FORMAL SYSTEMATIC REVIEW

WILEY

Diagnostic performance of COVID-19 serological assays during early infection: A systematic review and meta-analysis of 11 516 samples

John J. Y. Zhang[1] 🟢 | Keng Siang Lee[2] 🟢 | Chee Wui Ong[1] 🟢 | Mae Yee Chan[1] 🟢 | Li Wei Ang[3] 🟢 | Yee Sin Leo[1,3,4,5,6] 🟢 | Mark I-Cheng Chen[3,6] 🟢 | David Chien Boon Lye[1,3,4,5] 🟢 | Barnaby Edward Young[3,4,5] 🟢

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]Bristol Medical School, Faculty of Health Sciences, University of Bristol, Bristol, UK

[3]National Centre for Infectious Diseases, Singapore, Singapore

[4]Tan Tock Seng Hospital, Singapore, Singapore

[5]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[6]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Correspondence
John J. Y. Zhang, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. Postal Address: 1E Kent Ridge Rd, Singapore 119228, Singapore.
Email: zhangjohnjy@gmail.com

Barnaby Edward Young, National Centre for Infectious Diseases, 16 Jalan Tan Tock Seng, Singapore 308442, Singapore.
Email: barnaby_young@ncid.sg

Funding information
National Medical Research Council

Abstract

Objective: The use of coronavirus disease 2019 (COVID-19) serological testing to diagnose acute infection or determine population seroprevalence relies on understanding assay accuracy during early infection. We aimed to evaluate the diagnostic performance of serological testing in COVID-19 by providing summary sensitivity and specificity estimates with time from symptom onset.

Methods: A systematic search of Ovid MEDLINE, Embase, Cochrane Central Register of Controlled Trials (CENTRAL) and PubMed was performed up to May 13, 2020. All English language, original peer-reviewed publications reporting the diagnostic performance of serological testing vis-à-vis virologically confirmed SARS-CoV-2 infection were included.

Results: Our search yielded 599 unique publications. A total of 39 publications reporting 11 516 samples from 8872 human participants met eligibility criteria for inclusion in our study. Pooled percentages of IgM and IgG seroconversion by Day 7, 14, 21, 28 and after Day 28 were 37.5%, 73.3%, 81.3%, 72.3% and 73.3%, and 35.4%, 80.6%, 93.3%, 84.4% and 98.9%, respectively. By Day 21, summary estimate of IgM sensitivity was 0.872 (95% CI: 0.784-0.928) and specificity 0.973 (95% CI: 0.938-0.988), while IgG sensitivity was 0.913 (95% CI: 0.823-0.959) and specificity 0.960 (95% CI: 0.919-0.980). On meta-regression, IgM and IgG test accuracy was significantly higher at Day 14 using enzyme-linked immunosorbent assay (ELISA) compared to other methods.

Conclusions: Serological assays offer imperfect sensitivity for the diagnosis of acute SARS-CoV-2 infection. Estimates of population seroprevalence during or shortly after an outbreak will need to adjust for the delay between infection, symptom onset and seroconversion.

John J. Y. Zhang and Keng Siang Lee are co-first authors and contributed equally to this article.

1 | INTRODUCTION

On March 12, 2020, the World Health Organisation (WHO) declared the outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and associated coronavirus disease 2019 (COVID-19) as a pandemic. Unprecedented control measures have been put in place in an effort to reduce transmission, which at their peak is estimated to have covered a third of the global population.[1-3] Despite these measures, by July 15, 2020, 13.5 million confirmed cases had been reported worldwide with 581 221 deaths—a crude case fatality rate of 4.3%.[4] However, the true number of infections (and deaths) is likely to be substantially higher due to the large proportion of infections which are undiagnosed because of atypical, mild or absent symptoms, or unconfirmed because testing was not available.[5-8]

Serological assays have the potential to play an important role in the surveillance of COVID-19. The results of early seroprevalence studies have indicated that during the first wave of the COVID-19 pandemic, up to 10% of the population of Wuhan, China may have been infected, and up to 33% in cities of other countries have experienced large outbreaks.[9,10] Serology can also be used for the diagnosis of acute infection and can form an important tool in containment strategies by identifying and linking clusters of infection.[11] If a serological immune correlate of protection can be found, these assays may also form part of exit strategies from control measures.[12]

While serological assays have been reported to have high sensitivity and specificity for SARS-CoV-2 infection, this reflects diagnostic performance during convalescence.[13] Understanding antibody kinetics during early SARS-CoV-2 infection is critical for assessing the accuracy of diagnostic serological results and for interpreting the results of seroprevalence studies. To address this issue, we conducted a systematic review and meta-analysis to evaluate the diagnostic performance of serological assays in early COVID-19, when compared to polymerase chain reaction (PCR) as the gold standard.

2 | MATERIALS AND METHODS

This review was conducted in accordance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[14]

2.1 | Search strategy and study selection

A search string comprising synonyms of "COVID-19" and "serological assays" was applied to the following databases: Ovid MEDLINE, Embase, Cochrane Central Register of Controlled Trials (CENTRAL) and PubMed from November 1, 2019 to May 13, 2020 (Table S1). Studies were screened independently by two reviewers (JJYZ and

KSL) with disagreements resolved by consensus or appeal to a third senior reviewer (BEY). Agreement between the reviewers on study inclusion was evaluated using Cohen's κ.[15]

All English language, original peer-reviewed publications reporting the diagnostic performance of serological testing in comparison with virologically confirmed SARS-CoV-2 infection were included. Specific inclusion and exclusion criteria are outlined in Table S2.

2.2 | Risk of bias assessment

The quality of included studies was assessed using QUADAS-2 (Table S3 and Figure S1).[16] In summary, the QUADAS-2 tool consists of four key domains that discuss patient selection, index test, reference standard and flow of patients through the study and timing of the index tests and reference standard. Two researchers (KSL and CWO) assessed the quality of all included studies and discussed discrepancies until consensus was reached.

2.3 | Data extraction and outcome measures

Data were extracted on the following variables: study, sample and patient details, method of diagnosis, type of blood sample and immunoassay, commercial name of test kit, cut-off values adopted, percentage seroconversion for IgM and IgG and the time after symptom onset of serological testing. Outcome measures used were sensitivity and specificity, or true positive (TP), false negative (FN), false positive (FP) and true negative (TN) values. Prevalence, positive predictive value (PPV), negative predictive value (NPV) and F1 scores were calculated from TP, FN, FP and TN values.

