

Evelyne Huyghe, Hilde Jansens, Veerle Matheussen, Ilse Hoffbauer, Herman Goossens and Bart Peeters*

Performance of three automated SARS-CoV-2 antibody assays and relevance of orthogonal testing algorithms

<https://doi.org/10.1515/cclm-2020-1378>

Received September 11, 2020; accepted November 3, 2020;

published online November 19, 2020

Abstract

Objectives: Development and implementation of SARS-CoV-2 serologic assays gained momentum. Laboratories keep on investigating the performance of these assays. In this study, we compared three fully automated SARS-CoV-2 antibody assays.

Methods: A total of 186 samples from 84 PCR-positive COVID-19 patients and 120 control samples taken before the SARS-CoV-2 pandemic were analyzed using commercial serologic assays from Roche, Siemens and DiaSorin. Time after the positive COVID-19 PCR result and onset of symptoms was retrieved from the medical record. An extended golden standard, using the result of all three assays was defined, judging if antibodies are present or absent in a sample. Diagnostic and screening sensitivity/specificity and positive/negative predictive value were calculated.

Results: Diagnostic sensitivity (ability to detect a COVID-19 positive patient) ≥ 14 days after positive PCR testing was 96.7% (95% CI 88.5–99.6%) for DiaSorin, 93.3% (95% CI 83.8–98.2%) for Roche and 100% (95% CI 94.0–100%) for Siemens. Lower diagnostic sensitivities were observed < 14 days after onset of symptoms for all three assay. Diagnostic specificity (ability to detect a COVID-19 negative patient) was 95.0% (95% CI 89.4–98.1%) for DiaSorin, 99.2% (95% CI 95.4–99.9%) for Roche and 100% (95% CI

97.0–100%) for Siemens. The sensitivity/specificity for detecting antibodies (ability of detecting absence (specificity) or presence (sensitivity) of COVID-19 antibodies) was 92.4% (95% CI 86.4–96.3%)/94.9% (95% CI 90.5–97.6%) for DiaSorin, 97.7% (95% CI 93.5–99.5%)/97.1% (95% CI 93.5–99.1%) for Roche and 98.5% (95% CI 94.6–99.8)/97.1% (95% CI 93.5–99.1%) for Siemens.

Conclusions: This study revealed acceptable performance for all three assays. An orthogonal testing algorithm using the Siemens and Roche assay achieved the highest positive predictive values for antibody detection in low seroprevalence settings.

Keywords: COVID-19; SARS-CoV-2; serology.

Introduction

The COVID-19 (coronavirus disease 2019) pandemic caused by SARS-CoV-2019 (severe acute respiratory syndrome coronavirus 2), has forced many governments to apply draconic measures [1–4]. As these measures like community quarantine and social distancing are impossible to hold for a longer period, countries are chasing a strategy to tapering off the measures without encouraging the viral spread. Here, serology could play a role in different ways [5, 6]. It can be tested when polymerase chain reaction (PCR) techniques are unavailable or additionally to PCR. It can give epidemiological information as it can help mapping viral spread in a given population and identifying subjects who became already PCR-negative [7]. Furthermore, serology helps in selecting subjects who are suitable as donor for convalescent plasma therapy [6, 8]. Later on, serology testing will be essential in evaluating immune response to vaccination.

Four structural proteins of the coronaviruses could be a possible target of immunocompetent antibodies, namely the S (Spike), N (Nucleocapsid), E (Envelope), and M (Membrane) protein [9]. Among these proteins, the S-protein and N-protein are the most known targets for serology testing in coronaviruses [10]. In fact, antibodies against the RBD (Receptor Binding Domain), located on the

*Corresponding author: Bart Peeters, Department of Laboratory Medicine, Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem, Belgium, Phone: +32 3 821 39 21, Fax: +32 3 825 11 48, E-mail: bart.peeters@uza.be

Evelyne Huyghe, Hilde Jansens, Veerle Matheussen, Ilse Hoffbauer and Herman Goossens, Department of Laboratory Medicine, University Hospital Antwerpen, Antwerpen, Belgium, E-mail: evelyne.huyghe@uza.be (E. Huyghe), hilde.jansens@uza.be (H. Jansens), veerle.matheussen@uza.be (V. Matheussen), ilse.hoffbauer@uza.be (I. Hoffbauer), herman.goossens@uza.be (H. Goossens)

S1 domain of the Spike protein [11], have shown to neutralize the SARS-CoV-2 virus in mouse models [12]. These findings give the testing for RBD immunology great opportunities as it could give information about immunity against SARS-CoV-2. The purpose of this study was a thorough analytical validation and comparison of the clinical and screening performances of three SARS-CoV-2 antibody assays on fully automated platforms of Siemens, DiaSorin and Roche. By using three validated platforms, the design of an extended golden standard was possible for judging if antibodies are present or absent and predictive value of a combined testing strategy (orthogonal testing set-up) was evaluated as a function of SARS-CoV-2 seroprevalence.

Materials and methods

Study design

This retrospective validation study was performed in the laboratory of clinical biology of the University Hospital Antwerpen from 01/06/2020 to 19/06/2020. All sera (n=306) originated from diagnostic laboratory requests and were saved from removal after five days. Recovered sera were stored in the laboratory at -20°C . This study has approval of the Ethical Committee of the UZ Antwerpen (Reference Number: 20/27/356).

