

Differences in Performance Characteristics Among Four High-Throughput Assays for the Detection of Antibodies Against SARS-CoV-2 Using a Common Set of Patient Samples

David M. Manthei, MD, PhD, Jason F. Whalen, Lee F. Schroeder, MD, PhD, Anthony M. Sinay, C(ASCP), Shih-Hon Li, MD, PhD, Riccardo Valdez, MD, Donald A. Giacherio, PhD, and Carmen Gherasim, PhD

From the Department of Pathology, University of Michigan, Ann Arbor.

Key Words: SARS-CoV-2; Serology; Immunoassay; Specificity; Sensitivity; COVID-19; Coronavirus

Am J Clin Pathol February 2021;155:267-279

DOI: 10.1093/AJCP/AQAA200

ABSTRACT

Objectives: Serologic testing for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has experienced a changing landscape of available assays coupled with uncertainty surrounding performance characteristics. Studies are needed to directly compare multiple commercially available assays.

Methods: Residual serum samples were identified based on SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR) testing, clinical test results, and collection dates. Serum samples were analyzed using assays from four different manufacturers: DiaSorin anti-SARS-CoV-2 S1/S2 IgG, EUROIMMUN anti-SARS-CoV-2 IgG ELISA, Roche Elecsys anti-SARS-CoV-2, and Siemens SARS-CoV-2 Total antibody assays.

Results: Samples from SARS-CoV-2 RT-PCR-positive patients became increasingly positive as time from symptom onset increased. For patients with latest sample 14 or more days after symptom onset, sensitivities reached 93.1% to 96.6%, 98.3%, and 96.6% for EUROIMMUN, Roche, and Siemens assays, respectively, which were superior to the DiaSorin assay

Key Points

- Anti-severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) total antibody assays from Roche and Siemens performed comparable to each other and superior to DiaSorin and EUROIMMUN IgG assays in detecting prior infection by SARS-CoV-2.
- The use of a common set of serum samples on each assay provides a direct comparison of performance, including variability in detection limits and repeatedly false-positive samples.
- Understanding the strengths and limitations of serology testing platforms is necessary if tiered or confirmatory testing is performed.

at 87.7%. The specificity of Roche and Siemens assays was 100% and superior to DiaSorin and EUROIMMUN assays, which ranged from 96.1% to 97.0% and 86.3% to 96.4%, respectively.

Conclusions: Laboratories should be aware of the advantages and limitations of serology testing options for SARS-CoV-2. The specificity and sensitivity achieved by the Roche and Siemens assays would be acceptable for testing in lower-prevalence regions and have the potential of orthogonal testing advantages if used in combination.

Clinical laboratory serologic testing for antibodies directed against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), has been beset by a series of problems. Serologic tests were initially allowed to be distributed by any manufacturer, although some claims were lacking in quality, leading to removal from the US Food and Drug Administration (FDA) “notification list” of assays with pending or approved emergency use authorization.¹ As more manufacturers with established in vitro diagnostics history have released serology assays with larger premarket studies, the presumption is that more consistent and high-quality assays are available. Many SARS-CoV-2 serology assay manufacturers report impressive performance results, which are summarized by the FDA based on review of submitted data.² However, given a paucity of real-world and consistent evaluations of these assays, the FDA has partnered with the National Institutes of Health and Centers for Disease Control and Prevention (CDC) to perform limited independent evaluations of some assays.³ Although there are multiple evaluations of SARS-CoV-2 serology assays,^{4,22} there is a relative paucity of robust evaluations of more recently released tests. As such, there is need for greater direct comparisons across multiple, different platforms. Many laboratories have little real-world information about performance limitations when choosing among assays.

The best use of SARS-CoV-2 serologic testing remains an open question.²³ Regardless of whether clinical or epidemiologic use is planned, the predictive values are important to consider, particularly in areas with low pretest probability (eg, mass screening). Both the throughput of an assay and in particular its specificity are crucial parameters for clinical laboratory implementation if large-scale serology testing is desirable in a low-prevalence environment.

Approaches for anti-SARS-CoV-2 immunoassays involve differing antigens (eg, nucleocapsid vs spike protein and full-length vs subdomains) and/or immunoglobulin isotypes (eg, immunoglobulin G [IgG], immunoglobulin M [IgM], total). A preferred combination may evolve depending on the desired purpose of testing as our understanding of COVID-19 matures, including breadth and duration of antibody responses, neutralizing effects of antibodies, and antigen choices for vaccines. It is possible that appropriate assays for detection of prior viral infection may be different from those to confirm a vaccine response.

Given the uncertain and shifting landscape for clinical laboratories, we sought to compare a common set of serum samples across four commercial assays as part of our evaluation for potential implementation of

high-throughput testing. These findings will help inform the strengths and limitations of each assay in a directly comparable manner.

Materials and Methods

Serum Collection and Clinical Characteristics

Studies were performed as part of standard clinical laboratory assay evaluation and quality assurance studies and in accordance with the Declaration of Helsinki. We identified frozen serum samples routinely saved for laboratory quality assurance purposes that were collected at prepandemic time points, presumptively negative for SARS-CoV-2. Sample selection for potential cross-reactivity was based on FDA suggestions for cross-reactivity evaluations, including viral (eg, seasonal coronaviruses, influenza A/B, hepatitis B) and autoimmune conditions.²⁴ We identified residual presumptively negative clinical samples from individuals with prior non-SARS-CoV-2 viral infections or abnormal clinical autoimmune testing. Clinical reverse transcription polymerase chain reaction (RT-PCR) testing for SARS-CoV-2 was performed by the Abbott Realtime SARS-CoV-2 assay on m2000 or DiaSorin Simplexa COVID-19 Direct Kit on LIAISON MDX. Residual serum samples from inpatients known to be SARS-CoV-2 RT-PCR positive were identified by scripted queries in the laboratory information system with confirmation of molecular diagnosis, symptom onset, and potentially relevant comorbidities or medications performed by manual review of the electronic health record. In addition, we retained serum samples from outpatients with known results of SARS-CoV-2 RT-PCR testing, in which samples drawn for clinical serologic testing were due to current symptoms or investigation at a convalescent time point.

Of the RT-PCR-confirmed COVID-19 patients with residual clinical serum specimens, a single sample was chosen for analysis to primarily characterize serologic responses at approximately 10 to 14 days after symptom onset. In addition, a subset of hospitalized COVID-19 patients had at least four serial samples available, which were all analyzed to determine the profile of their serologic responses.

Testing Methods

Serum samples were analyzed on each of four commercial platforms (briefly described below and in **Table 1**) according to manufacturers' instructions. If a sample was analyzed multiple times as part of precision or other studies, the first result was used in this analysis for

comparison. Testing was performed across multiple days, and serum samples were frozen (-20°C) until used with up to two freeze-thaw cycles in total. A schematic of methods is shown in **Figure 1**.

IgG

The anti-SARS-CoV-2 IgG ELISA (EUROIMMUN) uses recombinant S1 protein subunit as a capture antigen with subsequent IgG antibody detection based on binding of anti-human IgG conjugated to an enzyme to catalyze a colorimetric reaction.²⁵ Testing was performed with controls and patient samples in singlet. Enzyme-linked immunosorbent assays (ELISAs) were performed manually with normalization to the single calibrator.