2.4 | Statistical analysis

Random effects models were used for meta-analyses of variables and end points.[17] Pooled proportions were computed with the inverse variance method using the variance-stabilizing Freeman-Tukey double arcsine transformation.[18] Confidence intervals (CI) for individual studies were calculated using the Wilson Score confidence interval method with continuity correction. The I^2 statistic was used to present between-study heterogeneity, where $I^2 \leq 30\%$, between 30% and 50%, between 50% and 75%, and ≥75% were considered to indicate low, moderate, substantial and considerable heterogeneity, respectively.[19] P values for the I^2 statistic were derived from the chi-square distribution of Cochran Q test. For pooling of means of numerical variables, we computed missing means and standard deviations (SDs) from medians, ranges and interquartile ranges using the methods proposed by Hozo

et al and Wan et al[20,21] Standard logit confidence intervals for PPV and NPV were calculated as proposed by Mercaldo et al[22]

Bivariate summary receiver operating characteristic (SROC) curves and point estimates of sensitivity and specificity were computed with the approach proposed by Reitsma et al,[23] using a linear mixed model with known variances of the random effects. Bivariate meta-regression with additional likelihood-ratio tests was performed to evaluate for any significant effects of covariates. In addition, univariate meta-analysis was also done using random effect estimation with the DerSimonian-Laird method to produce pooled diagnostic odds ratios.[24]

Publication bias was assessed using funnel plots and Egger's regression test, based on a weighted linear regression of the treatment effect on its standard error.[25,26]

All statistical analyses were performed using R software version 3.4.3 , with the packages *meta* and *mada*.[27,28] *P* values less than .05 were considered statistically significant.

3 | RESULTS

3.1 | Study characteristics

Our search yielded 599 unique publications, of which 72 were reviewed in full text. A total of 39 publications reporting 11 516 samples taken from 8872 human participants met criteria for inclusion in our study (Figure S2).[10,29-66] Reliability of study selection between observers was substantial at both the title and abstract screening stage (Cohen's κ = 0.91) and the full-text review stage (Cohen's κ = 0.94).

Of the 39 publications, 55 distinct studies were conducted. Studies were regarded as distinct if they had different study designs, evaluated different cohorts or analysed different test kits with results reported separately. Of the 55 studies, 37 were retrospective and 18 were prospective. 24 were case-control studies, 19 were case series of COVID-19 patients, and 12 were cohort studies including both COVID-19 and non-COVID-19 participants. The majority of studies were from China (n = 34, 61.8%), followed by Italy (n = 7, 12.7%) and France (n = 6, 10.9%). There were three studies from Hong Kong and one study each from the United States, Germany, Sweden, Japan and Taiwan. Characteristics of included studies are summarized in Table S4.

The type of immunoassay used was specified in 47 studies. Among these, 18 used immunochromatographic assay (ICA), 15 used enzyme-linked immunosorbent assay (ELISA), 13 used chemiluminescent immunoassay (CLIA) and one used dry fluorescence. Forty eight studies tested both IgM and IgG, four studies tested IgM only, two studies tested IgG only and one study tested IgA only.

3.2 | Quality assessment with QUADAS-2

Among the 39 publications, the proportion of studies with low, high and unclear risk of bias and concerns regarding applicability are summarized in Table S3 and Figure S1. More than half of the studies were found to be at high risk of bias under the domains "Patient Selection and Index Test." The main causes for high risk of bias were due to non-cohort study designs and cut-off values not being reported for the respective serological assays.

3.3 | Characteristics of patients and controls

Of the 11 516 samples analysed in total, 5743 were taken from laboratory-diagnosed COVID-19 patients, 5265 were from healthy controls and 508 were from patients infected with human coronaviruses (229E, HKU1, NL63, OC43 and SARS-CoV (using convalescent samples)), influenza A and B, and other respiratory pathogens.

A total of 5743 serum samples were acquired from 3630 patients with COVID-19. 27 studies reported patient gender, of which 50.5% were females (962 out of 1906) and 49.5% were males (944 out of 1906). Pooled mean age of patients across the 22 studies that reported mean and SD of age was 53.8 years (95% CI: 49.7-57.8). The most commonly reported comorbidities were diabetes mellitus, malignancy and hypertension (pooled percentages of 11.6%, 13.4% and 20.0% reported across seven, five and four studies, respectively).

3.4 | IgM Seroconversion

A total of 44 studies reported percentage of IgM seroconversion across 4026 samples, with 30 studies specifying the time of serological testing. 20 studies performed serological testing within 28 days of symptom onset while 10 performed testing within a time interval that extended beyond 28 days up to Day 50. Pooled percentage of IgM seroconversion was 75.3% (95% CI: 69.7-80.6; Figure S3). Study heterogeneity was considerable ($I^2 = 91.5\%$, $P < .0001$). Pooled percentages of IgM seroconversion by Day 7, 14, 21, 28 and after Day 28 were 37.5%, 73.3%, 81.3%, 72.3% and 73.3%, respectively (Table 1 and Figure 1). No evidence of publication bias for IgM seroconversion was identified on Egger's regression test ($P = .69$; Figure S4). Subgroup meta-analysis identified no significant difference in IgM seroconversion rates between the various types of immunoassay ($P = .44$; Figure S5).

3.5 | IgG Seroconversion

A total of 43 studies reported percentage of IgG seroconversion across 4211 samples, with 27 studies specifying the time of serological testing. 17 studies performed serological testing within 28 days of symptom onset while 10 performed testing within a time interval that extended beyond 28 days up to Day 50. Pooled percentage of IgG seroconversion was 85.8% (95% CI: 78.6-92.0). Study heterogeneity was considerable ($I^2 = 96.4\%$, $P < .0001$; Figure S6). Pooled percentages of IgG seroconversion by Day 7, 14, 21, 28 and after Day 28 were 35.4%, 80.6%, 93.3%, 84.4% and 98.9%, respectively