Population and sample collection

Samples used for evaluation of sensitivity (n=186) were collected during the COVID-19 outbreak (March–April 2020) from 84 symptomatic patients (29 women; 55 men) with a mean age of 64 years (range 19–90 years) with a positive COVID-19 SARS-CoV-2 PCR result on a respiratory sample. Samples (n=120) used in the experiments for specificity were collected before September 2019. All blood samples were collected in serum tubes (Becton Dickinson) and centrifuged at 3,000 rpm for 10 min. Sera were used for diagnostic laboratory requests and stored at $2-8^{\circ}\text{C}$ for five days. Selected sera were saved from removal and an aliquot was stored at -20°C awaiting serological testing. Frozen samples were thawed at room temperature and analyzed the same day on all three platforms. Aliquot volume was sufficient in order to analyze the same sample on all three platforms.

Analytical procedures

Three automated SARS-CoV-2 antibody assays were evaluated and compared in accordance with the manufacturer's instructions [13–15]. The Atellica IM SARS-CoV-2 assay (Siemens) and Elecsys Anti-SARS-CoV-2 assay (Roche) are total (IgG and IgM), qualitative assays based on detection of antibodies directed against the S1 receptor binding domain and the nucleocapsid antigen, respectively. Results are expressed as reactive (≥ 1.0) or non-reactive (< 1.0) index values. The LIAISON SARS-CoV-2 IgG assay (DiaSorin), performed on the Liaison XL analyzer is a quantitative IgG-assay based on detection of antibodies directed against two subunits of the virus spike protein (S1 and S2). Results are expressed as positive (≥ 15.0 AU/mL), doubtful

(12.0–15.0 AU/mL) or negative (< 12.0 AU/mL). In this study samples with a doubtful result were considered as negative.

The SARS-CoV-2 PCR was performed on nasopharyngeal swabs or lower respiratory tract specimens. Multiple validated platforms were used next to each other: the Xpert Xpress SARS-CoV-2 molecular assay (Cepheid), the BD MAX BioGX SARS-CoV-2 molecular assay (Becton Dickinson) and an in-house assay based on a publication of Corman et al. [16].

Procedures for assessment of analytical and clinical performance

Trueness and precision were evaluated according to the Clinical and Laboratory Standards Institute (CLSI) EP15-A3 document [17] using the reported performance by the manufacturer as a criterion (Table 2). Two levels of quality controls (QC) provided by the manufacturer were used to evaluate trueness by comparing the average value (± 1.96 SD) obtained on 25 replicates to the target values. Repeatability and reproducibility were evaluated using QC provided by the manufacturer and a positive/negative serum pool. The positive serum pool consisted of COVID-19 patient samples collected ≥ 14 days after a positive SARS-CoV-2 PCR result. The negative serum pool consisted of patient samples collected before September 2019. After pooling and thoroughly mixing, both serum pools were analyzed on the three platforms, confirming desired values (negative and low-positive), and divided in 1 mL aliquots which were stored at -20°C awaiting analysis. Each level was run five times for five consecutive days.

Signal/noise experiments were performed according to CLSI EP17 [18] where limit of blank (LOB) was determined by a 30-fold analysis of a diluent provided by the manufacturer with $\text{LOB} = \text{mean}_{\text{blank}} + [1.645 \times \text{SD}_{\text{blank}}]$. Limit of detection (LOD) was determined by a 30-fold analysis of the negative serum pool with $\text{LOD} = \text{LOB} + [1.645 \times \text{SD}_{\text{negative pool}}]$. Limit of quantification was calculated according to equations of the COFRAC [19] with $\text{LOQ} = \text{mean}_{\text{blank}} + [30 \times \text{SD}_{\text{blank}}]$. LOB and LOD were acceptable if they were smaller than 0.5 times the cut-off for positivity and the LOQ should be smaller than the cut-off for positivity.

Linearity was assessed according to CLSI EP06 [20]. A high positive sample was run in triplicate and five dilutions were made in a two-fold dilution series using a manufacturer diluent. Each dilution was run in triplicate.

Carry-over was evaluated using the method (three replicates of a strong positive sample, followed by three replicates of a negative sample) and formula of Broughton [21]. Carry-over $< 1\%$ was considered as insignificant.

For the assessment of diagnostic sensitivity sera collected at 0–5 days (n=73), 6–13 days (n=53) and ≥ 14 days (n=60) since the diagnostic positive SARS-CoV-2 PCR were analyzed. For the same sample cohort the date of onset of COVID-19 symptoms was retrieved from patient's records. These sera were collected at 0–5 days (n=20), 6–13 days (n=69) and ≥ 14 days (n=97) after onset of COVID-19 symptoms.

For the assessment of diagnostic specificity, four different challenging panels consisting of sera collected before September 2019 were compiled (n=120 in total). A first panel of 50 randomly selected sera kept in an historical biobank (no further information). A second panel of 15 sera from patients with a non-COVID-19 strain infection (HCoV 229E [n=4]; NL63 [n=4]; OC43 [n=4]; HKU1 or unidentified [n=3]). A third panel of 10 rheumatoid factor (RF) positive sera with RF

Table 1: Selected serum samples for assessing cross-reactivity of three SARS-CoV-2 serologic assays.

