AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

VIROLOGY

Evaluation of Three Commercial SARS-CoV-2 Serologic Assays and Their Performance in Two-Test Algorithms

Sarah E. Turbett,[a,b] Melis Anahtar,[a] Anand S. Dighe,[a] Wilfredo Garcia Beltran,[a] Tyler Miller,[a] Hannah Scott,[c] Sienna Marie Durbin,[c] Maheetha Bharadwaj,[c] Jason Thomas,[c] Tasos S. Gogakos,[a] Michael Astudillo,[a] Jochen Lennerz,[a] Eric S. Rosenberg,[a,b] John A. Branda[a]

[a]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
[b]Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
[c]Undergraduate Medical Education Program, Harvard Medical School, Boston, Massachusetts, USA

ABSTRACT Sensitive and specific severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) serologic assays are needed to inform diagnostic, therapeutic, and public health decision-making. We evaluated three commercial serologic assays as stand-alone tests and as components of two-test algorithms. Two nucleocapsid antibody tests (Abbott IgG and Roche total antibody) and one spike protein antibody test (DiaSorin IgG) were included. We assessed sensitivity using 128 serum samples from symptomatic PCR-confirmed coronavirus disease 2019 (COVID-19)-infected patients and specificity using 1,204 samples submitted for routine serology prior to COVID-19's emergence, plus 64 pandemic-era samples from SARS-CoV-2 PCR-negative patients with respiratory symptoms. Assays were evaluated as stand-alone tests and as components of a two-test algorithm in which positive results obtained using one assay were verified using a second assay. The two nucleocapsid antibody tests were more sensitive than the spike protein antibody test overall (70% and 70% versus 57%; $P \leq 0.003$), with pronounced differences observed using samples collected 7 to 14 days after symptom onset. All three assays were comparably sensitive ($\geq$89%; $P \geq 0.13$) using samples collected >14 days after symptom onset. Specificity was higher using the nucleocapsid antibody tests (99.3% and 99.7%) than using the spike protein antibody test (97.8%; $P \leq 0.002$). When any two assays were paired in a two-test algorithm, the specificity was 99.9% ($P < 0.0001$ to 0.25 compared with the individual assays), and the positive predictive value (PPV) improved substantially, with a minimal effect on the negative predictive value (NPV). In conclusion, two nucleocapsid antibody tests outperformed a spike protein antibody test. Pairing two different serologic tests in a two-test algorithm improves the PPV, compared with the individual assays alone, while maintaining the NPV.

KEYWORDS SARS-CoV-2, COVID-19, antibody, serology, coronavirus

Several commercial severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) serologic tests have received FDA emergency-use authorization. Serology appears to complement direct viral RNA detection as a diagnostic tool: RNA detection is most sensitive within the first few days after symptom onset, dropping below 50% after 1 week of symptoms (1–3); in contrast, total antibody is detectable in ~50% of patients after 1 week of symptoms, and sensitivity exceeds 90% after 2 weeks (2, 4). Thus, serology is best suited for (i) supporting the diagnosis of coronavirus disease 2019 (COVID-19) infection in RNA-negative symptomatic patients, (ii) identifying potential convalescent-phase plasma donors, and (iii) establishing seroprevalence in population studies (5, 6). Serology may also prove useful in determining immunity, which could inform return-to-work decisions and other public health measures (5, 6).

Citation Turbett SE, Anahtar M, Dighe AS, Garcia Beltran W, Miller T, Scott H, Durbin SM, Bharadwaj M, Thomas J, Gogakos TS, Astudillo M, Lennerz J, Rosenberg ES, Branda JA. 2021. Evaluation of three commercial SARS-CoV-2 serologic assays and their performance in two-test algorithms. J Clin Microbiol 59:e01892-20. https://doi.org/10.1128/JCM.01892-20.

Editor Angela M. Caliendo, Rhode Island Hospital

Address correspondence to John A. Branda, branda.john@mgh.harvard.edu.

Received 21 July 2020
Returned for modification 13 August 2020
Accepted 3 October 2020

Accepted manuscript posted online 5 October 2020
Published 17 December 2020

Considering the relatively low prevalence of COVID-19 infection in many tested populations and the implications of false-positive results for patient care and public health measures, the Centers for Disease Control and Prevention (CDC) has determined that highly specific (≥99.5%) serologic tests are required to provide adequate positive predictive value (PPV) (7). Although high specificity is reported for many commercial SARS-CoV-2 serologic assays, not all of them consistently meet this specificity threshold (see Table S1 in the supplemental material) (8–10). A potential approach to ensure consistently high specificity involves the application of a two-test algorithm in which reactivity using one assay is confirmed using a different (orthogonal) assay. This strategy is employed in the serodiagnosis of several common infectious diseases, including syphilis, human immunodeficiency virus (HIV) infection, and Lyme disease, and exploration of this approach in SARS-CoV-2 antibody testing was recommended by U.S. national public health officials in a recent testing blueprint for state and local laboratories (11–13).

Here, we evaluated the performance of the Abbott SARS-CoV-2 nucleocapsid IgG test (Abbott Laboratories, Abbott Park, IL), the Liaison SARS-CoV-2 spike protein S1/S2 IgG test (DiaSorin, Centralino, Italy), and the Roche Elecsys anti-SARS-CoV-2 nucleocapsid total antibody test (Roche Diagnostics, Indianapolis, IN) as stand-alone assays and as components of two-test algorithms.

MATERIALS AND METHODS

This study was approved by the Partners Healthcare institutional review board.