TABLE 1 Percentages of IgM and IgG seroconversion by weeks from symptom onset

Outcome	No. of studies reporting variable	No. of samples analysed	Pooled percentage of samples (95% CI)	I^2 (%)	P value from χ^2 test
IgM seroconversion	44	4026	75.3 (69.7-80.6)	91.5	<.0001
By Day 7	19	491	37.5 (30.9-44.4)	43.0	.0246
By Day 14	26	706	73.3 (64.7-81.2)	76.4	<.0001
By Day 21	17	349	81.3 (69.7-91.1)	79.1	<.0001
By Day 28	8	213	72.3 (48.8-91.7)	84.3	<.0001
After Day 28	7	179	73.3 (51.5-90.8)	87.1	<.0001
IgG seroconversion	43	4211	85.8 (78.6-92.0)	96.4	<.0001
By Day 7	19	486	35.4 (23.9-47.7)	82.5	<.0001
By Day 14	24	686	80.6 (70.0-89.7)	86.7	<.0001
By Day 21	16	337	93.3 (86.1-98.4)	66.7	<.0001
By Day 28	7	212	84.4 (68.1-96.4)	79.3	<.0001
After Day 28	7	181	98.9 (95.6-100.0)	12.8	.3318
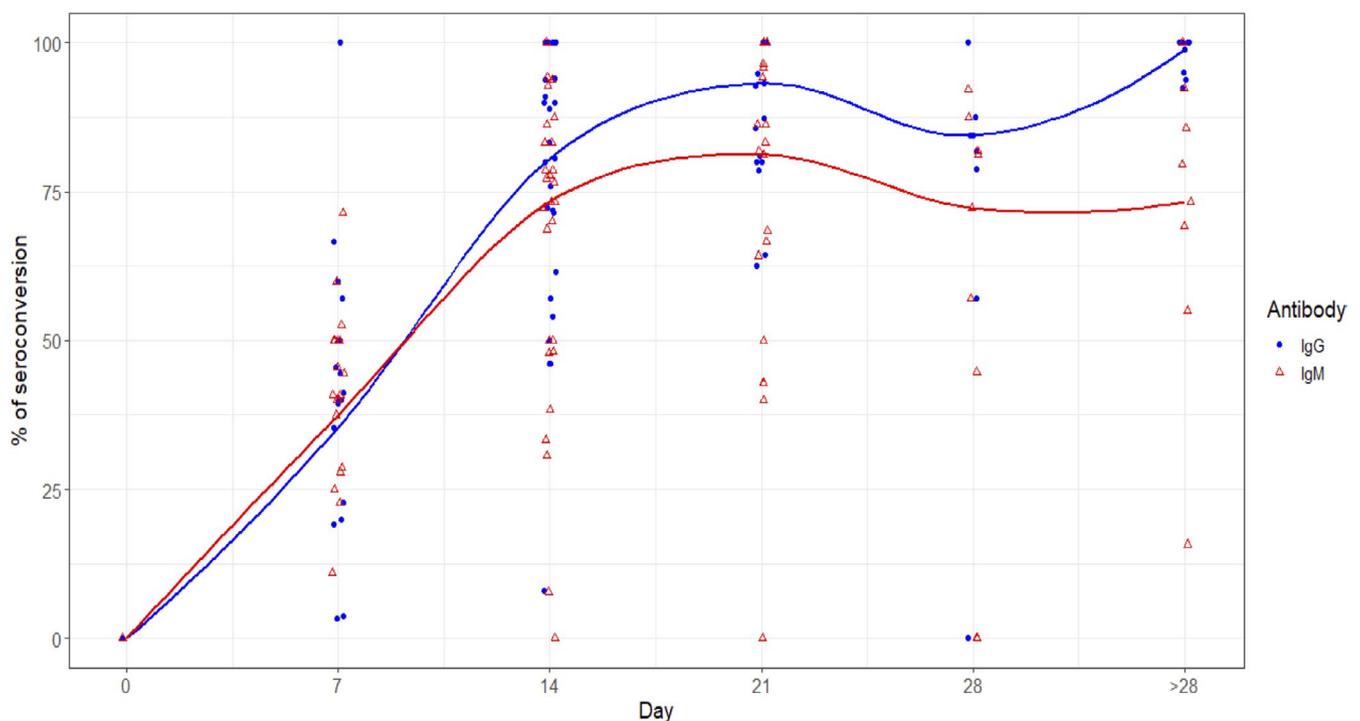
FIGURE 1 Graph of IgM and IgG seroconversion against time from symptom onset

(Table 1 and Figure 1). No evidence of publication bias for IgG seroconversion was identified on Egger's regression test (P = .70; Figure S4). Subgroup meta-analysis identified no significant difference in IgG seroconversion rates between the various types of immunoassay (P = .56; Figure S7).

3.6 | Diagnostic accuracy of IgM testing

TP, FN, TN and FP values for IgM testing were reported in 24 studies. Figure 2 demonstrates a coupled forest plot of sensitivity and specificity values reported across all 24 studies, stratified by time

from symptom onset. On univariate meta-analysis, diagnostic odds ratio for IgM testing was 41.4 (95% CI: 16.1-106.6). Study heterogeneity was negligible ($I^2 = 0.0\%$, P = .54).