Number of sam- ples (n=45)	Positive serology for:					
	CMV IgM	EBV VCA IgG	EBV VCA IgM	VZV IgG	HBsAg	HBsAb
9		X				
8						x
7					x	
4				x		
4		X	X			
2				x		x
2		X				x
2		X	X			x
2	X	X				
1		X		x		x
1		X	X	x		
1	X	X		x		
1	X	X				x
1	X					x

CMV, cytomegalovirus; EBV, Epstein-Barr virus; VCA, viral capsid antigens; VZV, varicella-zoster virus; HBsAg, hepatitis B surface antigen; HBsAb, hepatitis B surface antibody.

concentrations ranging from 34 to 566 IU/mL were selected, seven of these samples had a positive ANA-screening with titers ranging from 1:80 to 1:5,120. A last panel of sera (n=45) with a positive viral serologic result were included to assess cross-reactivity (Table 1).

For assessing capability of detecting the absence (screening specificity) or presence (screening sensitivity) of COVID-19 antibodies, an extended golden standard was defined and applied on all analyzed samples (n=306). Antibodies were considered to be present if at least two out of three assays gave a positive result (n=131). Antibodies are considered to be absent if at least two out of three assays gave a negative result (n=175). Sensitivity and specificity of each assay was calculated compared to the result of the extended golden standard.

Statistical analysis

Statistics were performed in MedCalc Statistical Software version 17.5.5 (MedCalc Software, Ostend, Belgium). ‘Diagnostic sensitivity’ reflects the proportion of positive serological results from patients who

were diagnosed with COVID-19 by a positive respiratory SARS-CoV-2 PCR result. Diagnostic sensitivity is assessed as a function of time (in days) since the positive SARS-CoV-2 PCR result. ‘Diagnostic specificity’ was defined as the total proportion of negative serological results from the sera selected in the four challenging panels. ‘Detection sensitivity/specificity’ is calculated from the proportion of positive/negative serological results in which the extended golden standard indicated presence/absence of antibodies, respectively.

Positive and negative predictive value (PPV, NPV) was calculated and a graph with (combined) PPV and NPV as a function of prevalence was designed.

Results

Trueness and precision

Trueness of the three assays could only be assessed using the reference range of the insert provided with the QC from the manufacturer. None of the assays had QC results outside this reference range and so the mean ± 1.96 SD of each method was well inside their reference range (Table 2). Repeatability and reproducibility results using the manufacturer QC material were in line with CV%’s found in the LIAISON SARS-CoV-2 IgG and Atellica IM SARS-CoV-2 reagent inserts. CV%’s were much higher using a positive and negative serum pool, only the Roche assay achieved similar CV%’s in the serum pool experiment. Although the CV%’s in pooled serum were higher, they were still acceptable for routine use of the assays taking into account their mean (Table 2).

Limit of blank, detection and quantification, linearity and carry-over

LOB, LOD and LOQ were acceptable for all assays according to our arbitrary criteria (Table 2). Only the LIAISON SARS-CoV-2 IgG insert reports a LOD of 3.8 AU/mL. The LOD found in this study is well below this value (1.02 AU/mL). No statements were made by the manufacturers

$$PPV = \frac{\text{sensitivity} \times \text{prevalence}}{(\text{sensitivity} \times \text{prevalence}) + [(1 - \text{specificity}) \times (1 - \text{prevalence})]}$$

$$NPV = \frac{\text{specificity} \times (1 - \text{prevalence})}{[(1 - \text{sensitivity}) \times \text{prevalence}] + [\text{specificity} \times (1 - \text{prevalence})]}$$

$$\text{CombinedPPV}_{\text{test a+test b}} = \frac{\text{sensitivity (a)} \times \text{sensitivity (b)} \times \text{prevalence}}{(\text{sensitivity (a)} \times \text{sensitivity (b)} \times \text{prevalence}) + [(1 - \text{specificity (a)}) \times (1 - \text{specificity (b)}) \times (1 - \text{prevalence})]}$$

$$\text{CombinedNPV}_{\text{test a+test b}} = \frac{(1 - \text{specificity (a)}) \times \text{specificity (b)} \times (1 - \text{prevalence})}{[\text{sensitivity (a)}(1 - \text{sensitivity (b)}) \times \text{prevalence}] + [(1 - \text{specificity (a)}) \times \text{specificity (b)} \times (1 - \text{prevalence})]}$$

Table 2: Results of analytical validation of three SARS-CoV-2 serologic assays.