COVID-19-infected patients. To evaluate serologic test sensitivity, two sets of serum samples from patients with laboratory-confirmed COVID-19 infection were assembled. The retrospective COVID-19-positive serum set ($n = 101$) was assembled by reviewing medical records of 177 unique patients for whom a serum procalcitonin test had been ordered during the COVID-19 pandemic. Among these 177 patients, 101 met both criteria for inclusion: (i) at least one positive SARS-CoV-2 RNA test result prior to collection and (ii) a date of COVID-19 symptom onset (DOSO) that was obtainable through medical record review. For each included patient, the remnant frozen serum sample (stored at −80°C for approximately 1 month after collection) was used; each sample underwent only one freeze-thaw cycle prior to analysis. The prospective COVID-19-positive serum set ($n = 27$) was assembled prospectively from 31 hospitalized patients with confirmed COVID-19 infection based on RNA detection. Four patients were excluded because the DOSO could not be determined. These samples were stored at 4°C for less than 48 h prior to analysis.

The DOSO was defined as the earliest date that at least one of the following COVID-19 symptoms was reported: fever, chills, loss of smell or taste, myalgias, rhinorrhea, nasal congestion, sore throat, cough, and shortness of breath. All DOSOs were confirmed by a second independent reviewer; discrepancies were adjudicated by a third independent reviewer. If the DOSO could not be determined with confidence, the sample was excluded from the analysis.

Control subjects. To evaluate assay specificity, two sets of control samples were assembled, each comprised of frozen (stored at −80°C for up to 18 months after collection) archived serum samples collected from unique COVID-19-negative subjects. Each sample underwent only one freeze-thaw cycle prior to analysis.

The pre-COVID-19-era control serum set included remnant serum samples submitted during the pre-COVID-19 era (December 2018 to May 2019) from 1,204 unique subjects. Of those samples, 1,107 (92%) had originally been submitted for routine rubella virus antibody testing; these samples had been archived for an extended period according to standard laboratory protocols and thus were available for this study. A convenience sample of 97 additional sera (8% of the pre-COVID-19-era control serum set) was chosen based on a search for available archived samples that had been previously tested and shown to contain one or more of the following potentially cross-reactive factors: HIV antigen/antibody, hepatitis C virus (HCV) IgG antibody, treponemal antibody, hepatitis B virus (HBV) surface antibody, or rheumatoid factor (RHF). A laboratory record review was also performed on the other 1,107 samples in this serum set (those originally submitted for rubella virus antibody testing); if the index sample had been tested for any of these potentially cross-reactive factors as a part of routine clinical care and the result was positive, this was noted. See Table 4 for the number of samples within the pre-COVID-19-era control serum set that were found to contain potentially cross-reactive factors.

The symptomatic control serum set ($n = 64$) included remnant serum samples collected during the COVID-19 pandemic for serum procalcitonin testing from SARS-CoV-2 RNA-negative patients with respiratory symptoms and stored at 4°C for <48 h prior to analysis.

SARS-CoV-2 serology. Serum samples were analyzed using three commercial serologic assays: Abbott SARS-CoV-2 IgG, DiaSorin Liaison SARS-CoV-2 S1/S2 IgG, and Roche Elecsys anti-SARS-CoV-2. The antigen targets and assay methods are described in Table 1. Batches of unblinded samples were tested on all three assays on the same day, over a 1-week period, by senior medical technologists. All testing was performed according to the manufacturers' instructions.

TABLE 1 Commercial SARS-CoV-2 serologic assays used in this study[a]

Manufacturer	Format	Principle	Antigen(s)	Isotype detected	Claimed overall sensitivity (%)[b] (*P* value)	Claimed sensitivity in convalescence (%)[b] (*P* value)	Claimed overall specificity (%)[b] (*P* value)
Abbott	CMIA	Indirect ELISA	Nucleocapsid	IgG only	89.3 (**0.0003**)	100 (0.5)	99.8 (0.4)
DiaSorin	CLIA	Indirect ELISA	Spike (S1, S2)	IgG only	70.9 (**0.02**)	97.6 (0.23)	99.3 (**0.004**)
Roche	ECLIA	Sandwich immunoassay	Nucleocapsid	Total antibody	80.0 (0.16)	100 (0.29)	99.8 (0.33)

[a]Abbreviations: CLIA, chemiluminescence immunoassay; ELISA, enzyme-linked immunosorbent assay; S1, spike protein subunit 1; S2, spike protein subunit 2; IgG, immunoglobulin G; CMIA, chemiluminescent microparticle immunoassay; ECLIA, electrochemiluminescence immunoassay. *P* values reflect the comparison between the claimed sensitivity or specificity and the sensitivity or specificity observed in this study, as reported in Tables 2 and 3. *P* values that indicate significant differences are shown in boldface type.
[b]As reported in the assay package insert.

Data analysis. Sensitivity and specificity were calculated for each assay alone and for pairs of assays used in two-test algorithms. In the two-test approach, the overall result was considered positive if the sample tested positive by both assays and negative if the sample was negative using one or both assays. When calculating the positive predictive value (PPV) or negative predictive value (NPV) at various prevalence rates, assay sensitivity was defined as the observed sensitivity using samples collected >14 days after symptom onset (when sensitivity is reported to be its highest) (8–10). Specificity was defined (for PPV and NPV calculations) as the value obtained when all control samples were combined. Differences between proportions were considered statistically significant if the 2-tailed *P* value was <0.05 as determined using McNemar's test or Fisher's exact test.