Using a bivariate SROC curve, the overall summary estimates of IgM sensitivity and specificity were 0.727 (95% CI: 0.658-0.786) and 0.918 (95% CI: 0.805-0.969), respectively (Figure S8). Sensitivity and specificity values of IgM testing by Day 7, 14 and 21 are shown in Table 2. By Day 21, summary estimate of IgM sensitivity was 0.872 (95% CI: 0.784-0.928) and specificity was 0.973 (95% CI: 0.938-0.988). Prevalence, PPV, NPV and F1 scores of studies reporting IgM testing stratified by time from symptom onset are shown in Table S5.
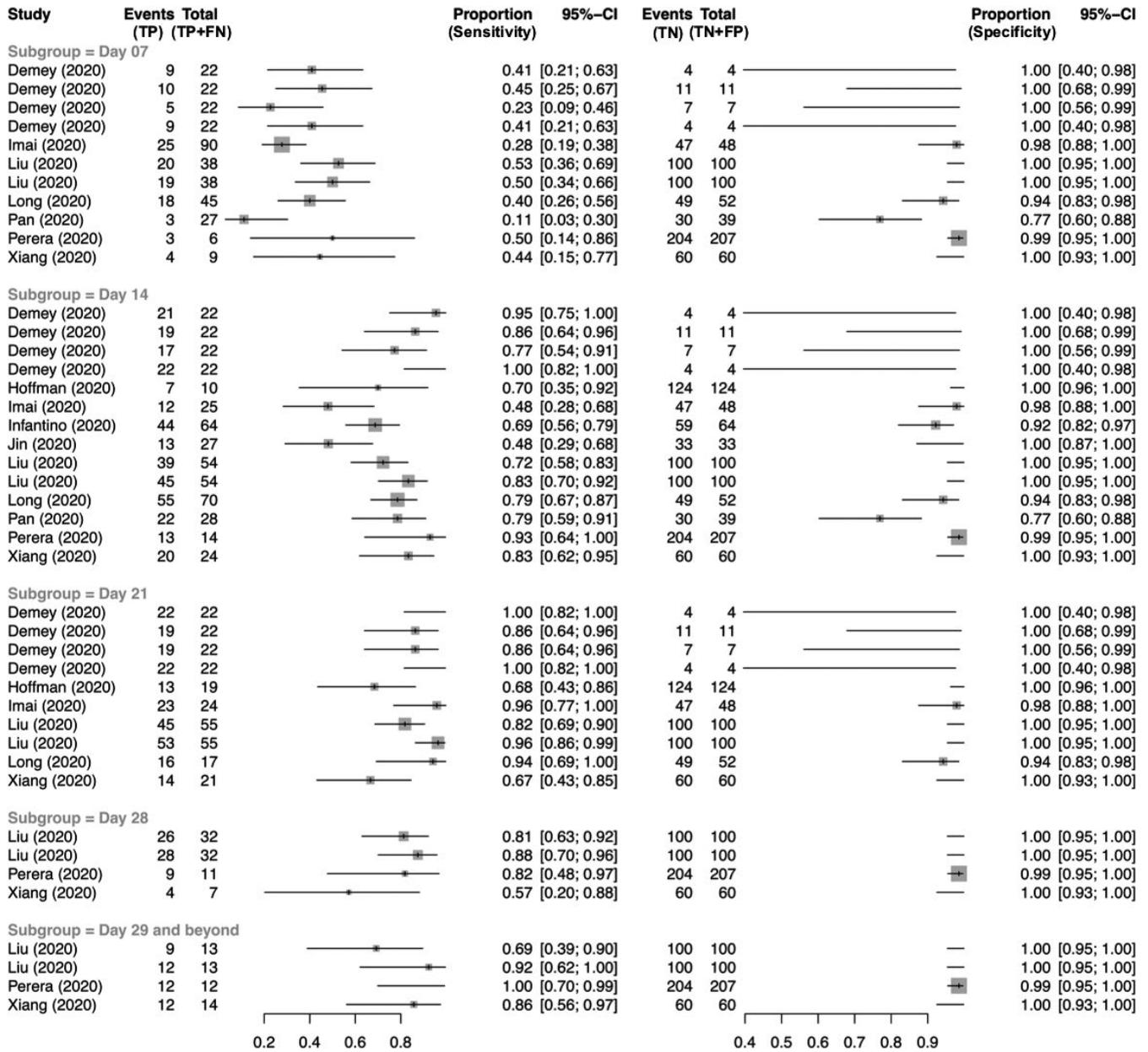
FIGURE 2 Coupled forest plot showing sensitivity and specificity of IgM testing stratified by time from symptom onset

On meta-regression, specificity of IgM testing by Day 14 was significantly higher using ELISA than CLIA (*P* = .006), with similar sensitivity values (*P* = .11). The likelihood-ratio test further suggested a significant difference in test accuracy values with ELISA versus CLIA (χ^2 = 10.55, *P* = .0051). ELISA also produced a significantly higher specificity than ICA in IgM testing by Day 14 (*P* = .014), with similar sensitivity values (*P* = .83). On further comparison using the likelihood-ratio test, however, the difference in diagnostic accuracy between the two tests was only close to statistical significance (*P* = .071).

3.7 | Diagnostic accuracy of IgG testing

TP, FN, TN and FP values for IgG testing were reported in 27 studies. Figure 3 demonstrates a coupled forest plot of sensitivity and

specificity values reported across all 27 studies. On univariate meta-analysis, diagnostic odds ratio for IgG testing was 87.4 (95% CI: 31.0-246.2). Study heterogeneity was insignificant (I^2 = 13.6%, *P* = .26).

Using a bivariate SROC curve, the summary estimates of IgG sensitivity and specificity were 0.788 (95% CI: 0.684-0.865) and 0.948 (95% CI: 0.882-0.978), respectively (Figure S9). Sensitivity and specificity values of IgG testing by Day 7, 14 and 21 are shown in Table 2. By Day 21, summary estimate of IgG sensitivity was 0.913 (95% CI: 0.823-0.959) and specificity was 0.960 (95% CI: 0.919-0.980). Prevalence, PPV, NPV and F1 scores of studies reporting IgG testing stratified by time from symptom onset are shown in Table S6.

Meta-regression demonstrated that IgG testing performed by Day 14 using ELISA had a higher specificity than ICA that was close to statistical significance (*P* = .056), with comparable sensitivity values (*P* = .19).

TABLE 2 Summary of pooled diagnostic performance using IgM, IgG and combined IgM and IgG assays

Serological test and time of test	No. of studies reporting variable	No. of samples analysed	Diagnostic odds ratio	I^2 (%)	P value from χ^2 test	Sensitivity (95% confidence interval)	Specificity (95% confidence interval)	AUC
IgM assay	24	3719	41.4 (16.1-106.6)	0.0	.541	0.727 (0.658-0.786)	0.918 (0.805-0.969)	0.826
By Day 7	11	973	15.3 (4.4-53.3)	0.0	.617	0.381 (0.300-0.469)	0.962 (0.917-0.983)	0.696
By Day 14	14	1311	89.5 (38.7-206.9)	0.0	.558	0.752 (0.674-0.816)	0.962 (0.921-0.982)	0.913
By Day 21	10	789	349.5 (142.2-859.1)	0.0	.825	0.872 (0.784-0.928)	0.973 (0.938-0.988)	0.976
IgG assay	27	6452	87.4 (31.0-246.2)	13.6	.264	0.788 (0.684-0.865)	0.948 (0.882-0.978)	0.922
By Day 7	11	973	10.1 (2.5-40.3)	0.0	.444	0.317 (0.200-0.463)	0.951 (0.875-0.981)	0.730
By Day 14	14	1311	119.9 (33.8-424.6)	0.0	.894	0.793 (0.660-0.883)	0.959 (0.902-0.983)	0.948
By Day 21	10	789	304.2 (129.7-713.5)	0.0	.485	0.913 (0.823-0.959)	0.960 (0.919-0.980)	0.979
IgM or IgG assay	16	3079	51.8 (20.9-128.4)	12.5	.310	0.740 (0.636-0.823)	0.936 (0.887-0.965)	0.924
By Day 7	5	641	15.6 (1.2-207.2)	2.3	.393	0.409 (0.224-0.624)	0.953 (0.763-0.992)	0.719
By Day 14	6	698	100.8 (26.4-384.4)	24.3	.252	0.849 (0.700-0.931)	0.946 (0.801-0.987)	0.944
By Day 21	4	451	1079.1 (269.4-4323.1)	0.0	.541	0.949 (0.900-0.974)	0.975 (0.903-0.994)	0.952

