure and post symptom onset. *Eur Respir J*. 2020;56(2):2000763.
36. Jin Y, Wang M, Zuo Z, et al. Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019. *Int J Infect Dis.* 2020;94:49–52.
37. Zhao J, Yuan Q, Wang H, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin Infect Dis.* 2020;71(16):2027–2034.
38. Gao Y, Yuan Y, Li TT, 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. (https://doi.org/10.1101/2020.03.26.20042044). Accessed November 8, 2020.
39. Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA*. 2020;323(22):2249–2251.
40. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn*. 2020;20(5):453–454.
41. Ai T, Yang Z, Hou H, et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020;296(2):E32–E40.
42. Bai HX, Hsieh B, Xiong Z, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology*. 2020;296(2):E46–E54.
43. Lassaunière R, Frische A, Harboe ZB, et al. Evaluation of nine commercial SARS-CoV-2 immunoassays. *medRxiv*. 2020. (https://doi.org/10.1101/2020.04.09.20056325). Accessed November 8, 2020.
44. Wang Q, Du Q, Guo B, et al. A method to prevent SARS-CoV-2 IgM false positives in gold immunochromatography and enzyme-linked immunosorbent assays. *J Clin Microbiol*. 2020;58(6):e00375-20.