RESULTS

Clinical sensitivity of individual serologic assays. When results obtained using all 128 serum samples from confirmed COVID-19 cases were considered together, the Abbott IgG and Roche total antibody assays were comparably sensitive (70% for each; $P = 1.0$), whereas the DiaSorin IgG assay was significantly less sensitive than the others (57%; $P \leq 0.003$) (Table 2). Overall sensitivity values reported here are significantly lower than the manufacturers' claimed sensitivities for the Abbott and DiaSorin assays, whereas our findings confirm the sensitivity claims of the Roche assay (Table 1).

When the samples were stratified into three subcategories based on DOSO, the sensitivity of each assay was directly related to symptom duration (Table 2). Although the DiaSorin assay was numerically less sensitive than the other two assays in all three subcategories, the differences were significant only for samples collected 7 to 14 days after symptom onset. In this subcategory, the Abbott and Roche assays were 62% and 64% sensitive, respectively ($P = 1.0$), while the DiaSorin assay was 42% sensitive ($P \leq 0.02$). All three assays were highly sensitive using samples collected >14 days after symptom onset (≥89%; $P \geq 0.13$) (Table 2), and the results were in agreement with those claimed by the manufacturers for samples collected during a similar time frame (Table 1).

Specificity of individual serologic assays. When results from all COVID-19-negative subjects ($n = 1{,}268$) were considered together, the Abbott and Roche assays had similar overall specificities (99.3% and 99.7%; $P = 0.23$) (Table 3). The DiaSorin assay's overall specificity was significantly lower than those of the other two assays (97.8%; $P \leq 0.002$) and the manufacturer's claimed specificity (Table 1).

TABLE 2 Clinical sensitivity of three commercial SARS CoV-2 serologic assays in symptomatic patients with confirmed COVID-19[d]

Sample collection time frame (no. of samples)	Sensitivity (%) (95% CI)			*P* values
	Abbott	Roche	DiaSorin	
<7 days after symptom onset (20)	20 (8–42)	20 (8–42)	10 (3–30)	0.68,[a] 0.48,[b] 0.68[c]
7–14 days after symptom onset (53)	62 (49–74)	64 (51–78)	42 (29–55)	**0.02**,[a] 1.0,[b] **0.01**[c]
>14 days after symptom onset (55)	96 (88–99)	93 (83–97)	89 (78–95)	0.13,[a] 0.48,[b] 0.62[c]
All samples (128)	70 (62–78)	70 (61–77)	57 (48–65)	**0.002**,[a] 1.0,[b] **0.003**[c]

[a]*P* values reflect the comparison between the Abbott and DiaSorin assays.
[b]*P* values reflect the comparison between the Abbott and Roche assays.
[c]*P* values reflect the comparison between the DiaSorin and Roche assays.
[d]CI, confidence interval. *P* values that indicate significant differences are shown in boldface type.

TABLE 3 Specificity of three commercial SARS CoV-2 serologic assays in control subjects[e]

Control cohort (no. of samples)	Specificity (%) (95% CI)			*P* values
	Abbott	Roche	DiaSorin	
Pre-COVID-19-era control serum set (1,204)	99.3 (99–100)	99.8 (99–100)	98.0 (97–99)	**0.007**,[a] 0.11,[b] **<0.001**[c]
Symptomatic control serum set (64)	98.4[d] (91–100)	96.9[d] (89–100)	93.8[d] (85–98)	0.25,[a] 1.00,[b] 0.63[c]
All samples (1,268)	99.3 (99–100)	99.7 (99–100)	97.8 (97–98)	**0.002**,[a] 0.23,[b] **<0.001**[c]

[a]*P* values reflect the comparison between the Abbott and DiaSorin assays.
[b]*P* values reflect the comparison between the Abbott and Roche assays.
[c]*P* values reflect the comparison between the DiaSorin and Roche assays.
[d]One sample from the symptomatic control serum set was positive using all 3 assays. The sample was from a healthy male in his 30s who presented in April 2020 with fever, anosmia, shortness of breath, myalgias, and acute cardiomyopathy. Multiple respiratory SARS-CoV-2 RNA tests were negative. For the purposes of this study, the subject remained categorized as COVID-19 negative, but clinical suspicion remained high.
[e]CI, confidence interval. *P* values that indicate significant differences are shown in boldface type.

When results from the pre-COVID-19-era control serum set ($n = 1,204$) were considered separately from those of the symptomatic control serum set ($n = 64$), each assay was numerically more specific in the former group than in the latter; none of the differences was statistically significant (Table 3). Similarly, when only results obtained from the symptomatic control serum set were compared across the three assays, there were numerical differences in specificity (Abbott, 98.4%; Roche, 96.9%; DiaSorin, 93.8%), but none was statistically significant.

When results from the pre-COVID-19-era control serum set were analyzed, the DiaSorin assay was 98.0% specific, which was significantly lower than that of the Abbott or Roche assay (99.3% and 99.8%; $P \leq 0.007$) (Table 3). The difference in specificity between the Roche and Abbott tests was nonsignificant ($P = 0.11$). False-positive rates among samples known to contain potentially cross-reactive factors are presented in Table 4. Notably, 17.6% of samples yielding false-positive SARS-CoV-2 antibody test results (6/34) contained more than one potential cross-reactive antibody (Table 4). The highest false-positive rates were produced in samples containing HIV antibody/antigen (Abbott, 3/25 [12.0%]; DiaSorin, 2/25 [8%]) or HCV IgG antibody (DiaSorin, 4/44 [9.1%]). The Roche assay produced few false-positive results in samples containing antibodies directed against non-SARS-CoV-2 infectious agents but produced one false-positive result in a sample containing autoantibodies to both rheumatoid factor and antinuclear antibody (ANA).