The LIAISON SARS-CoV-2 S1/S2 IgG assay (DiaSorin) uses recombinant spike protein S1 and S2

subunits on magnetic particles as capture antigens for a chemiluminescent immunoassay, with mouse anti-human IgG conjugated with an isoluminol derivative for detection.²⁶ The assay output ranges from 3.8 to 400 arbitrary units (AU)/mL.

Total Antibody

The Elecsys anti-SARS-CoV-2 assay (Roche Diagnostics) uses recombinant biotinylated and ruthenium-labeled nucleocapsid protein in an electrochemiluminescent immunoassay.²⁷ Streptavidin-captured antigen-antibody complexes provide chemiluminescent signal after electrical stimulation. No manufacturer controls were available; pooled negative serum samples and a known strongly positive sample diluted in negative serum samples were used as negative and

Table 1
Characteristics of Commercial Anti-SARS-CoV-2 Assays Used in This Comparative Study

Manufacturer	Assay	Analyte	Capture Antigen ^a	Method	Analyzer	Output	Cutoff Values		
							Negative	Indeterminate	Positive
EUROIMMUN	Anti-SARS-CoV-2 IgG ELISA	IgG	S1	ELISA	BioTek 800 TS	Ratio to a single calibrator	<0.8	0.8 to <1.1	≥1.1
DiaSorin	LIAISON SARS-CoV-2 S1/S2 IgG	IgG	S1 and S2	CLIA	LIAISON XL	Index based on 2 calibrators	<15	NA	≥15
Roche	Elecsys Anti-SARS-CoV-2	Total immunoglobulins	Nucleocapsid	ECLIA	cobas e411	Index based on 2 calibrators	<1.0	NA	≥1.0
Siemens	SARS-CoV-2 Total (COV2T)	Total immunoglobulins	RBD	CLIA	Centaur XP	Index based on 2 calibrators	<1.0	NA	≥1.0

CLIA, chemiluminescence immunoassay; ECLIA, electrochemiluminescence immunoassay; ELISA, enzyme-linked immunosorbent assay; IgG, immunoglobulin G; NA, not applicable; RBD, receptor-binding domain; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

^aS1 and S2 are subunits of the spike protein; the RBD is a domain within the S1 subunit.

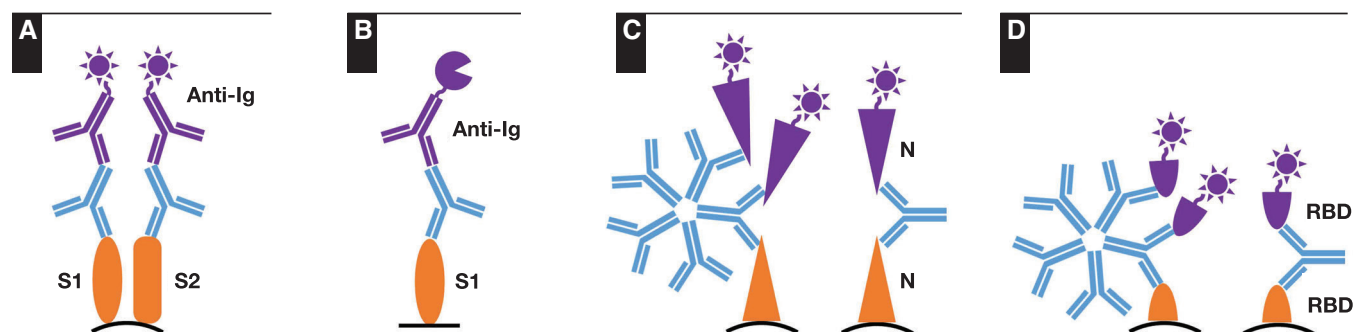


Figure 1 Schematic of methods for DiaSorin (A), EUROIMMUN (B), Roche (C), and Siemens (D) assays. Capture antigens are orange, anti-SARS-CoV-2 antibodies from patient serum samples are blue, and conjugated detection methods are purple, showing luminescent reporters for the DiaSorin, Roche, and Siemens assays (A, C, D) and an enzymatic reporter for the EUROIMMUN assay (B). S1 and S2 indicate subunits of Spike protein and RBD denotes receptor-binding domain of S1. N denotes nucleocapsid protein.

positive quality control (QC) material, respectively, with aliquots frozen for daily QC.

The SARS-CoV-2 Total assay (Siemens) uses recombinant S1 subunit receptor-binding domain (RBD) as biotinylated and acridinium ester-conjugated antigens for a chemiluminescent immunoassay.²⁸ Streptavidin-captured sandwich complexes generate light after excitation. Testing was bracketed by daily cleaning procedures. The output spans an index value of 0.05 to 10.00.

Statistical Analysis

Assay results were compared for categorical agreement by group based on expected serology result and SARS-CoV-2 RT-PCR status. For some graphical comparisons, the numeric value was divided by the cutoff (15 AU/mL and 1.1 for the DiaSorin and EUROIMMUN assays, respectively) for normalization with an upper limit of 10 to align with Siemens result limits. The χ^2 and Wilcoxon rank-sum tests were performed as appropriate with two tails and significance set at $P = .05$. Confidence intervals were calculated by the Clopper-Pearson method for sensitivity and specificity; confidence intervals for positive predictive values (PPVs) were calculated from

bootstrapped likelihood ratios in the R statistical environment using v.3.5.2, using the bootLR package.²⁹

Results

Patient Demographics and Sample Collection

In total, 310 individual serum specimens were analyzed across all platforms; 11 additional samples were compared across a subset of platforms due to volume limitations. Seventy-four prepandemic presumptive SARS-CoV-2–negative samples were identified and 63 additional routine samples were selected for potential cross-reactivity, which were also archived before widespread COVID-19 (Table 2). Samples from 53 outpatients with RT-PCR–negative results and clinically collected SARS-CoV-2 serology samples were separately identified (Table 2).

A total of 131 serum samples from 68 patients who were RT-PCR positive for SARS-CoV-2 were analyzed, with the range of days after symptom onset shown in Table 2. Of the 131 samples, 95 were from 33 hospitalized patients; all of these samples were collected less than

Table 2
Characteristics of Patients With Serum Samples Used in This Comparative Study

Sample Type (Year Collected)	Samples (Unique Patients), No.	Sex, Male/Total, No.	Age, Range (Median) y	Days After Symptom Onset, Range (Median)	Hospital Admission, No.
Prepandemic presumptive negative (2016-2019)	74 (74)				
Samples with demographics	31 (31)	17/31	12-79 (46)		
Cross-reactivity panel (2020)	63 (63)				
Seasonal coronavirus	13 (13)	6/13	2-83 (55)		
HKU1	7				
OC43	2				
NL63	2				
229E	2				
Hepatitis B	10 (10)	5/10	22-72 (48)		
Hepatitis C	10 (10)	7/10	20-74 (62)		
Autoimmune	30 (30)	4/30	12-71 (40)		
ANA	10				
ENA	7				
Anti-dsDNA	17				
RF	6				
Increased IgG	5				
M-protein	1				
Anti-EBV	1				
Anti-HSV 1 and 2	1				
Clinical RT-PCR negative (2020)	53 (53)	13/53	16-64 (40)	0-52 ^a (4)	
Symptom onset >14 d	25 (25)	6/25	17-61 (41)	15-52 (32)	
Symptom onset <14 d	28 (28)	7/28	16-64 (38)	0-4 ^a (0)	
Clinical RT-PCR positive (2020)	131 (68)	29/68	23-65 (49)	5-59 (20)	35
Retrospective serology	36 (35)	10/35	23-65 (38)	18-59 ^b (41)	2
Prospective medical care	95 (33)	19/33	27-85 (63)	5-29 (14)	33

ANA, anti-nuclear antibody by immunofluorescence; dsDNA, double-stranded DNA; EBV, Epstein-Barr virus; ENA, anti-nuclear antibody by multiplex extractable nuclear antigen; HSV, herpes simplex virus; RF, rheumatoid factor; RT-PCR, reverse transcription polymerase chain reaction.