3.8 | Combined value of IgM and IgG testing

TP, FN, FP and TN values for serological testing using IgM or IgG were reported in 16 studies. On univariate meta-analysis, diagnostic odds ratio for testing using IgM or IgG was 51.8 (95% CI: 20.9-128.4). Study heterogeneity was negligible ($I^2 = 12.5\%$, $P = .31$).

Using a bivariate SROC curve, the summary estimates of sensitivity and specificity were 0.740 (95% CI: 0.636-0.823) and 0.936 (95% CI: 0.887-0.965), respectively (Figure S10). Sensitivity and specificity values of testing for IgM or IgG by Day 7, 14 and 21 are shown in Table 2. By Day 21, summary estimate of sensitivity using IgM or IgG testing was 0.949 (95% CI: 0.900-0.974) and specificity was 0.975 (95% CI: 0.903-0.994).

On meta-regression, ELISA was found to have higher specificity values than CLIA ($P = .011$), with similar sensitivity values ($P = .67$). The likelihood-ratio test further suggested a significant difference in test accuracy values with CLIA versus ELISA ($\chi^2 = 6.62$, $P = .036$). ELISA also demonstrated higher specificity values than ICA ($P = .021$), with comparable sensitivity values ($P = .51$). On further comparison using the likelihood-ratio test, however, the difference in diagnostic accuracy between the two tests was only close to statistical significance ($P = .072$).

4 | DISCUSSION

Our meta-analysis indicated that the overall sensitivity and specificity of serological testing for the diagnosis of COVID-19 were 0.727 (95% CI: 0.658-0.786) and 0.918 (95% CI: 0.805-0.969) for IgM, and 0.788 (95% CI: 0.684-0.865) and 0.948 (95% CI: 0.882-0.978) for IgG, respectively. The performance of serological testing was strongly influenced by the duration of infection: sensitivity of IgM increased from 0.381 (95% CI: 0.300-0.469) at Day 7 to 0.872 (95% CI: 0.784-0.928) at Day 21, while sensitivity of IgG increased from 0.317 (95% CI: 0.200-0.463) at Day 7 to 0.913 (95% CI: 0.823-0.959) at Day 21. Combined testing for IgM and IgG offered similar results to IgG alone with a sensitivity of 0.949 (95% CI: 0.900-0.974) at Day 21.

Similarly, rates of IgM and IgG seroconversion were low in the first week of infection at only 37.5% and 35.4%, respectively. For IgM, this rose to 81.3% at Day 21 then declined to 73.3% after Day 28, while for IgG, this rose to 93.3% at Day 21 and 98.9% beyond Day 28. This delayed time to seroconversion indicates that the serological assays included in this review can only play a limited role in the early diagnosis of COVID-19. An early, accurate diagnosis is important for clinical management and necessary for prompt isolation of the infected individual, and contact tracing and quarantine of exposed contacts. An ELISA-based IgG serological test beyond Day 28 from symptom onset will yield accurate results, but such a delayed diagnosis is outside the window for targeting effective control measures. The low sensitivity of IgM reported in our meta-analysis may reflect serological tests which have been calibrated to offer high specificity (>95% from Day 7 to 21) at the cost of sensitivity in