Clinical sensitivity of two-test algorithms using pairs of serologic assays. Results obtained by testing the COVID-19-positive sample cohort using each individual assay were reanalyzed to determine the effect of applying a two-test algorithm in which the overall result was recorded as positive if a positive result was produced using both assays in a pair (Table 5). If one or both assays in a pair produced a negative result, the two-test result was recorded as negative. Among the three possible two-test

TABLE 4 False-positive rates among samples containing potentially cross-reactive factors[a]

Category	No. of samples tested	No. of positive samples (%)		
		Abbott	Roche	DiaSorin
HIV Ag/Ab positive	25	3[b] (12.0)	0 (0)	2[c,d] (8.0)
HCV IgG positive	44	0 (0)	0 (0)	4[c,e] (9.1)
HBV sAb positive	141	4[b] (2.8)	0 (0)	5[c,d,e] (3.6)
Syphilis Ab positive	27	0 (0)	0 (0)	1[c] (3.7)
RHF positive	2	0 (0)	1[f] (50.0)	0 (0)
ANA positive	6	0 (0)	1[f] (16.7)	0 (0)

[a]False-positive rates were calculated for each potential interference factor. Some samples contained multiple potential interference factors.
[b]Two samples contained both human immunodeficiency virus antigen/antibody (HIV Ag/Ab) and hepatitis B virus surface antibody (HBV sAb).
[c]One sample contained HBV sAb, hepatitis C virus immunoglobulin G (HCV IgG), HIV Ag/Ab, and syphilis antibody (treponemal antibody).
[d]One sample contained both HIV Ag/Ab and HBV sAb.
[e]One sample contained HCV IgG and HBV sAb.
[f]One sample contained both antinuclear antibody (ANA) and rheumatoid factor (RHF).

TABLE 5 Sensitivity and specificity of two-test algorithms using pairs of commercial SARS-CoV-2 serologic assays[h]

Cohort (no. of patients)	Abbott + DiaSorin		Abbott + Roche		DiaSorin + Roche		*P* values
	Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)	Sens (95% CI)	Spec (95% CI)	
COVID-19 patients							
<7 days after symptom onset (20)	0 (0–14)		15 (5–42)		0 (0–14)		0.13,[a] 0.50,[b] 1.00,[c] 1.00,[d] 0.50,[e] 0.13[f]
7–14 days after symptom onset (53)	36 (24–49)		57 (43–69)		38 (26–51)		**0.001**,[a] 0.25,[b] 0.25,[c] 0.13,[d] 0.50,[e] **0.001**[f]
Samples collected >14 days after symptom onset (55)	89 (78–95)		93 (82–98)		87 (78–95)		0.13,[a] 1.0,[b] 0.48,[c] 1.0,[d] 1.0,[e] 0.25[f]
Total (128)	53 (45–62)		66 (57–73)		53 (45–62)		**<0.0001**,[a] 0.06,[b] **0.03**,[c] 0.06,[d] 0.06,[e] **0.008**[f]
Control cohort							
Pre-COVID-19-era control serum set (1,204)		100 (100)		100 (100)		100 (100)	**0.01**,[a] **<0.0001**,[b] **0.01**,[c] 0.50,[d] **<0.0001**,[e] 0.50[f]
Symptomatic control serum set (64)		98 (91–100)		98 (91–100)		98 (91–100)	1.0,[a] 0.25,[b] 1.0,[c] 1.0,[d] 0.25,[e] 1.0[f]
Total (1,268)		99.9[g] (100)		99.9[g] (100)		99.9[g] (100)	**0.01**,[a] **<0.0001**,[b] **0.01**,[c] 0.25,[d] **<0.0001**,[e] 0.25[f]

[a]*P* values reflect the comparison between the Abbott assay and a 2-test algorithm pairing the Abbott and DiaSorin assays.
[b]*P* values reflect the comparison between the DiaSorin assay and a 2-test algorithm pairing the Abbott and DiaSorin assays.
[c]*P* values reflect the comparison between the Abbott assay and a 2-test algorithm pairing the Abbott and Roche assays.
[d]*P* values reflect the comparison between the Roche assay and a 2-test algorithm pairing the Abbott and Roche assays.
[e]*P* values reflect the comparison between the DiaSorin assay and a 2-test algorithm pairing the DiaSorin and Roche assays.
[f]*P* values reflect the comparison between the Roche assay and a 2-test algorithm pairing the DiaSorin and Roche assays.
[g]One sample from the symptomatic control serum set was positive using all 3 assays. The sample was from a healthy male his 30s who presented in April 2020 with fever, anosmia, shortness of breath, myalgias, and acute cardiomyopathy. Multiple respiratory SARS-CoV-2 RNA tests were negative. For the purposes of this study, the subject remained categorized as COVID-19 negative, but clinical suspicion remained high.
[h]Abbreviations: Sens, sensitivity; Spec, specificity. *P* values that indicate significant differences are shown in boldface type.

combinations, the highest overall sensitivity was achieved when the Abbott and Roche assays were paired (66%) (Table 5). However, pairing these assays resulted in lower sensitivity than with the Abbott test alone (70% [90/128]; $P = 0.03$). This pairing was also numerically less sensitive than the Roche assay alone (70% [89/128]), although the difference did not quite reach statistical significance ($P = 0.06$).