^aOne patient without symptoms.

^bRange excluding first sample on day 0 of a patient with two samples.

1 month after symptom onset. Thirty-five separate outpatients had serum samples collected after prior positive SARS-CoV-2 RT-PCR testing, with time between serology sample collection and symptom onset being generally greater than 1 month.

Assay Reproducibility

QC material was run daily with each assay, and all assays had coefficients of variation (CVs) of less than 5% for positive controls. Specifically, the means and CVs were 39.6 AU/mL and 3.5% for DiaSorin, 3.23 AU/mL and 4.1% for EUROIMMUN, 3.82 AU/mL and 3.7% for Roche, and 1.93 AU/mL and 4.7% for Siemens. The EUROIMMUN negative control averaged 0.10 with a CV of 14.6%, and the Roche negative control averaged 0.08 with a CV of 8.9%. For DiaSorin, since all negative samples were less than 3.8 AU/mL, the relative light units (RLUs) were alternatively analyzed, with negative control averaging 1,093 RLUs and a CV of 9.7%. Similarly, for Siemens, the average RLUs and CV for negative controls were 13,284 and 4.4%, respectively.

Specificity

When characterizing the 53 outpatient SARS-CoV-2 RT-PCR–negative samples, two samples were positive for SARS-CoV-2 antibodies across all four platforms. On additional review, one patient with serum collected at 19 days after symptom onset (cough, gastrointestinal symptoms, loss of taste/smell, muscle aches) had negative day 1 SARS-CoV-2 RT-PCR testing but also had a positive day 1 anti-SARS-CoV-2 IgG send-out result (performed at Viracor Eurofins using Diazyme DZ-Lite SARS-CoV-2 IgG CLIA), indicating likely exposure to SARS-CoV-2. The other patient had serum collected at 38 days after symptom onset (fevers, cough, congestion body aches) with day 1 testing negative for SARS-CoV-2

RT-PCR, influenza A/B, and respiratory syncytial virus testing. Taken together, we considered these two patients to have false-negative or postacute RT-PCR test results rather than false-positive serology results, and these two samples were excluded from subsequent calculations.

As such, of the remaining expected negative samples, neither the Roche nor Siemens assays had any false-positive results, yielding specificities of 100% (Table 3). The DiaSorin assay had specificity between 96.1% and 97.0% across groupings of negative samples. The EUROIMMUN assay specificity ranged from 86.3% to 96.4%, depending on the group of negative samples and whether indeterminate results were interpreted as positive or negative. The specificity of the Roche and Siemens assays was significantly better than that of both DiaSorin and EUROIMMUN assays ($P = .01$ each compared to DiaSorin, $P < .001$ and $P = .007$ each compared with EUROIMMUN with indeterminate results as positive or negative, respectively; χ^2 proportions). The specificity of the DiaSorin assay was greater than that of the EUROIMMUN assay when considering indeterminate results as positive ($P = .03$; χ^2 proportions). The distribution of assay values for all expected negative samples is shown in Figure 2.

Nineteen samples yielded falsely positive or indeterminate results by DiaSorin, EUROIMMUN, or both assays; none of these 19 samples were positive by Roche or Siemens assays (Table 4). A subset of false-positive results was common to both DiaSorin and EUROIMMUN assays; three of the four DiaSorin assay false-positive results from prepandemic and cross-reactivity panels were either positive or indeterminate by the EUROIMMUN assay. In addition, two other positive or indeterminate samples by the EUROIMMUN assay also had a numeric value on the DiaSorin assay between 10 and 15 AU/mL; while these are not positive results, they are in a higher range than typical negative results from the DiaSorin assay. Of

Table 3
Specificity of Assays^a

Manufacturer	All Expected Negative Samples ^b		Prepandemic and Cross-Reactivity		Clinical RT-PCR Negative ^b	
	Positive/Samples, No.	Specificity (95% CI)	Positive/Samples, No.	Specificity (95% CI)	Positive/Samples, No.	Specificity (95% CI)
DiaSorin	6/185	96.8 (93.1-98.8)	4/134	97.0 (92.5-99.2)	2/51	96.1 (86.5-99.5)
EUROIMMUN, indeterminate = positive	16/188	91.5 (86.6-95.1)	9/137	93.4 (87.9-97.0)	7/51	86.3 (73.7-94.3)
EUROIMMUN, indeterminate = negative	7/188	96.3 (92.5-98.5)	5/137	96.4 (91.7-98.8)	2/51	96.1 (86.5-99.5)
Roche	0/188	100 (98.1-100)	0/137	100 (97.3-100)	0/51	100 (93.0-100)
Siemens	0/188	100 (98.1-100)	0/137	100 (97.3-100)	0/51	100 (93.0-100)

CI, confidence interval; RT-PCR, reverse transcription polymerase chain reaction.

^aSpecificity is calculated for all samples expected to be negative and pertinent subgroups. For EUROIMMUN, calculations are provided for categorizing indeterminate results as positive or negative.

^bExcluding two samples that were positive on all assays and adjudicated as presumptive RT-PCR testing false negatives.