Study	Events (TP)	Total (TP+FN)	Proportion (Sensitivity)	95%–CI	Events (TN)	Total (TN+FP)	Proportion (Specificity)	95%–CI
Subgroup = Day 07								
Demey (2020)	9	22	0.41	[0.21; 0.63]	4	4	1.00	[0.40; 0.98]
Demey (2020)	10	22	0.45	[0.25; 0.67]	10	11	0.91	[0.57; 1.00]
Demey (2020)	5	22	0.23	[0.09; 0.46]	7	7	1.00	[0.56; 0.99]
Demey (2020)	9	22	0.41	[0.21; 0.63]	4	4	1.00	[0.40; 0.98]
Imai (2020)	3	90	0.03	[0.01; 0.10]	48	48	1.00	[0.91; 1.00]
Liu (2020)	15	38	0.39	[0.24; 0.57]	100	100	1.00	[0.95; 1.00]
Liu (2020)	19	38	0.50	[0.34; 0.66]	100	100	1.00	[0.95; 1.00]
Long (2020)	18	45	0.40	[0.26; 0.56]	49	52	0.94	[0.83; 0.98]
Pan (2020)	1	27	0.04	[0.00; 0.21]	24	39	0.62	[0.45; 0.76]
Perera (2020)	3	6	0.50	[0.14; 0.86]	207	207	1.00	[0.98; 1.00]
Xiang (2020)	4	9	0.44	[0.15; 0.77]	57	60	0.95	[0.85; 0.99]
Subgroup = Day 14								
Demey (2020)	21	22	0.95	[0.75; 1.00]	4	4	1.00	[0.40; 0.98]
Demey (2020)	22	22	1.00	[0.82; 1.00]	10	11	0.91	[0.57; 1.00]
Demey (2020)	20	22	0.91	[0.69; 0.98]	7	7	1.00	[0.56; 0.99]
Demey (2020)	22	22	1.00	[0.82; 1.00]	4	4	1.00	[0.40; 0.98]
Hoffman (2020)	9	10	0.90	[0.54; 0.99]	123	124	0.99	[0.95; 1.00]
Imai (2020)	2	25	0.08	[0.01; 0.28]	48	48	1.00	[0.91; 1.00]
Infantino (2020)	46	64	0.72	[0.59; 0.82]	64	64	1.00	[0.93; 1.00]
Jin (2020)	24	27	0.89	[0.70; 0.97]	30	33	0.91	[0.75; 0.98]
Liu (2020)	39	54	0.72	[0.58; 0.83]	100	100	1.00	[0.95; 1.00]
Liu (2020)	41	54	0.76	[0.62; 0.86]	100	100	1.00	[0.95; 1.00]
Long (2020)	63	70	0.90	[0.80; 0.96]	49	52	0.94	[0.83; 0.98]
Pan (2020)	16	28	0.57	[0.37; 0.75]	24	39	0.62	[0.45; 0.76]
Perera (2020)	10	14	0.71	[0.42; 0.90]	207	207	1.00	[0.98; 1.00]
Xiang (2020)	20	24	0.83	[0.62; 0.95]	57	60	0.95	[0.85; 0.99]
Subgroup = Day 21								
Demey (2020)	22	22	1.00	[0.82; 1.00]	4	4	1.00	[0.40; 0.98]
Demey (2020)	22	22	1.00	[0.82; 1.00]	10	11	0.91	[0.57; 1.00]
Demey (2020)	22	22	1.00	[0.82; 1.00]	7	7	1.00	[0.56; 0.99]
Demey (2020)	22	22	1.00	[0.82; 1.00]	4	4	1.00	[0.40; 0.98]
Hoffman (2020)	18	19	0.95	[0.72; 1.00]	123	124	0.99	[0.95; 1.00]
Imai (2020)	15	24	0.62	[0.41; 0.80]	48	48	1.00	[0.91; 1.00]
Liu (2020)	48	55	0.87	[0.75; 0.94]	100	100	1.00	[0.95; 1.00]
Liu (2020)	51	55	0.93	[0.82; 0.98]	100	100	1.00	[0.95; 1.00]
Long (2020)	17	17	1.00	[0.77; 0.99]	49	52	0.94	[0.83; 0.98]
Xiang (2020)	17	21	0.81	[0.57; 0.94]	57	60	0.95	[0.85; 0.99]
Subgroup = Day 28								
Liu (2020)	28	32	0.88	[0.70; 0.96]	100	100	1.00	[0.95; 1.00]
Liu (2020)	27	32	0.84	[0.66; 0.94]	100	100	1.00	[0.95; 1.00]
Perera (2020)	9	11	0.82	[0.48; 0.97]	207	207	1.00	[0.98; 1.00]
Xiang (2020)	4	7	0.57	[0.20; 0.88]	57	60	0.95	[0.85; 0.99]
Subgroup = Day 29 and beyond								
Liu (2020)	13	13	1.00	[0.72; 0.99]	100	100	1.00	[0.95; 1.00]
Liu (2020)	12	13	0.92	[0.62; 1.00]	100	100	1.00	[0.95; 1.00]
Perera (2020)	12	12	1.00	[0.70; 0.99]	207	207	1.00	[0.98; 1.00]
Xiang (2020)	14	14	1.00	[0.73; 0.99]	57	60	0.95	[0.85; 0.99]

Sensitivity axis: 0.2 0.4 0.6 0.8 Specificity axis: 0.4 0.5 0.6 0.7 0.8 0.9

FIGURE 3 Coupled forest plot showing sensitivity and specificity of IgG testing stratified by time from symptom onset

order to overcome the lower affinity and relative non-specific nature of IgM binding.

Improving IgM sensitivity with attendant lower specificity may offer acceptable accuracy in an outbreak scenario where the prevalence of COVID-19 is high. However, this situation is likely to change. An unintended consequence of physical efforts to reduce SARS-CoV-2 transmission is historically low rates of influenza infections in Singapore, Australia and New Zealand.[67,68] However, with the easing of lockdowns and as winter approaches in the Northern Hemisphere, circulation of influenza and other respiratory viruses is likely to increase. In this situation, the prevalence of COVID-19 among symptomatic individuals will decline assuming there is a significant additional number of acute respiratory infections from other causes. Tests with inadequate specificity will hence yield more false positive results and lower positive predictive values. Furthermore,

while cross-reactivity with other human coronaviruses was not evident from the diagnostic performance of assays included in our review, this may become an issue with less specific but more sensitive testing.

An additional question is how these tests will perform in the event of further waves of COVID-19 where a substantial proportion of the population may already have been infected. IgM can persist for months, reducing the specificity of serological diagnosis.[69] With wide variation in seroprevalence between different regions of a country, age groups and contact clusters, the epidemiological context for test interpretation becomes extremely complex. Ongoing seroprevalence studies will also require repeated iterations given the delay in time to seroconversion.

Our findings are in accordance with those from two previous systematic reviews. Lisboa Bastos et al reviewed 40 studies that

investigated the diagnostic performance of serological testing for COVID-19 and concluded that current evidence does not support the continued use of existing point-of-care serological tests.[70] Similarly, a Cochrane Database review by Deeks et al found that the sensitivity of antibody tests is too low in the first week since symptom onset to have a primary role for the diagnosis of COVID-19, and at present, there exists very little evidence beyond 35 days after symptom onset.[71] In addition to confirming the above findings with a more updated search, we present a meta-regression analysis that suggested superiority of ELISA over ICA and CLIA in diagnosing COVID-19 infection.

There were several limitations of our study. First, only studies published in English were included, which may have introduced selection bias. Second, a majority of studies were from Chinese institutions (61.8%); therefore, this may not accurately reflect the performance of serological testing in non-Asian continents. Third, the diagnostic performance of the various serological tests may have been influenced by clinical and host factors, including the severity of disease, age, comorbidities and medications. Clinical symptoms and treatment details were scarcely reported in the included studies; therefore, our analysis was unable to adjust for these factors. Fourth, quantification of the amount of IgM or IgG, and not just a categorical positive or negative result is important to analyse. However, there was too much variability in study designs and testing methods to assess for this. Fifth, there was a paucity of data for serological tests done beyond 28 days of disease onset; hence, further studies would be needed to elucidate the diagnostic performance of antibody testing in individuals beyond 28 days. Finally, despite RT-PCR being deemed to be the gold standard for COVID-19 diagnosis, it may itself have imperfect sensitivity, especially when the viral load is low or when patients are tested late in the course of the infection.[72-74] This leads to possible false-negative PCR results and should be considered as a limitation of our findings.

