

Verena Haselmann*, Mustafa K. Özçürümez, Frank Klawonn, Volker Ast, Catharina Gerhards, Romy Eichner, Victor Costina, Gerhard Dobler, Wolf-Jochen Geilenkeuser, Roman Wölfel and Michael Neumaier

Results of the first pilot external quality assessment (EQA) scheme for anti-SARS-CoV2-antibody testing

<https://doi.org/10.1515/cclm-2020-1183>

Received June 24, 2020; accepted August 7, 2020; published online August 27, 2020

Abstract

Objectives: Assessment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection prevalence and immunity is cornerstones in the fight against CoViD-19 pandemic. For pandemic control, reliable assays for the detection of anti-SARS-CoV-2 antibodies are required. This pilot external quality assessment (EQA) scheme aimed to independently assess the participants' clinical performance of anti-SARS-CoV-2 testing, to identify shortcomings in clinical practice and to evaluate the suitability of the scheme format.

Methods: The EQA scheme consisted of eight serum samples with variable reactivity against SARS-CoV-2 intended for the analysis of anti-SARS-CoV-2 immunoglobulin (Ig)G, IgA, and IgM antibodies. Laboratories reported: (1) results

for each sample and the respective method, (2) raw data from replicate testing of each sample.

Results: The 16 selected pilot EQA participants reported 294 interpreted results and 796 raw data results from replicate testing. The overall error rate for the anti-SARS-CoV-2 IgG, IgA, and IgM tests was 2.7, 6.9, and 16.7%, respectively. While the overall diagnostic specificity was rated as very high, sensitivity rates between 67 and 98% indicate considerable quality differences between the manufacturers, especially for IgA and IgM.

Conclusions: Even the results reported by the small number of participants indicate a very heterogeneous landscape of anti-SARS-CoV-2 serological testing. Differences of available tests and the individual performance of laboratories result in a success rate of 57.1% with one laboratory succeeding for all three antibody-classes. These results are an incentive for laboratories to participate in upcoming open EQA schemes that are needed to achieve a harmonization of test results and to improve serological testing.

Keywords: COVID-19; external quality assessment scheme; proficiency testing; ring trial; SARS-CoV-2.

Roman Wölfel and Michael Neumaier jointly supervised this work.

*Corresponding author: Dr. med. Verena Haselmann, MD, Department of Clinical Chemistry, University Medical Center, Mannheim, Germany. Phone: +49 621 383 2222, Fax: +49 621 383 3819, E-mail: verena.haselmann@umm.de

Mustafa K. Özçürümez, Department of Laboratory Medicine of the Medical Clinic at the University Medical Center Knappschafts Krankenhaus Bochum, Ruhr University, Bochum, Germany

Frank Klawonn, Biostatistics, Helmholtz Centre for Infection Research, Braunschweig, Germany; Department of Computer Science, Ostfalia University of Applied Sciences, Wolfenbüttel, Germany

Volker Ast, Catharina Gerhards, Romy Eichner, Victor Costina and Michael Neumaier, Department of Clinical Chemistry, University Medicine Mannheim, Medical Faculty Mannheim of the University of Heidelberg, Mannheim, Germany

Gerhard Dobler and Roman Wölfel, Bundeswehr Institute of Microbiology, Munich, Germany; German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany

Wolf-Jochen Geilenkeuser, Reference-Institute for Bioanalytics, German Society for Clinical Chemistry and Laboratory Medicine (DGKL), Bonn, Germany

Introduction

Detection of SARS-CoV-2 virus infection or immunological response is an integral part of the fight against the CoViD-19 pandemic. In addition to behavioral and organizational measures, reliable detection systems are particularly needed for effective pandemic control.

Molecular and serological diagnostics of SARS-CoV-2 infections are no longer limited to reference laboratories and a few highly specialized research institutions [1, 2], but are offered by numerous laboratories for standard care. Automated antibody test systems are particularly important for the determination of SARS-CoV-2 seroprevalence at population level [3]. A second application, which is important from a health economic perspective but not yet scientifically fully validated, is the use of antibody testing

to demonstrate immune protection or to check the vaccination effectiveness [4].

As immunoassays for detection of anti-SARS-CoV-2 immunoreactivity are gaining growing attention, the number of commercially available SARS-CoV-2 antibody tests is rapidly increasing [5, 6]. The diversity of these immunoassays using different viral components as ligand-specific binding reagents already implies a certain variability of results. Moreover, given the speed of industrial development and the emergency use authorization (EUA), validation requirements for new test systems had been relaxed [7], resulting in a questionable quality of some assays. Thus, the clinical accuracy and performance of these tests systems are currently being elucidated in several external validation studies [8–16].

However, to date, no reference material is available and no external quality control by the mean of external quality assessment (EQA) was available for quality assurance of anti-SARS-CoV-2 serological testing. EQA schemes represent a key strategy for comparing analytical test performance among laboratories [17, 18], for obtaining a reliable estimation whether laboratory performance meets the required proficiency standard for patient care [19], and for helping to harmonize tests methods [17, 20] and to improve the quality of serological testing [19–21].

In order to meet this demand, the Reference Institute for Bioanalytics (RfB) conducted the first pilot EQA scheme for detection of anti-SARS-CoV-2-specific antibodies in characterized serum samples with the aim to (1) assess the current quality level of laboratory-based SARS-CoV-2 antibody testing and (2) prepare the implementation of an open EQA scheme.