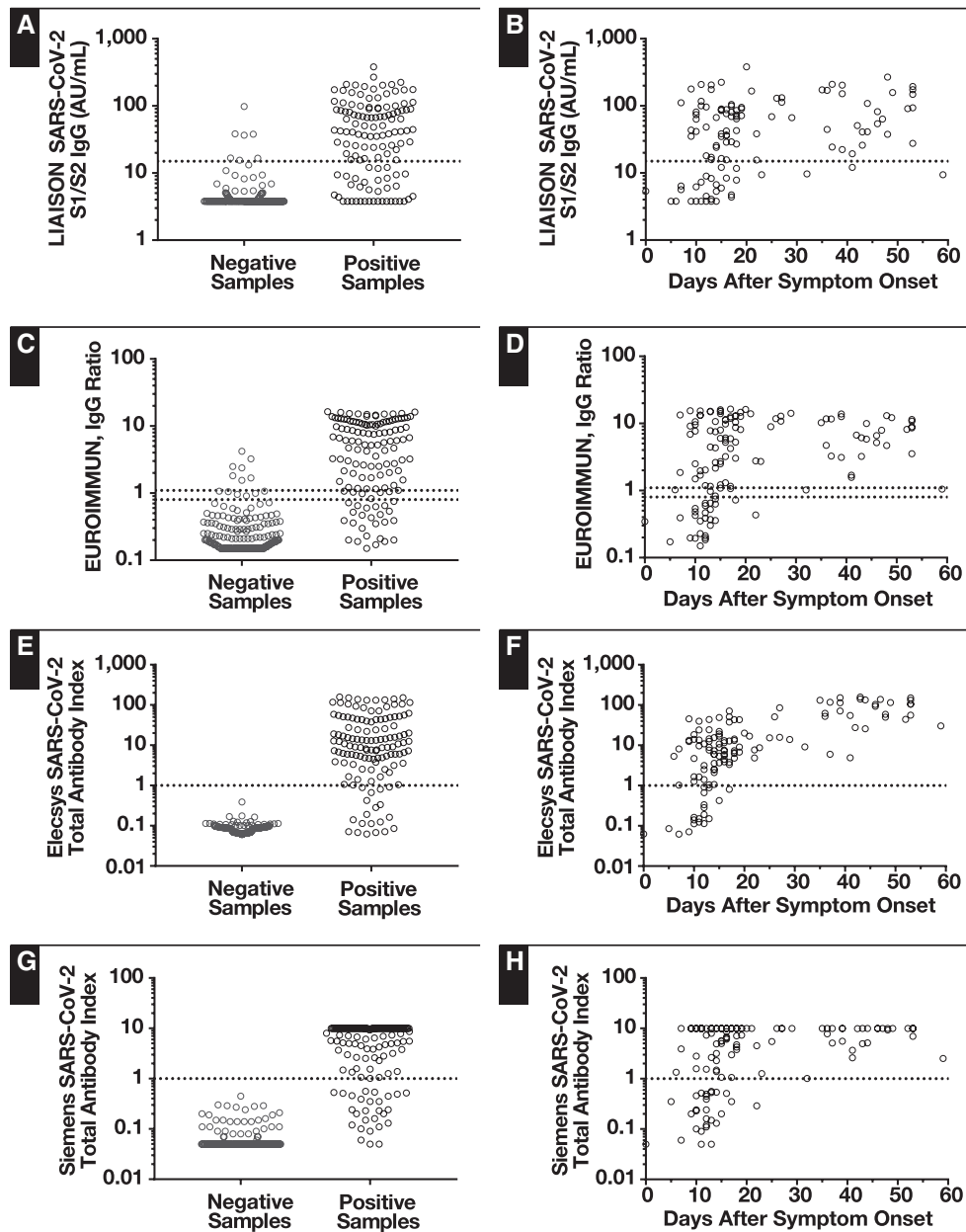


Figure 2 Serology assay values are stratified by expected result (A, C, E, G). Expected positives are SARS-CoV-2 RT-PCR positive while expected negatives include RT-PCR negative patients, pre-pandemic samples, or samples assayed for potential cross-reactivity. Assay results for RT-PCR-positive samples are also shown based on days after symptom onset (B, D, F, H). Thresholds for individual assays are denoted by dotted lines.

the 51 remaining outpatient cases negative by RT-PCR, the DiaSorin assay read out two positive results while the EUROIMMUN assay yielded seven positive and indeterminate results on separate samples (Table 4). Two of these patients who were tested near symptom onset (0 and 4 days after onset) also had repeat serologic testing 15 to 20 days later that was negative by alternative platforms, further supporting the classification as false-positive results. There were no differences between rates of false

positives within the DiaSorin or EUROIMMUN assays across categories of expected negative samples (cross-reactivity, prepandemic, and RT-PCR negative samples; $P = .66$ and $P = .34$, respectively; χ^2).

Sensitivity

Sensitivities for all assays improved as time from symptom onset to sample collection increased. Including

Table 4
Samples With Discordant Results or Immunosuppression^a

Sample Type	DiaSorin	EUROIMMUN	Roche	Siemens	Age, y	Sex	Days After Symptom Onset	Other
	Arbitrary Units (15.00) ^b	Ratio (0.80-1.10) ^b	Index (1.00) ^b	Index (1.00) ^b				
Negatives								
Prepandemic	<3.80	1.68	0.09	<0.05	NA	NA	NA	
Prepandemic	98.2	0.41	0.10	<0.05	NA	NA	NA	
Prepandemic	36.7	<i>0.96</i>	0.08	<0.05	NA	NA	NA	
Prepandemic	16.6	2.49	0.09	<0.05	NA	NA	NA	
Prepandemic	10.9	1.55	0.10	<0.05	56	M	NA	
Prepandemic	4.79	<i>1.07</i>	0.09	<0.05	30	F	NA	Also positive for rubella IgG
Prepandemic	8.60	4.29	0.10	<0.05	53	M	NA	Neurofibromatosis type 1 and adrenal adenoma
Cross-reactive	11.5	3.24	0.13	<0.05	11	F	NA	NL63 seasonal coronavirus
Cross-reactive	15.5	<i>1.07</i>	0.09	<0.05	22	F	NA	Hepatitis B
Cross-reactive	9.25	<i>0.91</i>	0.08	<0.05	36	F	NA	ANA speckled at 1:640
RT-PCR negative	<3.80	1.81	0.08	<0.05	16	M	NA	No symptoms, preprocedure test, and Noonan syndrome
No RT-PCR test	37.9	0.29	0.06	0.08	17	M	39	Foot rash of concern for possible post-COVID-19 sequelae
RT-PCR negative ^c	16.7	0.36	0.09	<0.05	49	F	4	Eosinophilic esophagitis, hypertension, liver mass
RT-PCR negative	<3.80	<i>0.95</i>	0.07	0.11	28	F	35	
RT-PCR negative ^d	5.07	2.38	0.07	0.11	60	F	0	Pernicious anemia
RT-PCR negative	<3.80	<i>1.01</i>	0.07	0.11	49	F	34	Gilbert syndrome
RT-PCR negative	<3.80	<i>0.89</i>	0.06	<0.05	51	M	19	
RT-PCR negative	8.20	<i>0.99</i>	0.06	0.15	37	M	16	
RT-PCR negative	4.01	<i>1.05</i>	0.06	0.15	33	F	15	
Positives								
RT-PCR positive	9.71	<i>1.02</i>	9.15	1.01	37	F	32	
RT-PCR positive	9.42	<i>1.05</i>	30.4	2.53	27	F	59	
RT-PCR positive	38.4	0.43	4.83	0.29	50	F	22	
Immunosuppression								
RT-PCR positive	41	5.83	25.8	5.15	28	F	44	Systemic lupus erythematosus on hydroxychloroquine
RT-PCR positive	268	13.06	58.5	>10	63	F	48	Rheumatoid arthritis on hydroxychloroquine and tofacitinib (tofacitinib paused during symptoms)
RT-PCR positive	12.1	1.57	4.88	2.60	59	F	41	Unclear rheumatologic condition on hydroxychloroquine
RT-PCR positive	7.29	1.70	9.71	>10	60	M	11	Lung transplant on tacrolimus
RT-PCR positive	24.0	<i>0.83</i>	15.5	1.48	59	M	14	Kidney transplant on mycophenolate and tacrolimus
	90.5	8.11	43.2	>10			19	

ANA, anti-nuclear antibody by immunofluorescence; COVID-19, coronavirus disease 2019; IgG, immunoglobulin G; RT-PCR, reverse transcription polymerase chain reaction.