5 | CONCLUSIONS

Our study demonstrated that the utility of serological assays is limited to IgG more than 14-21 days after symptom onset, when most infected individuals have seroconverted and high sensitivity and specificity values are attained. Diagnostic testing is hence likely to continue to require virological confirmation such as PCR, while seroprevalence studies will significantly underestimate the proportion of a population infected if conducted too early in the epidemic curve. Longer-term studies would be beneficial to improve our understanding of the diagnostic performance of serological testing for COVID-19 in other epidemiological contexts such as endemic infection or future epidemics.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

John J. Y. Zhang: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Keng Siang Lee:** Data curation (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Chee Wui Ong:** Data curation (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Validation (equal); Visualization (equal); Writing-review & editing (equal). **Mae Yee Chan:** Data curation (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Validation (equal); Visualization (equal); Writing-review & editing (equal). **Li Wei Ang:** Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing-review & editing (equal). **Yee Sin Leo:** Funding acquisition (equal); Investigation (equal); Methodology (equal); Resources (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing-review & editing (equal). **Mark I-Cheng Chen:** Funding acquisition (equal); Investigation (equal); Methodology (equal); Resources (equal); Supervision (equal); Validation (equal); Writing-review & editing (equal). **David Chien Boon Lye:** Funding acquisition (equal); Investigation (equal); Methodology (equal); Resources (equal); Supervision (equal); Validation (equal); Writing-review & editing (equal). **Barnaby Edward Young:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Resources (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing-review & editing (equal).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/irv.12841.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

ORCID

John J. Y. Zhang (ID) https://orcid.org/0000-0002-8567-3807
Keng Siang Lee (ID) https://orcid.org/0000-0003-2308-0579
Chee Wui Ong (ID) https://orcid.org/0000-0002-0470-9725
Mae Yee Chan (ID) https://orcid.org/0000-0002-1537-2945

Li Wei Ang [iD] https://orcid.org/0000-0002-0036-255X

Yee Sin Leo [iD] https://orcid.org/0000-0003-4978-5825

Mark I-Cheng Chen [iD] https://orcid.org/0000-0001-9369-5830

David Chien Boon Lye [iD] https://orcid.org/0000-0003-0324-0205

Barnaby Edward Young [iD] https://orcid.org/0000-0003-1010-2230

REFERENCES

1. Pan A, Liu L, Wang C, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA*. 2020;323(19):1915.

2. Htun HL, Lim DW, Kyaw WM, et al. Responding to the COVID-19 outbreak in Singapore: staff protection and staff temperature and sickness surveillance systems. *Clin Infect Dis*. 2020;71(8):1947-1952.

3. Buchholz K. What share of the world population is already on COVID-19 lockdown? *Statista*. 2020.

4. WHO.Coronavirus disease (COVID-2019) situation reports. 2020.

5. Bai Y, Yao L, Wei T, et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA*. 2020;323(14):1406.

6. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020;382:1199-1207.

7. Lee VJ, Chiew CJ, Khong WX. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. *J Travel Med*. 2020;27(3):taa039.

8. Zhang JJY, Lee KS, Ang LW, Leo YS, Young BE. Risk factors of severe disease and efficacy of treatment in patients infected with COVID-19: a systematic review, meta-analysis and meta-regression analysis. *Clin Infect Dis*. 2020;71:2199-2206.

9. Ioannidis J. The infection fatality rate of COVID-19 inferred from seroprevalence data. *medRxiv*. 2020. https://doi.org/10.1101/2020.2005.2013.20101253

10. Wu X, Fu B, Chen L, Feng Y. Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China. *J Med Virol*. 2020;92(10):1795-1796.

11. Pung R, Chiew CJ, Young BE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *Lancet*. 2020;395(10229):1039-1046.

12. Altmann DM, Douek DC, Boyton RJ. What policy makers need to know about COVID-19 protective immunity. *Lancet*. 2020;395(10236):1527-1529.

13. Yong SEF, Anderson DE, Wei WE, et al. Connecting clusters of COVID-19: an epidemiological and serological investigation. *Lancet Infect Dis*. 2020;20(7):809-815.

14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.

15. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measur*. 1960;20(1):37-47.

16. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.

17. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111.

18. Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. *Arch Public Health*. 2014;72(1):39.

19. Higgins JPT, Greeb S. Cochrane handbook for systematic reviews of interventions version 5.1.0. *The Cochrane Collaboration*. 2011.

20. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5(1):13.

21. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14(1):135.

22. Mercaldo ND, Lau KF, Zhou XH. Confidence intervals for predictive values with an emphasis to case-control studies. *Stat Med*. 2007;26(10):2170-2183.

23. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.

24. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.

25. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.

26. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634.

27. Schwarzer G. meta: an R package for meta-analysis. *R News*. 2007;7(3):40-45.

28. Doebler P, Holling H. Meta-analysis of diagnostic accuracy with mada. *R Packag*. 2015;1:15.

29. Bryan A, Pepper G, Wener MH, et al. Performance characteristics of the Abbott architect SARS-CoV-2 IgG assay and seroprevalence in Boise, Idaho. *J Clin Microbiol*. 2020;58(8):e00941–20.

30. Cai XF, Chen J, Hu JL, et al. A peptide-based magnetic chemiluminescence enzyme immunoassay for serological diagnosis of coronavirus disease 2019 (COVID-19). *J Infect Dis*. 2020;222(2):189–193.

31. Cassaniti I, Novazzi F, Giardina F, et al. Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department. *J Med Virol*. 2020;92(10):1724-1727.

32. Chen Z, Zhang Z, Zhai X, et al. Rapid and sensitive detection of anti-SARS-CoV-2 IgG, using lanthanide-doped nanoparticles-based lateral flow immunoassay. *Anal Chem*. 2020;92(10):7226-7231.

33. Demey B, Daher N, François C, et al. Dynamic profile for the detection of anti-SARS-CoV-2 antibodies using four immunochromatographic assays. *J Infect*. 2020;81(2):e6-e10.

34. Döhla M, Boesecke C, Schulte B, et al. Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity. *Public Health*. 2020;182:170-172.

35. Du Z, Zhu F, Guo F, Yang B, Wang T. Detection of antibodies against SARS-CoV-2 in patients with COVID-19. *J Med Virol*. 2020;92(10):1735-1738.

36. Fu W, Chen Q, Wang T. Letter to the Editor: Three cases of re-detectable positive SARS-CoV-2 RNA in recovered COVID-19 patients with antibodies. *J Med Virol*. 2020;92(11):2298-2301.