When samples were stratified into three subcategories based on DOSO, sensitivity was either numerically the same or lower for all pairs of assays than with the individual component assays alone, for all sample collection time frames (Table 5). However, significant differences in sensitivity were found only using samples collected 7 to 14 days after symptom onset. In this time frame, pairing the DiaSorin test with either the Abbott or Roche test resulted in a lower overall sensitivity (36% or 38%, respectively) than with the Abbott or Roche assay alone (62% or 64%, respectively; $P \leq 0.01$). Sensitivity was not significantly reduced, compared with individual assays alone, by pairing any two assays in a two-test algorithm using samples collected >14 days after symptom onset (Table 5).

Pairing either the Abbott or Roche assay with the DiaSorin assay in a two-test algorithm resulted in lower overall sensitivity (53% for either pair) than with pairing the Abbott and Roche assays together (66%; $P = 0.0001$). Pairs involving the DiaSorin assay were also less sensitive overall (53%) than either the Abbott or the Roche assay alone (70%; $P \leq 0.008$ for both comparisons). The sensitivities of these pairs (DiaSorin plus Abbott or DiaSorin plus Roche) were not significantly lower than that of the DiaSorin assay alone (53% versus 57%; $P = 0.06$).

Specificity of two-test algorithms using pairs of serologic assays. For each of the three possible two-test pairs, the specificity was 99.9% when results from all control

TABLE 6 Positive and negative predictive values of three commercial SARS-CoV-2 serologic assays used alone or in two-test algorithms assuming a range of hypothetical prevalence rates[a]

Parameter	Value for test					
	Abbott	Roche	DiaSorin	Abbott + DiaSorin	Abbott + Roche	DiaSorin + Roche
PPV (%) if SARS-CoV-2 prevalence is 25%	97.9	99.0	93.1	99.7	99.7	99.7
NPV (%) if SARS-CoV-2 prevalence is 25%	98.7	97.7	96.4	96.5	97.7	95.8
PPV (%) if SARS-CoV-2 prevalence is 10%	93.8	97.2	81.8	99.0	99.0	99.0
NPV (%) if SARS-CoV-2 prevalence is 10%	99.6	99.2	98.8	98.8	99.2	98.6
PPV (%) if SARS-CoV-2 prevalence is 3%	81.0	90.6	55.6	96.5	96.6	96.4
NPV (%) if SARS-CoV-2 prevalence is 3%	99.9	99.9	99.8	99.7	99.9	99.8
PPV (%) if SARS-CoV-2 prevalence is 1.5%	67.6	82.5	38.1	93.3	93.4	93
NPV (%) if SARS-CoV-2 prevalence is 1.5%	99.9	99.9	99.8	99.8	99.9	99.8
PPV (%) if SARS-CoV-2 prevalence is 0.1%	12.1	23.7	3.9	47.1	48.2	46.6
NPV (%) if SARS-CoV-2 prevalence is 0.1%	100	100	100	100	100	100

[a]Abbreviations: PPV, positive predictive value; NPV, negative predictive value.

samples were considered together (Table 5). The specificity of the two-test approach (99.9%) was significantly higher than that of the Abbott (99.3%; $P = 0.01$) or DiaSorin (97.8%; $P < 0.0001$) assay alone but not the Roche assay (99.9% versus 99.7%; $P = 0.25$). The same trend was seen within the pre-COVID-19-era control serum set subcategory ($n = 1,204$). Each assay pair produced 100% specificity, which was significantly higher than the specificity of the Abbott (99.3%; $P = 0.01$) or DiaSorin (98.0%; $P < 0.0001$) assay alone but not the Roche assay alone (99.8%; $P = 0.50$).

Each pair of assays was 98% specific in the symptomatic control serum set subcategory ($n = 64$) (Table 5). This specificity was equal to that of the Abbott assay alone (98%; $P = 1.0$) and comparable to that of the Roche (97%; $P = 1.0$) or DiaSorin (94%; $P = 0.25$) assay alone.

Positive predictive value of serologic assays used alone or in two-test algorithms. For illustrative purposes, the PPV was calculated using five hypothetical COVID-19 disease prevalence rates (Table 6), some of which reflect reported prevalences in different regions of the United States at the time of writing (Queens County, NY, 3%; Massachusetts, 1.5%; Alaska, 0.1%) (14) and others of which reflect higher hypothetical prevalence rates. The PPV for pairs of assays used in two-test algorithms was substantially higher than those for the individual assays alone (Table 6). For example, when the disease prevalence was assumed to be 3%, the PPV was ≥96.4% for any pair of assays, whereas the PPV for the individual assays alone ranged from 55.6 to 90.6%.

Negative predictive value of serologic assays used alone or in two-test algorithms. Even at the highest assumed disease prevalence (25%), each individual assay produced NPVs of ≥96.4% (range, 96.4 to 98.7%) (Table 6), and the NPV increased as the assumed disease prevalence decreased. When pairs of assays were used in two-test algorithms, NPVs did not change substantially compared with those of each component assay alone; all two-test combinations produced NPVs of ≥95.8% when a 25% disease prevalence was assumed (range, 95.8 to 97.7%) (Table 6).

DISCUSSION

Accurate serologic tests are needed to inform diagnostic, therapeutic, and public health decisions. We evaluated three commercial SARS-CoV-2 serologic assays, assessing their performance as stand-alone tests and as components of two-test algorithms. The nucleocapsid antibody tests included in our study (Abbott IgG and Roche total antibody assays) were significantly more sensitive and specific than the spike protein antibody test (DiaSorin IgG assay) when used as stand-alone tests. Pairing any individual assay with another in a two-test algorithm in which an initial positive result is verified with a second orthogonal test resulted in a substantially increased PPV compared with those of the component assays alone, while the NPV was minimally affected.