^aDiscordant results are classified as nonconforming with expected results and/or other assays. Bold type indicates discordance, and italicized type indicates indeterminate results.

^bThreshold for positivity (and indeterminate range for EUROIMMUN).

^cRepeat testing by DiaSorin was 15.5 AU/mL. Testing 15 days later by EUROIMMUN was negative with ratio of 0.38.

^dTwenty days later tested by DiaSorin with negative results, <3.8 AU/mL.

all serum samples from RT-PCR-positive patients, sensitivities ranged between approximately 70% and 85%: 70.2% by the DiaSorin assay; 74.1% to 79.4% by the EUROIMMUN assay with indeterminate results

as negative and positive, respectively; 84.5% by the Roche assay; and 78.0% by the Siemens assay (Table 5). Figure 2 displays individual assay values for SARS-CoV-2 RT-PCR-confirmed patients both categorically

Table 5
Sensitivity of Assays^a

Sample Group	DiaSorin		EUROIMMUN, Indeterminate = Positive		EUROIMMUN, Indeterminate = Negative		Roche		Siemens	
	Positive/Samples, No.	Sensitivity (95% CI)	Positive/Samples, No.	Sensitivity (95% CI)	Positive/Samples, No.	Sensitivity (95% CI)	Positive/Samples, No.	Sensitivity (95% CI)	Positive/Samples, No.	Sensitivity (95% CI)
All RT-PCR+	87/124	70.2 (61.3-78.0)	104/131	79.4 (71.5-86.0)	97/131	74.1 (65.7-81.3)	109/129	84.5 (77.1-90.3)	103/131	78.6 (70.6-85.3)
All RT-PCR+ symptoms ≥ 10 d	83/114	72.8 (63.7-80.7)	98/121	81.0 (72.9-87.6)	92/121	76.0 (67.4-83.3)	103/119	86.6 (79.1-92.1)	97/121	80.2 (71.9-86.9)
All RT-PCR+ symptoms ≥ 12 d	75/98	76.5 (66.9-84.5)	88/104	84.6 (76.2-90.9)	82/104	78.9 (69.7-86.2)	92/103	89.3 (81.7-94.6)	87/104	83.7 (75.1-90.2)
All RT-PCR+ symptoms ≥ 14 d	68/82	82.9 (73.0-90.3)	79/84	94.1 (86.7-98.0)	74/84	88.1 (79.2-94.1)	81/83	97.6 (91.6-99.7)	78/84	92.9 (85.1-98.7)
All RT-PCR+ symptoms ≥ 16 d	56/65	86.2 (75.3-93.5)	63/65	96.9 (89.3-99.6)	60/65	92.3 (83.0-97.5)	64/65	98.5 (91.7-99.5)	62/65	95.4 (87.1-99.0)
All RT-PCR+ symptoms ≥ 18 d	44/50	88.0 (75.7-95.5)	48/50	96.0 (86.3-99.5)	46/50	92.0 (80.8-97.8)	50/50	100 (92.9-100)	49/50	98.0 (89.4-99.9)
Unique patients	56/65	86.2 (75.3-93.5)	64/68	94.1 (85.6-98.4)	62/68	91.2 (81.8-96.7)	63/67	94.0 (85.4-98.4)	63/68	92.7 (83.7-97.6)
Unique patients' symptoms ≥ 14 d on last sample	50/57	87.7 (76.3-94.9)	56/58	96.6 (88.1-99.6)	54/58	93.1 (83.3-98.1)	57/58	98.3 (90.8-99.9)	56/58	96.6 (88.1-99.6)

CI, confidence interval; RT-PCR, reverse transcription polymerase chain reaction.

^aSensitivities of each assay are determined at varying thresholds for number of days after onset of symptoms. For EUROIMMUN, calculations are provided for categorizing indeterminate results as positive or negative.

and based on specimen collection interval after symptom onset. Limiting analysis to samples collected at least 14 days after symptom onset results in increased sensitivity in all assays: 82.9% by the DiaSorin assay; 88.1% to 94.1% by the EUROIMMUN assay with indeterminate results as negative and positive, respectively; 97.6% by the Roche assay; and 92.9% by the Siemens assay. At this time point, the sensitivity of the DiaSorin assay was inferior to the Roche, Siemens, and EUROIMMUN (with indeterminate results as positive) assays ($P = .0015$, $P = .048$, and $P = .024$, respectively; χ^2 proportions). Reanalysis was performed by excluding the eight samples that were not performed on every platform, which yielded similar results where the DiaSorin assay was less sensitive than each of the other assays. The Roche assay at this time point was also significantly more sensitive than the EUROIMMUN assay (indeterminate results as negative, $P = .018$; χ^2 proportions). In addition, although patients were neither selected nor excluded based on immunosuppression status, five patients were on immunosuppressive medications (Table 4). Positivity rates for each assay were not altered for the latest test whether stratifying by age (<50 vs ≥50 years, $P > .3$ for all four assays; χ^2 proportions) or by the presence of immunosuppressive medications ($P = .08$ for DiaSorin, $P > .5$ for other assays; χ^2 proportions).

Fifteen hospitalized patients had serial samples assayed to follow antibody responses over time (Figure 3). Repeat chronologic samples tended to become positive earlier on the Roche assay (mean and median days since symptom onset to first positive result: DiaSorin, 14, 14; EUROIMMUN, 13.5, 15; Roche, 12.7, 13; and Siemens, 13.5, 14) but was only significantly different from the DiaSorin assay ($P = .03$; Wilcoxon rank sum).

Positive Predictive Values

Positive predictive values are shown in Table 6 based on the above-determined values for specificity from all expected negatives as well as for sensitivity from both all RT-PCR-positive patient samples and also limited to those acquired 14 or more days after symptom onset.

Discussion

We compared a set of serum samples across four high-throughput commercial serology platforms for anti-SARS-CoV-2, including IgG-specific assays against spike proteins from EUROIMMUN (S1) and DiaSorin (S1, S2), a total antibody assay against spike protein (S1 RBD) from Siemens, and a total antibody assay against