37. Gao HX, Li YN, Xu ZG, et al. Detection of serum immunoglobulin M and immunoglobulin G antibodies in 2019-novel coronavirus infected cases from different stages. *Chin Med J*. 2020;133(12):1479–1480.

38. 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.

39. Hoffman T, Nissen K, Krambrich J, et al. Evaluation of a COVID-19 IgM and IgG rapid test; an efficient tool for assessment of past exposure to SARS-CoV-2. *Infect Ecol Epidemiol*. 2020;10(1):1754538.

40. Hou H, Wang T, Zhang B, et al. Detection of IgM and IgG antibodies in patients with coronavirus disease 2019. *Clin Transl Immunol*. 2020;9(5):e01136.

41. Imai K, Tabata S, Ikeda M, et al. Clinical evaluation of an immunochromatographic IgM/IgG antibody assay and chest

computed tomography for the diagnosis of COVID-19. *J Clin Virol.* 2020;128:104393.

42. Infantino M, Grossi V, Lari B, et al. Diagnostic accuracy of an automated chemiluminescent immunoassay for anti-SARS-CoV-2 IgM and IgG antibodies: an Italian experience. *J Med Virol.* 2020;92(9):1671-1675.

43. 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.

44. Lee YL, Liao CH, Liu PY, et al. Dynamics of anti-SARS-Cov-2 IgM and IgG antibodies among COVID-19 patients. *J Infect.* 2020;81(2):e55-e58.

45. 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.

46. Liu W, Liu L, Kou G, et al. Evaluation of nucleocapsid and spike protein-based ELISAs for detecting antibodies against SARS-CoV-2. *J Clin Microbiol.* 2020;58(6):e00461–20.

47. Long QX, Liu BZ, Deng HJ, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med.* 2020;26(6):845–848.

48. Padoan A, Sciacovelli L, Basso D, et al. IgA-Ab response to spike glycoprotein of SARS-CoV-2 in patients with COVID-19: a longitudinal study. *Clin Chim Acta.* 2020;507:164-166.

49. Padoan A, Cosma C, Sciacovelli L, Faggian D, Plebani M. Analytical performances of a chemiluminescence immunoassay for SARS-CoV-2 IgM/IgG and antibody kinetics. *Clin Chem Lab Med.* 2020;58(7):1081-1088.

50. 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.

51. Perera RA, Mok CK, Tsang OT, et al. Serological assays for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), March 2020. *Eurosurveillance.* 2020;25(16):2000421.

52. Qu J, Wu C, Li X, et al. Profile of IgG and IgM antibodies against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clin Infect Dis.* 2020;71(16):2255–2258.

53. Shen B, Zheng Y, Zhang X, et al. Clinical evaluation of a rapid colloidal gold immunochromatography assay for SARS-Cov-2 IgM/IgG. *Am J Transl Res.* 2020;12(4):1348-1354.

54. Shi F, Wu T, Zhu X, et al. Association of viral load with serum biomakers among COVID-19 cases. *Virology.* 2020;546:122-126.

55. Solodky ML, Galvez C, Russias B, et al. Lower detection rates of SARS-COV2 antibodies in cancer patients vs healthcare workers after symptomatic COVID-19. *Ann Oncol.* 2020;31(8):1087–1088.

56. Spicuzza L, Montineri A, Manuele R, et al. Reliability and usefulness of a rapid IgM-IgG antibody test for the diagnosis of SARS-CoV-2 infection: a preliminary report. *J Infect.* 2020;81(2):e53-e54.

57. To KK, Tsang OT, Leung WS, 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.

58. Wang B, Wang L, Kong X, et al. Long-term coexistence of SARS-CoV-2 with antibody response in COVID-19 patients. *J Med Virol.* 2020;92(9):1684-1689.

59. Xiang F, Wang X, He X, et al. Antibody detection and dynamic characteristics in patients with COVID-19. *Clin Infect Dis.* 2020;71(8):1930–1934.

60. Xie J, Ding C, Li J, et al. Characteristics of patients with coronavirus disease (COVID-19) confirmed using an IgM-IgG antibody test. *J Med Virol.* 2020;92(10):2004-2010..

61. Xu Y, Xiao M, Liu X, et al. Significance of Serology Testing to Assist Timely Diagnosis of SARS-CoV-2 infections: implication from a Family Cluster. *Emerg Microbes Infect.* 2020;9(1):924-927.

62. Yongchen Z, Shen H, Wang X, et al. Different longitudinal patterns of nucleic acid and serology testing results based on disease severity of COVID-19 patients. *Emerg Microbes Infect.* 2020;9(1):833-836.

63. Zeng H, Xu C, Fan J, et al. Antibodies in infants born to mothers with COVID-19 pneumonia. *JAMA.* 2020;323(18):1848–1849.

64. Zhang G, Nie S, Zhang Z, Zhang Z. Longitudinal change of SARS-Cov2 antibodies in patients with COVID-19. *J Infect Dis.* 2020;222(2):183–188.

65. 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.

66. Zhong L, Chuan J, Gong B, et al. Detection of serum IgM and IgG for COVID-19 diagnosis. *Sci China Life Sci.* 2020;63(5):777-780.

67. Chow A, Hein AA, Kyaw WM. Unintended consequence: influenza plunges with public health response to COVID-19 in Singapore. *J Infect.* 2020;81(2):e68-e69.

68. FluTracking. Tracking COVID-19. 2020.

69. Matheus S, Talla C, Labeau B, et al. Performance of 2 commercial serologic tests for diagnosing Zika virus infection. *Emerg Infect Dis.* 2019;25(6):1153-1160.

70. Lisboa Bastos M, Tavaziva G, Abidi SK, et al. Diagnostic accuracy of serological tests for COVID-19: systematic review and meta-analysis. *BMJ.* 2020;370:m2516.

71. Deeks JJ, Dinnes J, Takwoingi Y, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database of Systematic Reviews.* 2020;(6).

72. Bahreini F, Najafi R, Amini R, Khazaei S, Bashirian S. Reducing false negative PCR test for COVID-19. *Int J MCH AIDS.* 2020;9(3):408-410.

73. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *PLoS One.* 2020;15(12):e0242958.

74. Wikramaratna PS, Paton RS, Ghafari M, Lourenço J. Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR. *Eurosurveillance.* 2020;25(50).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.