Experiment	LIAISON® SARS-CoV-2 IgG		Elecsys anti-SARS-CoV-2		Atellica SARS-CoV-2 total	
	Acceptance criteria according to manufacturer	Results	Acceptance criteria according to manufacturer	Results	Acceptance criteria according to manufacturer	Results
Trueness						
Low QC level	<6 AU/mL	<3.8 AU/mL	<0.8 COI	0.098 ± 0.001 COI	<0.8 COI	<0.05 COI
High QC level	15–45 AU/mL	33.3 ± 1.57 AU/mL	1.76–4.64 COI	2.8 ± 0.17 COI	1.0–5.0 COI	2.8 ± 0.53 COI
Repeatability						
Low QC level	<3.5%	n/a (mean < 3.8 AU/mL)	Not reported	2.57%	<4.0%	n/a (mean < 0.05 COI)
High QC level	<3.3%	1.47%		0.99%	<1.8%	0.00%
Negative serum pool	<3.5%	n/a (mean < 3.8 AU/mL)		1.44% (mean 0.09 COI)	<4.0%	32.22% (mean 0.14 COI)
Positive serum pool	<3.3%	3.15% (mean 46.9 AU/mL)		1.44% (mean 3.72 COI)	<1.8%	2.16% (mean 8.27 COI)
Reproducibility						
Low QC level	<7.6%	n/a (mean < 3.8 AU/mL)	Not reported	2.41%	<5.7%	n/a (mean < 0.05 COI)
High QC level	<3.7%	1.54%		3.20%	<4.4%	9.52%
Negative serum pool	<7.6%	n/a (mean < 3.8 AU/mL)		1.41% (mean 0.09 COI)	<5.7%	34.51% (mean 0.14 COI)
Positive serum pool	<3.7%	18.23% (mean 46.9 AU/mL)		2.86% (mean 3.72 COI)	<4.4%	10.15% (mean 8.27 COI)
Signal/noise						
LOB	Not reported	0.32 AU/mL	Not reported	0.23 COI	Not reported	0.33 COI
LOD	3.8 AU/mL	1.02 AU/mL		0.23 COI		0.38 COI
LOQ	Not reported	4.57 AU/mL		0.25 COI		0.63 COI
Linearity assessment						
Type of regression	Linear	Polynomial regression	Not reported	Polynomial regression	Not reported	Polynomial regression
Carry-over	<1%	0.2%	<1%	0.002%	<1%	0.1%

QC, quality control; AU, arbitrary unit; COI, cut-off index; LOB, limit of blank; LOD, limit of detection; LOQ, limit of quantification; n/a, not applicable.

regarding LOB, LOD and LOQ for the other assays. All the assays had a polynomial regression (type $ax^2 + bx + c$) and only the LIAISON SARS-CoV-2 IgG insert stated that there was a linear regression after dilution of a positive sample. This was not confirmed in this study. None of the manufacturers suggests diluting samples with a result above the upper measuring range. Using the Broughton method and formula, no carry-over was observed in any assay (Table 2).

Clinical performance in a diagnostic or screening setting

Overall diagnostic sensitivity and sensitivity as a function of days after a positive SARS-CoV-2 PCR result stated in the reagent inserts were not statistically different from diagnostic sensitivities found in this study, however sensitivities found in this study seem lower for all three assays (Table 3). Diagnostic sensitivity of all assays increased significantly in time after a positive SARS-CoV-2 PCR result or after onset of COVID-19 symptoms. For all assays, diagnostic sensitivity is low (range 38.4–48.0%) and very low (range 15.0–20.0%) in the first five days after a positive SARS-CoV-2 PCR result or onset of COVID-19 symptoms, respectively. Diagnostic sensitivities of each assay were significantly higher ≥ 14 days after a positive SARS-CoV-2 PCR result (range 93.3–100%) or after onset of COVID-19 symptoms (range 85.6–95.9%). Although statistically not significant, the LIAISON SARS-CoV-2 IgG kit has lower diagnostic sensitivities compared to the other two platforms across all time periods. Overall diagnostic specificity was excellent for the Atellica IM SARS-CoV-2 (100%) and Elecsys Anti-SARS-CoV-2 (99.2%) assay. None of the 120 tested sera were positive using the Atellica IM assay and one sample strongly positive for VZV IgG (2,251 mIU/mL with cut-off 135 mIU/mL) cross-reacted with the Elecsys Anti-SARS-CoV-2 assay. Diagnostic specificity of the LIAISON SARS-CoV-2 IgG (95.0%) seems lower (not statistically significant) as three RF positive samples reacted as positive. Two of these samples had a modest RF concentration (34 and 60 IU/mL with cut-off 10 IU/mL) with a high titer (1:1,280) positive ANA-screen. One sample had a higher RF concentration (277 IU/mL) with a modest titer (1:60) positive ANA-screen. In addition, there were three false positive results in samples positive for Varicella-zoster IgG ($n=1$, 2,251 mIU/mL with cut-off 135 mIU/mL), EBV VCA IgG ($n=1$, 294 IU/mL with cut-off 20 IU/mL) and HBsAb ($n=1$, $>1,000$ IU/mL with cut-off 10 IU/mL). Adjusted cut-off values, optimizing diagnostic sensitivity and specificity were calculated using ROC-curve analysis being >0.14 COI for Elecsys with diagnostic sensitivity/

specificity being 86.6% (95% CI 80.8–91.1)/96.7% (95% CI 91.8–89.1), >0.50 COI for Atellica with diagnostic sensitivity/specificity being 78.5% (95% CI 71.9–84.2)/100% (95% CI 97.2–100) and >9.75 AU/mL for Liaison with diagnostic sensitivity/specificity being 72.6% (95% CI 65.6–78.9)/91.9% (95% CI 85.7–96.1). As there is no statistical significant higher diagnostic sensitivity/specificity using the adjusted cut-offs (except for diagnostic sensitivity of Elecsys), manufacturer's cut-offs were used in this study.