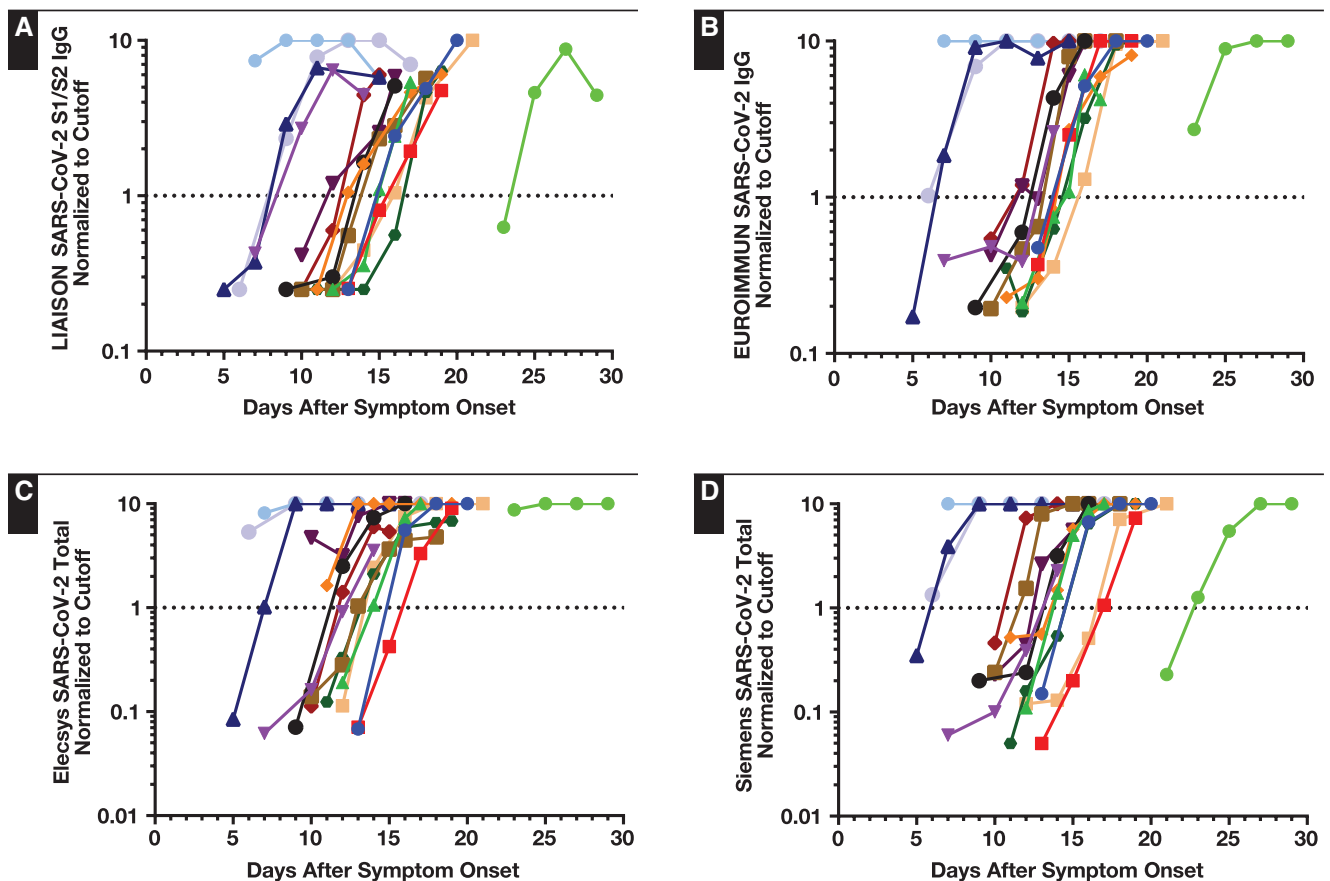


Figure 3 Normalized serology assay results for 15 SARS-CoV-2 RT-PCR–positive hospitalized patients with serial samples. Results were normalized to the positivity threshold (EUROIMMUN 1.1, DiaSorin 15 AU/mL) and limited to a maximum value of 10. Positivity corresponds to values ≥ 1.0 , denoted by dotted lines. The colors of individual patients are the same across DiaSorin (A), EUROIMMUN (B), Roche (C), and Siemens (D) assays.

nucleocapsid protein from Roche. We used a mix of RT-PCR–confirmed hospitalized COVID-19 patients and less ill outpatients. Our samples included multiple sources of potential cross-reactivity, including seasonal coronaviruses (Table 2). Despite limitations that symptoms could only be reviewed retrospectively, we noted increasingly positive results for samples collected at approximately 2 weeks after symptom onset on all four assays, similar to other studies.^{11,12,30-33} The sensitivity increased the further out in time serology was assessed, consistent with the biologic kinetic profile of a mounting adaptive immune response. Direct comparison with other studies is difficult due to differences in experimental design, including considerations for positivity (ie, days from symptom onset or RT-PCR positivity) and patient population characteristics that can influence when and whether someone might seek medical care. Assay specificity is more comparable but still limited by the number of samples (ie, statistical power) and selection of potentially cross-reactive samples.

While the Roche total antibody assay was often the first positive assay, the time before other assays crossed threshold was modest at approximately 1 day. This would not likely be clinically significant when testing patients at time points far from symptom onset to evaluate for history of COVID-19. There could be potential benefit of early serologic detection in the periconvalescent state in RT-PCR–negative patients, although this would be a limited use, and the utility as a primary diagnostic test seems ill-advised.³² As the Roche assay was the only nucleocapsid-based assay in this study, it is unclear whether detection differences were technical or biological. Similarly, it is possible that earlier detection with total antibody assays is due to assay principles. Both total antibody assays performed similar to each other despite use of different antigens, without significant difference in any assay characteristics at any time point.

Although we observed increasing sensitivity with samples collected at later time points, our study did not

Table 6
Estimated Positive Predictive Values at Varying Frequencies of COVID-19 Prevalence^a

Prevalence	DiaSorin PPV (95% CI), %		EUROIMMUN PPV (95% CI), Indeterminate = Positive, %		EUROIMMUN PPV (95% CI), Indeterminate = Negative, %		Roche PPV (95% CI), %		Siemens PPV (95% CI), %	
	All Samples	≥14 Days After Symptom Onset	All Samples	≥14 Days After Symptom Onset	All Samples	≥14 Days After Symptom Onset	All Samples	≥14 Days After Symptom Onset	All Samples	≥14 Days After Symptom Onset
1	17.9 (10.1-40.6)	20.5 (11.8-44.6)	8.6 (5.8-14.1)	10.4 (6.8-16.2)	16.7 (9.8-36.2)	19.3 (11.3-37.9)	100 (36.2-100)	100 (39.9-100)	100 (34.6-100)	100 (38.2-100)
5	53.2 (37.0-78.0)	57.4 (41.1-80.7)	32.9 (24.2-46.1)	36.8 (27.5-50.1)	51.1 (36.2-74.7)	55.5 (39.8-76.1)	100 (74.7-100)	100 (77.6-100)	100 (73.4-100)	100 (76.3-100)
10	70.6 (55.4-88.2)	74.0 (59.6-89.8)	50.9 (40.2-64.3)	55.1 (44.5-68.0)	68.8 (54.5-86.2)	72.4 (58.3-87.0)	100 (86.2-100)	100 (88.0-100)	100 (85.3-100)	100 (87.2-100)
20	84.4 (73.6-94.4)	86.5 (76.8-95.2)	70.0 (60.2-80.2)	73.4 (64.3-82.7)	83.3 (73.0-93.4)	85.5 (75.9-93.8)	100 (93.4-100)	100 (94.3-100)	100 (92.9-100)	100 (93.9-100)

CI, confidence interval; PPV, positive predictive value.