Reflecting the known kinetics of the SARS-CoV-2 antibody response (6), all three assays were poorly sensitive (≤20%) using samples collected <1 week after symptom onset; few patients seroconvert during this time frame. The two nucleocapsid antibody

tests were significantly more sensitive than the spike protein antibody test among samples collected from patients 7 to 14 days after symptom onset, a time frame during which seroconversion frequently occurs (1, 2). Previous studies of patients infected with SARS-CoV-1 indicate that a nucleocapsid antibody response develops faster than a spike protein antibody response; if a similar phenomenon occurs with SARS-CoV-2 infection, this could account for the observed sensitivity difference during this time frame (15, 16). Still, even the assay with the highest observed sensitivity in the 7- to 14-day period after symptom onset (64%; Roche assay) is inadequate for diagnostic purposes, emphasizing the limited utility of serology as a diagnostic tool for acute COVID-19 infection (6). Interestingly, the Abbott IgG nucleocapsid antibody test's sensitivity was not inferior to that of the Roche total nucleocapsid antibody test among the subcategories of patients with $<$1 week or 7 to 14 days of symptoms, suggesting that detection of IgM antibody in the Roche assay did not provide significant performance improvement during early infection.

No significant differences in sensitivity among the individual assays were seen in samples tested $>$14 days after symptom onset, with each assay demonstrating its highest sensitivity using samples collected during that time frame (range, 89 to 96%). Each assay's NPV was $\geq$96.4% when a disease prevalence of as high as 25% was assumed. Thus, any of the individual serologic tests evaluated here would be useful in ruling out prior exposure in asymptomatic individuals. The observed overall sensitivities of the Abbott and DiaSorin assays were significantly lower than those claimed by the manufacturers, potentially due to differences in populations tested between our evaluation and theirs. Most of the COVID-19-infected subjects in our study were hospitalized patients, perhaps selecting for a sicker cohort than with clinical evaluations referenced by the manufacturers.

There were important differences in specificity between the individual serologic assays. The DiaSorin assay's overall specificity (97.8%) was significantly lower than those of the Abbott and Roche assays (99.3% and 99.7% specificities, respectively; $P < 0.002$). While the difference in overall specificities between the Abbott and Roche assays was nonsignificant, only the Roche assay met the proposed threshold of $\geq$99.5% specificity suggested by the CDC as being adequate for use as a stand-alone test (7). The Roche assay's high specificity may owe not only to the target antigen but also to the method itself, as sandwich immunoassays are generally less prone to false-positive results than indirect enzyme-linked immunosorbent assays (ELISAs) (17).

Among the 1,204 control subjects whose samples were collected for routine serologic testing prior to the COVID-19 pandemic, false-positive results were found among samples known to contain potentially cross-reactive antibodies, especially HIV antibodies (DiaSorin and Abbott assays) and HCV antibodies (DiaSorin assay), as has been previously reported with other serologic assays (18, 19). Caution should be used when interpreting results in subjects known to have potentially cross-reactive serum antibodies. The Roche assay did not produce many false-positive results in samples collected from patients with antibodies against infectious agents but produced them in a sample from a patient with known autoantibodies, suggesting potential cross-reactivity in subjects with autoimmune conditions. Only a limited number of such samples were included in our study; more definitive conclusions cannot be drawn, and further investigation is required.

Among 64 samples from SARS-CoV-2 PCR-negative patients with respiratory symptoms during the pandemic, each assay produced lower specificity values than those obtained using the prepandemic control subject cohort. However, the sample size of the former cohort was small, and these differences did not reach statistical significance. It is possible that some control subjects with respiratory symptoms had true COVID-19 infection that was missed by RNA testing. Indeed, one sample—and only one—tested positive using all three serologic assays, suggesting the detection of COVID-19-specific antibodies. This sample was from a healthy male in his 30s who presented in April 2020 with fever, anosmia, shortness of breath, myalgias, and acute cardiomyopathy. Multiple respiratory SARS-CoV-2 RNA tests

TABLE 7 Guidance for selecting a SARS-CoV-2 serologic testing strategy depending on intended use and COVID-19 prevalence[a]

Intended use	Most valued test characteristic(s)	Rationale	Suggested testing strategy if disease prevalence is:			
			Low (≤1%)	Moderate (3%)	High (10%)	Very high (25%)
Routine diagnosis of COVID-19 in patients with recent URI symptoms and negative PCR result from respiratory tract sample(s)	High sensitivity in early infection; high NPV (≥95%)	Important not to miss true cases; false-positive results are not desirable but are more tolerable than false-negative results	Single Ab test	Single Ab test	Single Ab test	Single Ab test
Assessing candidacy for exptl COVID-19 therapeutics in PCR-negative patients with respiratory symptoms	High PPV (≥95%) and NPV (≥95%)	Need high NPV to ensure appropriate candidates get needed therapeutics; need high PPV so that inappropriate candidates do not receive drugs that are in short supply	2-test strategy	2-test strategy	2-test strategy	Single nucleocapsid Ab test or 2-test strategy
Identifying potential convalescent-phase plasma donors	Very high PPV (>99%)	Important that plasma donors have SARS-CoV-2 serum antibodies; high NPV is less important because potential plasma donors are abundant	2-test strategy	2-test strategy	2-test strategy	2-test strategy
Assessing for past exposure in population studies	High sensitivity in convalescence and high NPV (≥95%)	High PPV not critical because results will not be used to inform clinical decision-making; more important to identify all true cases	Single Ab test	Single Ab test	Single Ab test	Single Ab test
Assessing for past exposure to determine immune status	Strong correlation between detected antibody response and presence of neutralizing antibodies; high PPV (≥95%) and NPV (≥95%)	None of the tests evaluated in this study is claimed to be a tool for determining immunity to SARS-CoV-2 infection; however, studies suggest that spike protein antibodies correlate with neutralizing antibodies better than nucleocapsid antibodies do[b]	Assess for Ab response with a 2-test strategy including a spike protein Ab test; assess for immunity among seropositive patients using a neutralizing Ab test or another test known to correlate with a neutralizing Ab response	Assess for Ab response with a 2-test strategy including a spike protein Ab test; assess for immunity among seropositive patients using a neutralizing Ab test or another test known to correlate with a neutralizing Ab response	Assess for Ab response with a 2-test strategy including a spike protein Ab test; assess for immunity among seropositive patients using a neutralizing Ab test or another test known to correlate with a neutralizing Ab response	Assess for Ab response with a 2-test strategy including a spike protein Ab test; assess for immunity among seropositive patients using a neutralizing Ab test or another test known to correlate with a neutralizing Ab response