Detection capability (presence/absence) of COVID-19 antibodies was good for all methods with the highest sensitivity/specificity for the Atellica IM SARS-CoV-2 (98.5%/97.1%) and Elecsys Anti-SARS-CoV-2 (97.7%/97.1%) assay. Although statistically not significant, detection capability of the LIAISON SARS-CoV-2 IgG assay seemed lower (94.2%/94.9%). Using these performance characteristics, PPV and NPV were calculated and plotted as a function of seroprevalence (Figures 1 and 2). The combined use of the Roche and Siemens assay achieved the highest PPV in low seroprevalence settings (e.g. PPV $> 90\%$ if seroprevalence $\geq 1\%$). In order to optimize this testing strategy cut-off values were calculated generating a sensitivity/specificity statistically equal to 100%. If the Roche assay is used as a first test, a COI ≤ 0.849 has a sensitivity of 99.2% (95% CI 95.8–100%) and a COI > 3.29 has a specificity of 99.4% (95% CI 96.9–100%), samples with a COI between 0.849 and 3.29 should be confirmed by the Siemens assay. If the Siemens assay is used as a first test, a COI ≤ 0.66 has a sensitivity of 99.2% (95% CI 95.8–100%) and a COI > 5.33 has a specificity of 100% (95% CI 97.9–100%), samples with a COI between 0.66 and 5.33 should be confirmed by the Roche assay. If there are discordant interpretations by the Siemens and Roche assay, a new sample should be tested after one week or a third serologic assay can be applied to that sample in order to judge presence or absence of antibodies.

The orthogonal testing algorithm using Roche as a first test and Siemens as a confirmatory test using the optimized cut-off values was used in our hospital for a voluntary antibody screening of health workers as part of a local risk management strategy. Discordant results were resolved by testing with a third antibody test (DiaSorin). Of 1,151 tested samples (query date for serologic results in laboratory information system was 30/06/2020), 69 had a Roche COI ≥ 1.0 of which 7 had a COI between 1.01 and 3.29, 1,082 had a Roche COI < 0.1 of which 2 had a COI between 0.85 and 0.99. These nine samples (0.8% of total samples) were retested with the Siemens assay. The two negative samples also tested negative on Siemens and from the seven

Table 3: Results of clinical validation of three SARS-CoV-2 serologic assays.

Experiment	LIAISON® SARS-CoV-2 IgG		Elecys anti-SARS-CoV-2		Atellica SARS-CoV-2 Total	
	Acceptance criteria according to manufacturer	Results	Acceptance criteria according to manufacturer	Results	Acceptance criteria according to manufacturer	Results
Diagnostic sensitivity						
Overall (n=186)	71.1% [62.7–78.6]	66.7% [59.4–73.4]	77.0 [70.6–82.6]	71.0% [63.9–77.4]	84.8% [79.7–89.0]	72.0% [65.0–78.4]
After a positive PCR on respiratory sample						
0–5 days (n=73)	25% [14.6–39.4]	38.4% [27.2–50.5]	65.5% [56.1–74.1]	48.0% [36.1–60.0]	60.7% [49.7–70.9]	43.8% [32.2–56.0]
6–13 days (n=53)	90.4% [79.4–95.8]	71.7% [57.7–83.2]	88.1% [77.1–95.1]	77.4% [63.8–87.7]	97.5% [92.8–99.5]	79.3% [65.9–89.2]
≥14 days (n=60)	97.4% [86.8–99.5]	96.7% [88.5–99.6]	100% [88.1–100]	93.3% [83.8–98.2]	100% [91.6–100]	100% [94.0–100]
After onset of COVID-19 symptoms						
0–5 days (n=20)	Not reported	20.0% [5.7–43.7]	Not reported	15.0% [3.2–37.9]	Not reported	15.0% [3.2–37.9]
6–13 days (n=69)		53.6% [41.2–65.7]		59.4% [46.9–71.1]		55.1% [42.6–67.1]
≥14 days (n=97)		85.6% [77.0–91.9]		90.7% [83.1–95.7]		95.9% [89.8–98.9]
Diagnostic specificity						
Overall (n=120)	99.0% [98.3–99.5]	95.0% [89.4–98.1]	99.8% [99.7–99.9]	99.2% [95.4–99.9]	99.8% [99.3–100]	100% [97.0–100]
Random selected samples (n=50)	98.9% [94.0–99.8]	100% [92.9–100]	99.8% [99.6–99.9]	100% [92.9–100]	99.8% [99.3–100]	100% [92.9–100]
Samples with other (non-SARS-CoV-2) coronavirus antibodies (n=15)	100% [63.0–100]	100% [78.2–100]	100% [91.2–100]	100% [78.2–100]	Not reported	100% [78.2–100]
Samples with other viral antibodies (n=45)	98% [93.6–99.8]	93.3% [81.7–98.6]	100% [91.2–100]	97.8% [88.2–99.9]	100% [95.0–100]	100% [92.0–100]
Samples with positive rheumatoid factor (n=10)	90% [55.5–99.8]	70.0% [34.8–93.3]	Not reported	100% [69.1–100]	100% [39.8–100]	100% [69.1–100]
Detection capability of antibodies						
Sensitivity (presence of antibodies) (n=131)	Not reported	92.4% [86.4–96.3]	Not reported	97.7% [93.5–99.5]	Not reported	98.5% [94.6–99.8]
Specificity (absence of antibodies) (n=175)		94.9% [90.5–97.6]		97.1% [93.5–99.1]		97.1% [93.5–99.1]

PCR, polymerase chain reaction.