^aPPVs are shown as the point estimate (%) and 95% confidence interval based on bootstrapped likelihood ratios.

include samples beyond 2 months after symptom onset. It is unknown at what point anti-SARS-CoV-2 ceases to be reliably detected, as concentrations of anti-SARS-CoV-2 antibodies may decrease over time in convalescent patients at 2 to 3 months.^{34,35} We observed a small portion of patients at 1 to 2 months with relatively low but positive signals that are not explained by known comorbidities or medications (eg, immunosuppression) and could represent a waning response, although this seems like an early decline. In contrast, of five patients on immunosuppressive therapies, four generated moderate to robust antibody responses, and only one patient displayed a relatively weak response by all methods (although still positive on all assays except DiaSorin) approximately 1 month after symptom onset. It is worth noting that we did not test samples from any patients on strong immunomodulating agents (eg, ibrutinib, rituximab).

The high specificity of the Roche and Siemens assays may be the most important finding. Specificities are similar to those observed by other studies,^{2,12-14,19-22} and to our knowledge, our study included one of the larger number and variety of cross-reactivity samples. The specificities would remain high even if the two presumed false-negative RT-PCR patient samples were included in the analysis. There has been little literature regarding the Siemens assay, although recent reports indicate excellent specificity,^{13,20-22} which our findings further support with a wide array of samples. The specificities of the DiaSorin and EUROIMMUN assays are similar to, or modestly lower, than other evaluations,^{2-4,8-11,13-17,19,21,36} although some of the studies reporting better performance had very small number of samples or nonchallenging cross-reactivity panels. Differences may also be due to biases in sample selection such as inclusion of multiple low-positive samples in the current study and choosing time points at which detection rapidly increases, but they could also represent the performance in real-world clinical laboratory practice rather than well-controlled validation studies. In addition, although an immunoglobulin A (IgA) assay is available from EUROIMMUN, we have not included the evaluation of the test in the study due to the poor specificity ($\leq 90\%$) described by the manufacturer and other published reports.^{14-16,37,38}

Differences in performance of the total antibody assays may be explained by a variety of reasons. First, the total assays can have increased signal (ie, sensitivity) not only due to the number of antibody isotypes present but also due to the number of antigen-binding sites by including IgM, which could bind multiple reporter antigens. Potential advantages of increased signal may be relevant only early after infection while IgM (and IgA) are prevalent. In addition, by not relying on anti-IgG or anti-IgM

antibodies, the total assays are not expected to be susceptible to interferences such as heterophile antibodies or rheumatoid factor. The trade-off is that such assays do not provide isotype distinction, yet these assays appear robust even at 1 to 2 months when the antibody composition would be expected to be IgG predominant.

Initial concerns about potential cross-reactivity of seasonal coronaviruses with SARS-CoV-2 antibody assays resulted in FDA recommendations for inclusion of a disclaimer. However, pervasive issues across multiple assays would be likely if there were significant antigenic overlap due to high rates of a history of seasonal coronavirus infections in the general population. However, in this study, the EUROIMMUN assay had a positive result from one of two NL63 samples and was the only assay to have a positive result from the 13 seasonal coronavirus samples tested (Table 4). Similarly, there have been other reports with low, sporadic potential cross-reactivity of the EUROIMMUN assay to seasonal coronaviruses, including OC43 and NL63.^{16,31,38}

Despite the similar performance in detection of prior SARS-CoV-2 infection in this study, the importance of antinucleocapsid antibodies compared with antispike antibodies is not known. Although questions of immunity and neutralizing capacity remain open, antinucleocapsid antibodies may not mechanistically be linked to subsequent antiviral activity.³⁹ Since both antinucleocapsid and antispike responses appear correlated in most individuals, it is also possible that both antigens may correlate with any antiviral functional studies. There have been some reports of neutralizing antibodies correlating to values of DiaSorin, EUROIMMUN, and other spike-based serology assays.^{9,31,40,41} However, given the possibility of declining antibodies at convalescence,^{34,35} the composition of neutralizing and binding antibodies may be even more important for correlating outcomes when well into convalescence of many months. Potential benefits exist for detecting antibodies to nucleocapsid and spike antigens independently, where the distinction between vaccination vs prior infection may be possible and desirable. This is particularly relevant as the most likely possibilities for a future vaccine appear to be based on the spike protein, particularly the RBD portion of the spike S1 subunit.⁴²⁻⁴⁴ In addition, two different antigens would likely have the most fidelity to true orthogonal testing, as recommended by the CDC, if a secondary method of confirmation is desired.⁴⁵

There are drawbacks to each assay, and determining the right fit and acceptable performance will be an important decision for each clinical laboratory. The Siemens assay, as of this writing, requires bracketing washes or a limited test menu of other assays not using the common probes on the immunoanalyzer. The

Roche assay did not have QC material available and only claimed 3-day reagent stability at release, creating additional laboratory work to create QC material and potential for reagent waste depending on test volumes. Regarding concerns of throughput, there are other analyzers available (eg, cobas e801) with higher throughput compared to the e411. The DiaSorin assay on a Liaison XL is a high-throughput method but in our experience is more expensive than the Roche and Siemens assays. The EUROIMMUN assay may be amenable to a laboratory with ELISA capabilities but where random-access methods are not available, and it can be performed with less sample volume. However, the EUROIMMUN assay is complicated by the “Indeterminate” result, which clouds the interpretation of the assay. A clinical laboratory needs to be aware of which limitations are acceptable for its practice.

Potential testing algorithms advocate the use of multiple assays either concurrently or serially,⁴⁵ especially for assays with lower specificities. While the PPV is improved if both assays are positive, discordant results may require additional testing. This begs the question of how many assays the clinical laboratories should evaluate and whether the laboratories absorb the costs associated with the additional testing. Our data reveal that some samples may yield falsely positive results among similar test methods, which should be considered if choosing confirmatory testing. Conversely, a truly positive result from an assay with superior sensitivity could be negative by a secondary testing method. Testing at later time points by either method (eg, 1-2 weeks) would likely yield a true-positive result by both methods; however, that outcome would not be different from testing serially by a single platform in a suspected patient. Furthermore, although these are qualitative assays, an increasing index value over time provides confidence of increasing antibody concentrations, whereas some false-positive results may remain constant over time.

In summary, our study has provided a head-to-head evaluation of four commercial options for serology testing for anti-SARS-CoV-2 with a common set of samples. Both Roche and Siemens assays perform well and similarly to their stated claims, although there are still implementation issues with both platforms. The best choice of a serology assay is likely an individualized decision for each clinical laboratory, depending on practical issues and plans to manage potential false-positive or false-negative test results.

Corresponding author: David M. Manthei, MD, PhD;
dmanthei@med.umich.edu.