[a]Abbreviations: URI, upper respiratory tract infection; NPV, negative predictive value; PPV, positive predictive value; Ab, antibody.
[b]See reference 20.

were negative. For the purposes of this study, the subject remained categorized as COVID-19 negative, but clinical suspicion remained high.

When any of the commercial assays was paired with another in a two-test algorithm, specificity was 99.9% (1,267/1,268). In fact, only a single false-positive result was produced by any two-test pair, corresponding to the control subject mentioned above who tested positive using all three individual assays. This represents a significant increase in specificity for the Abbott and DiaSorin assays, but not for the Roche assay, compared with their use as stand-alone tests. Still, pairing any assay (including the Roche assay) with another in a two-test algorithm substantially increased the PPV compared with those of the individual assays alone, at assumed prevalence rates ranging from 0.1% to 25%. Interestingly, this improvement occurred even when the Abbott and Roche assays were paired, which use the same antigenic target (nucleocapsid protein). For this test pair, orthogonality may derive from the use of different test formats rather than different antigenic targets. Although pairing assays in two-test algorithms did reduce sensitivity compared with the constituent assays alone, particularly in samples collected from subjects 7 to 14 days after symptom onset, the effect on the NPV was minor. The difference between the NPV of each test pair and the NPVs of the individual component assays was 2.2% or lower at assumed prevalence rates of 0.1% to 25%.

When selecting a commercial SARS-CoV-2 serologic assay and deciding whether to verify positive results with an orthogonal assay, clinical laboratorians must consider many factors beyond the fixed test characteristics of sensitivity and specificity. Relevant variables include the COVID-19 disease prevalence in the tested population, the intended use of results, and practical considerations such as sample throughput, test complexity, reagent availability, and cost per reportable. Correlation with immunity should also be considered, depending on the assay's intended use: spike protein antibodies have been associated with neutralizing activity, although their ability to predict immunity to SARS-CoV-2 infection remains unknown (20). Table 7 provides proposed guidance for test selection and for choosing between a single-test and a two-test approach, based on our findings. In general, we believe that there are select circumstances under which a two-test approach would add value but that a single-test approach is adequate for the bulk of testing.

Our study has several limitations. First, the majority of serum samples were frozen prior to analysis. Second, the DOSO for the COVID-19-positive patients was determined by medical record review, which relies upon patient self-reporting that can be faulty. Errors in determining the DOSO would affect sensitivity determinations in our study, although efforts were made to ensure accuracy, including the use of a standardized DOSO definition and consensus between independent reviewers for each determination. Finally, cross-reactivity of the three assays in subjects with underlying autoimmune conditions was not fully explored.

In conclusion, the commercial nucleocapsid antibody tests evaluated in this study outperformed the spike protein antibody assay in terms of overall clinical sensitivity and specificity. With regard to sensitivity, differences were most pronounced in samples collected 7 to 14 days after symptom onset, whereas all three assays produced high (and comparable) sensitivity using samples collected $>$14 days after symptom onset. With regard to specificity, only the Roche nucleocapsid antibody assay reached the proposed target of $\geq$99.5% overall specificity suggested by the CDC as being adequate for use as a stand-alone test. Our findings demonstrate that pairing orthogonal tests in a two-test algorithm substantially improves the PPV compared with the individual assays alone, with little effect on the NPV.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We are grateful to Denise Zimmerman, Andrea Nixon, Janice Stefanski, Nancy Stanganelli, Chris Lofgren, and David Griggs for their excellent technical assistance.

This work was unfunded.

S.E.T. receives grant support for COVID-19-related work from the Centers for Disease Control and Prevention. J.A.B. has received research support from Zeus Scientific, bioMérieux, Immunetics, Alere, DiaSorin, the Bay Area Lyme Foundation (BALF), and the National Institute of Allergy and Infectious Diseases (NIAID) (award 1R21AI119457-01) for unrelated projects. J.A.B. has served as a paid consultant to T2 Biosystems, DiaSorin, and Roche Diagnostics. M.A. has equity and receives consulting fees from Day Zero Diagnostics. All other authors declare no competing interests.