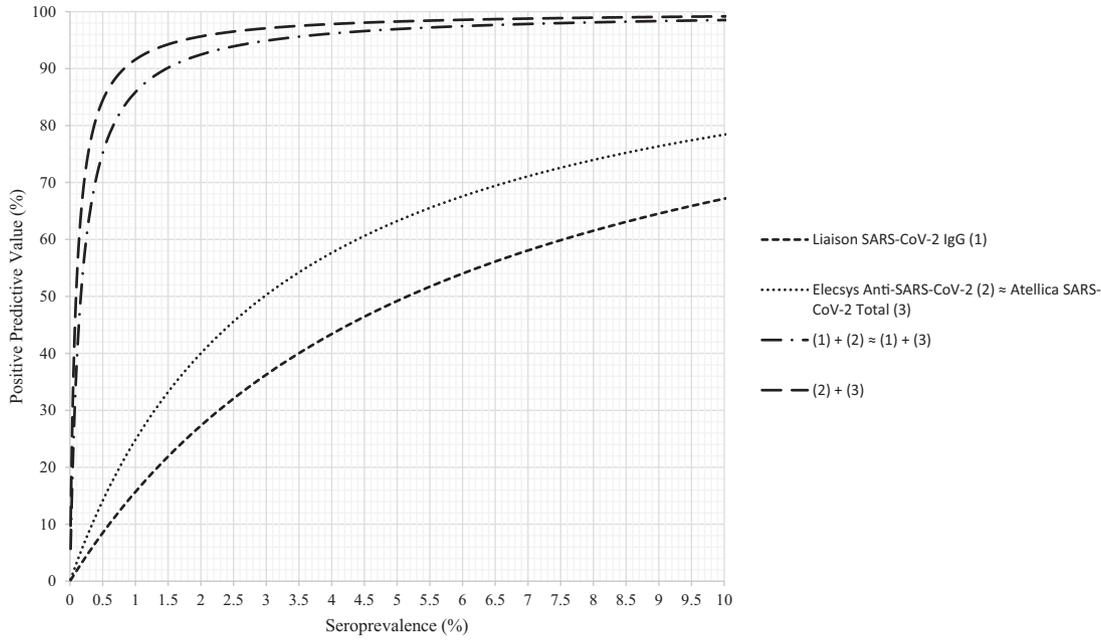


Figure 1: (Combined) positive predictive value as a function of prevalence of three SARS-CoV-2 serologic assays using sensitivity for detection of antibodies.

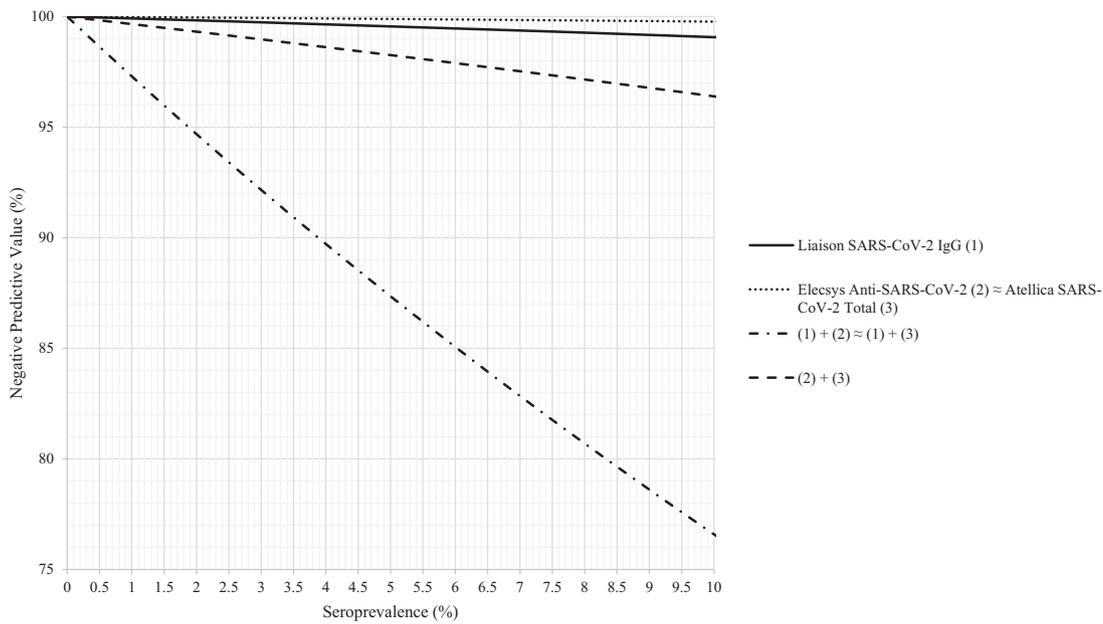


Figure 2: (Combined) negative predictive value as a function of prevalence of three SARS-CoV-2 serologic assays using specificity for detection of antibodies.

positive samples, three also tested positive on Siemens. In total 4 (0.3% of total samples) discordant Roche/Siemens results occurred and were tested with a third antibody assay (DiaSorin). The third antibody assay was negative in all four samples so the final conclusion was that these four

samples were negative for SARS-CoV-2 antibodies and the initial Roche result was false positive. Calculating COVID-19 seroprevalence using only the Roche assay would give 6.0% compared to 5.6% using an orthogonal testing algorithm.