References

- US Food and Drug Administration. FAQs on testing for SARS-CoV-2. www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/faqs-testing-sars-cov-2#nolonger-ivd. Accessed August 18, 2020.
- US Food and Drug Administration. EUA authorized serology test performance. www.fda.gov/medical-devices/emergency-situations-medical-devices/eua-authorized-serology-test-performance. Accessed July 24, 2020.
- US Food and Drug Administration. Independent evaluations of COVID-19 serological tests. open.fda.gov/apis/device/covid19serology/. Accessed July 24, 2020.
- Theel ES, Harring J, Hilgart H, et al. Performance characteristics of four high-throughput immunoassays for detection of IgG antibodies against SARS-CoV-2. *J Clin Microbiol*. 2020;58:e01243-20.
- Lassauniere R, Frische A, Harboe ZB, et al. Evaluation of nine commercial SARS-CoV-2 immunoassays. [published online April 9, 2020]. *medRxiv*.
- Whitman JD, Hiatt J, Mowery CT, et al. Test performance evaluation of SARS-CoV-2 serological assays. [published online May 17, 2020]. *medRxiv*.
- 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:e00941-20.
- Kruttgen A, Cornelissen CG, Dreher M, et al. Comparison of four new commercial serologic assays for determination of SARS-CoV-2 IgG. *J Clin Virol*. 2020;128:104394.
- Bonellia F, Sarasinib A, Zierolda C, et al. Clinical and analytical performance of an automated serological test that identifies S1/S2 neutralizing IgG in COVID-19 patients semiquantitatively. *J Clin Microbiol*. 2020;58:e01224-20.
- Charlton CL, Kanji JN, Johal K, et al. Evaluation of six commercial mid to high volume antibody and six point of care lateral flow assays for detection of SARS-CoV-2 antibodies. *J Clin Microbiol*. 2020;58:e01361-20.
- Tang MS, Hock KG, Logsdon NM, et al. Clinical performance of two SARS-CoV-2 serologic assays. *Clin Chem*. 2020;66:1055-1062.
- Favresse J, Euchet C, Elsen M, et al. Clinical performance of the Elecsys electrochemiluminescent immunoassay for the detection of SARS-CoV-2 total antibodies. *Clin Chem*. 2020;66:1104-1106.
- Public Health England. Evaluation of sensitivity and specificity of four commercially available SARS-CoV-2 antibody immunoassays. assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/898437/Evaluation_of_sensitivity_and_specificity_of_4_commercially_available_SARS-CoV-2_antibody_immunoassays.pdf. Accessed July 24, 2020.
- Herroelen PH, Martens GA, De Smet D, et al. Humoral immune response to SARS-CoV-2: comparative clinical performance of seven commercial serology tests [published online August 18, 2020]. *Am J Clin Pathol*.
- Tehrani ZR, Saadat S, Saleh E, et al. Specificity and performance of nucleocapsid and spike-based SARS-CoV-2 serologic assays [published online August 7, 2020]. *MedRxiv*.
- Jääskeläinen AJ, Kuivanen S, Kekäläinen E, et al. Performance of six SARS-CoV-2 immunoassays in comparison with microneutralisation. *J Clin Virol*. 2020;129:104512.
- Merrill AE, Jackson JB, Ehlers A, et al. Head-to-head comparison of two SARS-CoV-2 serology assays [published online July 27, 2020]. *J Appl Lab Med*.
- Suhandynata RT, Hoffman MA, Kelner MJ, et al. Multi-platform comparison of SARS-CoV-2 serology assays for the detection of COVID-19 [published online August 7, 2020]. *J Appl Lab Med*.
- Harb R, Remaley AT, Sacks DB. Evaluation of three commercial automated assays for the detection of anti-SARS-CoV-2 antibodies [published online August 6, 2020]. *Clin Chem*.
- Horber S, Soldo J, Relker L, et al. Evaluation of three fully-automated SARS-CoV-2 antibody assays [published online August 4, 2020]. *Clin Chem Lab Med*.
- Pfluger LS, Bannasch JH, Brehm TT, et al. Clinical evaluation of five different automated SARS-CoV-2 serology assays in a cohort of hospitalized COVID-19 patients. *J Clin Virol*. 2020;130:104549.
- Seegmiller JC, Kokaisel EL, Story SJ, et al. Method comparison of SARS-CoV-2 serology assays involving three commercially available platforms and a novel in-house developed enzyme-linked immunosorbent assay [published online August 11, 2020]. *Clin Biochem*.
- Theel ES, Slev P, Wheeler S, et al. The role of antibody testing for SARS-CoV-2: is there one? *J Clin Microbiol*. 2020;58:e00797-20.
- US Food and Drug Administration. Serology template for laboratories. www.fda.gov/media/137697/download. Accessed August 18, 2020.
- EUROIMMUN. Anti-SARS-CoV-2 ELISA (IgG) [insert sheet REF EI_2606G_A_US_C02.docx, 2020-05-04]. 2020.
- DiaSorin. LIAISON SARS-CoV-2 S1/S2 IgG [insert sheet REF 311460, 2020-04]. 2020.
- Roche Diagnostics. Elecsys Anti-SARS-CoV-2 [insert sheet REF 09203095190, 2020-05, V 2.0]. 2020.
- Siemens Healthcare Diagnostics. Siemens SARS-CoV-2 Total [insert sheet REF 11206922, 2020-05, Rev. 01]. 2020.
- Marill KA, Chang Y, Wong KF, et al. Estimating negative likelihood ratio confidence when test sensitivity is 100%: a bootstrapping approach. *Stat Methods Med Res*. 2017;26:1936-1948.
- Long QX, Liu BZ, Deng HJ, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med*. 2020;26:845-848.
- Okba NMA, Muller MA, Li W, et al. Severe acute respiratory syndrome coronavirus 2-specific antibody responses in coronavirus disease 2019 patients. *Emerg Infect Dis*. 2020;26:1478-1488.
- Deeks JJ, Dinnes J, Takwoingi Y, et al; Cochrane COVID-19 Diagnostic Test Accuracy Group. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev*. 2020;6:CD013652.
- 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;20:565-574.
- Long QX, Tang XJ, Shi QL, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med*. 2020;26:1200-1204.
- Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild Covid-19. *N Engl J Med*. 2020;383:1085-1087.
- Weidner L, Gansdorfer S, Unterwiesing S, et al. Quantification of SARS-CoV-2 antibodies with eight commercially available immunoassays. *J Clin Virol*. 2020;129:104540.

37. EUROIMMUN. Anti-SARS-CoV-2 ELISA (IgA) [insert sheet REF EI_2606A_A_UK_C01.docx, 2020-03-23]. 2020.
38. Beavis KG, Matushek SM, Abeleda APF, et al. Evaluation of the EUROIMMUN anti-SARS-CoV-2 ELISA assay for detection of IgA and IgG antibodies. *J Clin Virol*. 2020;129:104468.
39. Wu F, Wang A, Liu M, et al. Neutralizing antibody response to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications [published online April 20, 2020]. *medRxiv*.
40. Amanat F, Stadlbauer D, Strohmeier S, et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat Med*. 2020;26:1033-1036.
41. Jaaskelainen AJ, Kuivanen S, Kekalainen E, et al. Performance of six SARS-CoV-2 immunoassays in comparison with microneutralization. *J Clin Virol*. 2020;129:104512.
42. Lurie N, Saville M, Hatchett R, et al. Developing Covid-19 vaccines at pandemic speed. *N Engl J Med*. 2020;382:1969-1973.
43. Thanh Le T, Andreadakis Z, Kumar A, et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov*. 2020;19:305-306.
44. Amanat F, Krammer F. SARS-CoV-2 vaccines: status report. *Immunity*. 2020;52:583-589.
45. Centers for Disease Control and Prevention. Interim guidelines for COVID-19 antibody testing, updated May 23, 2020. www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html. Accessed July 24, 2020.