REFERENCES

1. Sethuraman N, Jeremiah JS, Ryo A. 2020. Interpreting diagnostic tests for SARS-CoV-2. JAMA 323:2249–2251. https://doi.org/10.1001/jama.2020.8259.
2. Wolfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Muller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brunink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C. 2020. Virological assessment of hospitalized patients with COVID-2019. Nature 581:465–469. https://doi.org/10.1038/s41586-020-2196-x.
3. Xiao AT, Tong YX, Zhang S. 19 April 2020. Profile of RT-PCR for SARS-CoV-2: a preliminary study from 56 COVID-19 patients. Clin Infect Dis https://doi.org/10.1093/cid/ciaa460.
4. Okba NMA, Müller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, Lamers MM, Sikkema RS, de Bruin E, Chandler FD, Yazdanpanah Y, Le Hingrat Q, Descamps D, Houhou-Fidouh N, Reusken CBEM, Bosch BJ, Drosten C, Koopmans MPG, Haagmans BL. 2020. Severe acute respiratory syndrome coronavirus 2-specific antibody responses in coronavirus disease 2019. Emerg Infect Dis 26:1478–1488. https://doi.org/10.3201/eid2607.200841.
5. Gronvall G, Connell N, Kobokovich A, West R, Warmbrod KL, Shearer MP, Mullen L, Inglesby T. 2020. Developing a national strategy for serology (antibody testing) in the United States. Johns Hopkins Center for Health Security, Baltimore, MD. https://www.centerforhealthsecurity.org/our-work/publications/developing-a-national-strategy-for-serology-antibody-testing-in-the-US. Accessed 5 June 2020.
6. Patel R, Babady E, Theel ES, Storch GA, Pinsky BA, St George K, Smith TC, Bertuzzi S. 2020. Report from the American Society for Microbiology COVID-19 International Summit, 23 March 2020: value of diagnostic testing for SARS-CoV-2/COVID-19. mBio 11:e00722-20. https://doi.org/10.1128/mBio.00722-20.
7. Centers for Disease Control and Prevention. 2020. Interim guidelines for SARS-CoV-2 antibody testing. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html. Accessed 30 June 2020.
8. Kohmer N, Westhaus S, Ruhl C, Ciesek S, Rabenau HF. 2020. Brief clinical evaluation of six high-throughput SARS-CoV-2 IgG antibody assays. J Clin Virol 129:104480. https://doi.org/10.1016/j.jcv.2020.104480.
9. Merrill AE, Jackson JB, Ehlers A, Voss D, Krasowski MD. 27 July 2020. Head-to-head comparison of two SARS-CoV-2 serology assays. J Appl Lab Med https://doi.org/10.1093/jalm/jfaa125.
10. Perkmann T, Perkmann-Nagele N, Breyer M-K, Breyer-Kohansal R, Burghuber OC, Hartl S, Aletaha D, Sieghart D, Quehenberger P, Marculescu R, Mucher P, Strassl R, Wagner OF, Binder CJ, Haslacher H. 10 August 2020. Side by side comparison of three fully automated SARS-CoV-2 antibody assays with a focus on specificity. Clin Chem https://doi.org/10.1093/clinchem/hvaa198.
11. Mead P, Petersen J, Hinckley A. 2019. Updated CDC recommendation for serologic diagnosis of Lyme disease. MMWR Morb Mortal Wkly Rep 68:703. https://doi.org/10.15585/mmwr.mm6832a4.
12. Workowski KA, Bolan GA, Centers for Disease Control and Prevention. 2015. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recommend Rep 64:1–137.
13. The White House, Centers for Disease Control and Prevention, US Food and Drug Administration. 2020. Testing blueprint: opening up America again. The White House, Washington, DC.
14. Centers for Disease Control and Prevention. 2020. CDC COVID-19 data tracker. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/covid-data-tracker/index.html. Accessed 6 June 2020.
15. Tan YJ, Goh PY, Fielding BC, Shen S, Chou CF, Fu JL, Leong HN, Leo YS, Ooi EE, Ling AE, Lim SG, Hong W. 2004. Profiles of antibody responses against severe acute respiratory syndrome coronavirus recombinant proteins and their potential use as diagnostic markers. Clin Diagn Lab Immunol 11:362–371. https://doi.org/10.1128/cdli.11.2.362-371.2004.
16. Woo PCY, Lau SKP, Wong BHL, Tsoi H, Fung AMY, Kao RYT, Chan K, Peiris JSM, Yuen K. 2005. Differential sensitivities of severe acute respiratory syndrome (SARS) coronavirus spike polypeptide enzyme-linked immunosorbent assay (ELISA) and SARS coronavirus nucleocapsid protein ELISA for serodiagnosis of SARS coronavirus pneumonia. J Clin Microbiol 43:3054–3058. https://doi.org/10.1128/JCM.43.7.3054-3058.2005.
17. Sakamoto S, Putalun W, Vimolmangkang S, Phoolcharoen W, Shoyama Y, Tanaka H, Morimoto S. 2018. Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. J Nat Med 72:32–42. https://doi.org/10.1007/s11418-017-1144-z.
18. Jonas G, Pelzer C, Beckert C, Hausmann M, Kapprell HP. 2005. Performance characteristics of the ARCHITECT anti-HCV assay. J Clin Virol 34:97–103. https://doi.org/10.1016/j.jcv.2005.08.001.
19. Hernández-Aguado I, Bolumar F, Moreno R, Pardo FJ, Torres N, Belda J, Espacio A, Valencian Study Group on HIV Epidemiology. 1998. False-positive tests for syphilis associated with human immunodeficiency virus and hepatitis B virus infection among intravenous drug abusers. Eur J Clin Microbiol Infect Dis 17:784–787. https://doi.org/10.1007/s100960050186.
20. Ju B, Zhang Q, Ge J, Wang R, Sun J, Ge X, Yu J, Shan S, Zhou B, Song S, Tang X, Yu J, Lan J, Yuan J, Wang H, Zhao J, Zhang S, Wang Y, Shi X, Liu L, Zhao J, Wang X, Zhang Z, Zhang L. 2020. Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature 584:115–119. https://doi.org/10.1038/s41586-020-2380-z.