Discussion

This study investigated the analytical and clinical performance of three commercial, fully-automated SARS-CoV-2 antibody assays. Analytical performance of all three assays was acceptable and comparable with results found in other studies [22–25]. Using the cut-off suggested by the manufacturer, diagnostic sensitivity and specificity did not differ significantly from performance specifications found in literature [22–25] or manufacturer's specifications. Diagnostic sensitivity increased in time after a positive SARS-CoV-2 PCR result or onset of symptoms with acceptable sensitivities ≥ 14 days after a positive respiratory SARS-CoV-2 PCR result or onset of symptoms.

Some authors [23, 24] suggested adapted cut-off values in order to optimize diagnostic sensitivity and specificity. Both publications suggested a lower cut-off for the Elecsys Anti-SARS-CoV-2 (>0.165 COI) and LIAISON SARS-CoV-2 IgG (>6.1 AU/mL) assay, increasing diagnostic sensitivity without loss of specificity. If these literature-based cut-off values were applied in our study cohort, a significant higher diagnostic sensitivity was achieved, 84.9% (95% CI 79.0–89.8) for the Elecsys assay and 81.2% (95% CI 74.8–86.5) for the Liaison assay with no significant loss of specificity was observed for the Elecsys assay, being 98.3% (95% CI 94.1–99.8) but for the Liaison assay a significant drop of specificity to 78.3% (95% CI 69.8–85.3) was observed. Adjusted cut-off values were also calculated in this study cohort, optimizing diagnostic sensitivity and specificity but their use did not translate in significant higher diagnostic sensitivity/specificity (except for diagnostic sensitivity of Elecsys). This shows that adapted cut-off values are specific for each laboratory and cannot be exchanged, as significant differences in sensitivity and specificity can be observed. A recent study underscored the importance of redefined cut-off values to improve negative likelihood ratios and overall inter-assay agreement and so a better harmonization of serological tests [26].

Because this study used three serological assays with different epitopes and assay principles, an extended golden standard could be defined, judging if COVID-19 antibodies are present or absent in a sample. Each assay was compared to this standard in order to calculate sensitivity (presence of antibodies) and specificity (absence of antibodies) for antibody detection. As more studies start to investigate seroprevalence in asymptomatic, general populations and many hospitals offer serological COVID-19 testing to health workers (without performing a respiratory SARS-CoV-2 PCR) as part of a local risk management, it is important to know the effect of

seroprevalence on PPV and NPV [27, 28]. In low prevalence settings, it is important to have a high PPV, this can be achieved by choosing a test with a very high specificity ($\geq 99.5\%$) for antibodies detection. In this study none of the assays had a specificity for antibody detection statistically $\geq 99.5\%$ resulting in low PPV (e.g. upto 25%) in low prevalence (e.g. 1%) settings (Figure 1). Another way to ensure a high PPV for detection of antibodies is using orthogonal testing algorithms based on confirmation of a positive result with a different (e.g. antigens) test. Figure 1 shows a significant increase in PPV if positive results are tested on a second platform with the highest combined PPV for a testing algorithm using the Elecsys Anti-SARS-CoV-2 and Atellica IM SARS-CoV-2 Total assay. Testing negative patient samples with a second test is not necessary if the initial assay has a high NPV (Figure 2). All assays evaluated in this study had high NPV across a broad seroprevalence range. The orthogonal test algorithm using optimized cut-off values with Roche as the first test and Siemens as a confirmation test was applied in a local voluntary screening project including 1,151 health workers. The difference in calculated seroprevalence using an orthogonal (seroprevalence 5.6%) or single-assay (seroprevalence 6.0%) test algorithm was small and one could say that seroprevalence is 6% in this cohort, regardless of the testing strategy. However, in some specific situations it is important to have the most correct COVID-19 antibody status for an individual e.g. inclusion criteria for studies investigating COVID-19 vaccine immune response, determination of protective titers or cut-off values once studies have proven protective properties of COVID-19 antibodies or studies in which seroprevalence must be determined very accurately.

Conclusions

To conclude, this study revealed acceptable analytical and clinical performance of three automated SARS-CoV-2 antibody assays. Adjusted cut-off values can be calculated by each laboratory in their specific population but cannot be generalized. In this study the diagnostic sensitivity/specificity did not significantly increase using adjusted cut-offs and manufacturer's defined cut-offs were applied. This study underscored the importance of high specificity if populations with lower seroprevalence are sampled and elaborated the impact of seroprevalence on PPV and NPV even if high sensitivity and specificity is achieved. The relevance of an orthogonal testing algorithms is explained and illustrated. Testing samples with the Elecsys Anti-

SARS-CoV-2 and confirming positive results with the Atellica IM SARS-CoV-2 Total assay (or vice versa) would give the highest PPV and NPV in populations with a lower seroprevalence.

Acknowledgments: The authors want to thank the laboratory staff of the University Hospital Antwerpen for technical assistance. Siemens provided all reagents for this study free of charge. We would like to thank the GRACE consortium for making available some ‘non-SARS-CoV-2’ coronavirus sera.

Research funding: None declared.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Ethical approval: This study has approval of the Ethical Committee of the UZ Antwerpen (reference number: 20/27/356).

References

- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33.
- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–3.
- Wang H, Li X, Li T, Zhang S, Wang L, Wu X, et al. The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis* 2020;39:1629–35.
- WHO Director-General’s opening remarks at the media briefing on COVID-19—11 March 2020. Available from: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020> [Accessed 25 June 2020].
- Younes N, Al-Sadeq DW, Al-Jighefee H, Younes S, Al-Jamal O, Daas HI, et al. Challenges in laboratory diagnosis of the novel coronavirus SARS-CoV-2. *Viruses* 2020;12:582.
- Winter AK, Hegde ST. The important role of serology for COVID-19 control. *Lancet Infect Dis* 2020;20:758–9.
- Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *Lancet* 2020;395:1039–46.
- Jin Y, Wang M, Zuo Z, Fan C, Ye F, Cai Z, et al. Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019. *Int J Infect Dis* 2020;94:49–52.
- Tang Y-W, Schmitz JE, Persing DH, Stratton CW. Laboratory diagnosis of COVID-19: current issues and challenges. *J Clin Microbiol* 2020;58:e00512–20.
- Meyer B, Drosten C, Müller MA. Serological assays for emerging coronaviruses: challenges and pitfalls. *Virus Res* 2014;194:175–83.
- Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining ‘host jump’ of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol* 2015;23:468–78.
- Wu Y, Wang F, Shen C, Peng W, Li D, Zhao C, et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* 2020;368:1274–8.
- Siemens. Atellica: COV2T. Siemens Healthcare Diagnostics; 2020; Rev. A.
- Roche. Cobas 8000: anti-SARS-CoV-2, V1.0; 2020.
- Crescentino. Liaison XL: SARS-CoV-2 S1/S2 IgG. DiaSorin; 2020.
- Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25:2000045.
- Wayne P. CLSI EP15-A3: user verification of precision and estimation of bias; approved guideline, 3rd ed. Wayne: Clinical and Laboratory Standards Institute; 2015.
- Neill Carey R, Hauck WW, Kallner A, Kondratovich MV, Guy Middles James J Pierson-Perry F, et al. Aparna Srinovasan. CLSI EP15-A3: user verification of precision and estimation of bias; approved guideline, 3rd ed. Wayne: Clinical and Laboratory Standards Institute; 2012, vol 34.
- Cofrac. Comité français d’accréditation. COFRAC. SH GTA 04. Guide technique d’accréditation de vérification (portée A)/ validation (portée B) des méthodes en biologie médicale. Available at <http://www.cofrac.fr> [Accessed 10 Sep 2020]. 2015.
- Daniel W, Tholen MK, Rex Astles J, Caffo AL, Happe TM, Krouwer J, et al. Evaluation of the linearity of quantitative measurement procedures: a statistical approach; approved guideline. Wayne: Clinical and Laboratory Standards Institute; 2003, vol 23.
- Broughton PM, Gowenlock AH, McCormack JJ, Neill DW. A revised scheme for the evaluation of automatic instruments for use in clinical chemistry. *Ann Clin Biochem* 1974;11:207–18.
- Favresse J, Eucher C, Elsen M, Marie TH, Dogné JM, Douxfils J. Clinical performance of the Elecsys electrochemiluminescent immunoassay for the detection of SARS-CoV-2 total antibodies. *Clin Chem* 2020;66:1104–6.
- Egger M, Bundschuh C, Wiesinger K, Gabriel C, Clodi M, Mueller T, et al. Comparison of the Elecsys® Anti-SARS-CoV-2 immunoassay with the EDI™ enzyme linked immunosorbent assays for the detection of SARS-CoV-2 antibodies in human plasma. *Clin Chim Acta* 2020;509:18–21.
- Tré-Hardy M, Wilmet A, Beukinga I, Dogné JM, Douxfils J, Blairon L. Validation of a chemiluminescent assay for specific SARS-CoV-2 antibody. *Clin Chem Lab Med* 2020;58:1357–64.
- Kohmer N, Westhaus S, Rühl C, Ciesek S, Rabenau HF. Brief clinical evaluation of six high-throughput SARS-CoV-2 IgG antibody assays. *J Clin Virol* 2020;129:104480.
- Plebani M, Padoan A, Negrini D, Carpinteri B, Sciacovelli L. Diagnostic performances and thresholds: the key to harmonization in serological SARS-CoV-2 assays? *Clin Chim Acta* 2020;509:1–7.
- CDC. Interim guidelines for COVID-19 antibody testing. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html> [Accessed 23 May 2020].
- Diamandis P, Prassas I, Diamandis EP. Antibody tests for COVID-19: drawing attention to the importance of analytical specificity. *Clin Chem Lab Med* 2020;58:1144–5.