Materials and methods

EQA design

The pilot EQA scheme was conducted by the RfB, a DIN EN ISO/IEC 17043:2010 accredited international EQA provider and initiated on behalf of the CoViD-19 Task Force of the German Society for Clinical Chemistry and Laboratory Medicine. Participation within this pilot scheme was restricted to Task Force member institutions. The EQA scheme consisted of eight serum samples from pseudonymized patients. Patients were recruited at University Medical Centre Mannheim, Germany, as part of the Immunitor study [22]. The study was approved by the Institutional Review Board, and informed written consent was obtained from each subject.

Each laboratory received a 350 μ L blind aliquot of each, characterized sample dedicated to detection of anti-SARS-CoV-2-IgG, anti-SARS-CoV-2-IgA, and anti-SARS-CoV-2-IgM antibodies. The samples

were distributed to all participants at 4 °C on 24th of April 2020. The transport temperature is in agreement with other EQA schemes offered by the RfB and the current recommendations of the European Center for Disease Prevention and Control (ECDC) [23]. Twenty-four hour express delivery was chosen, which is a time frame described by several manufacturers of anti-SARS-CoV-2 immunoassays as acceptable for storage at 4 °C [24, 25]. Each sample dispatch was accompanied by a covering letter giving basic instructions and a reporting sheet. Participants were asked to use their regular routine procedures for determination of the respective anti-SARS-CoV-2 antibodies. Additionally, all participants were instructed to perform triplicate testing, if feasible, provide their raw data for in-depth analyses and report results within 5 days. A general report summarizing the statistics and final results was sent to all participating laboratories, together with a certificate for anti-SARS-CoV-2 serological testing for each Ig class being analyzed correctly.

Preparation and characterization of EQA samples by the reference institute

The EQA samples were prepared according to standard operation procedures as described in the following.

After blood draw, serum samples were stored at ambient temperature for at least 1 h to allow appropriate clotting. Clotted samples were centrifuged at 2,000 g for 10 min at 18 °C within 4 h after sample collection, serum was pooled, aliquoted 350 μ L and finally stored at –80 °C. Aliquots of each sample were used for pre-characterization by the RfB scheme organizer's laboratories prior to sample dispatch.

At the scheme organizer's laboratories, serological testing was performed by two different commercially available immunoassay including the anti-SARS-CoV-2 IgG and anti-SARS-CoV-2 IgA ELISA (Euroimmun, Germany) as well as the EDI™ Novel Coronavirus COVID-19 IgG ELISA and EDI™ Novel Coronavirus COVID-19 IgM ELISA (Epitope Diagnostics). Both tests were validated according to in-house quality management requirements and in agreement with the ISO 15189. All sample stocks were analyzed in triplicates. Assays were performed according to manufacturer's instructions.

Additionally, for each sample, a virus micro-neutralization test (VNT) was performed in a biosafety level 3 laboratory at the Bundeswehr Institute of Microbiology. In detail, serum samples were heat-inactivated at 56 °C for 30 min. In 96-well plates, 50 μ L double serial dilutions, starting from 1:10 to 1:80, was mixed in duplicate with the same volume of a virus stock solution containing 100 TCID₅₀ of the SARS-CoV-2 strain 2019 MUC-IMB-1. The serum-virus mixtures were incubated at 37 °C for 1 h. After incubation, 50 μ L of a Vero E6 cell suspension containing 2×10^5 cells per mL was added to each well. The plates were incubated at 37 °C in a humidified atmosphere with 5% CO₂ and examined under a microscope for signs of CPE after 3 days. The CPE was made visible by staining with 0.1% crystal violet solution containing 13% formalin. The titer of the respective serum was the highest dilution that completely neutralized the challenge dose of SARS-CoV-2. In each run, positive and negative patient serum were included as controls. The concentration of the virus stock was also verified by back titration in each test plate.

Results obtained for sample characterization by the scheme organizer's laboratories are summarized Table 1.

Table 1: Results of sample characterization by the reference institute.

	VNT	IgG (Euroimmune/Epitope)	IgA	IgM	Clinical information
Sample 1	<1:10	Negative/negative	Negative	Negative	SARS-CoV-2 negative patient
Sample 2	1:20	Positive/positive	Positive	Positive	SARS-CoV-2 positive patient, blood-draw 22 days after symptom onset
Sample 3	1:80	Positive/positive	Negative	Negative	SARS-CoV-2 positive patient, blood-draw 36 days after symptom onset
Sample 4	<1:10	Negative/negative	Negative	Negative	SARS-CoV-2 negative patient
Sample 5	1:40	Positive/positive	Positive	Positive	SARS-CoV-2 positive patient, blood-draw 43 days after symptom onset
Sample 6	<1:10	Negative/negative	Negative	Negative	SARS-CoV-2 negative patient
Sample 7	1:20	Positive/positive	Positive	Positive	SARS-CoV-2 positive patient, blood-draw 40 days after symptom onset
Sample 8	<1:10	Negative/negative	Negative	Negative	SARS-CoV-2 negative patient

VNT, virus neutralization test.

Statistical analysis

Results of data analysis are presented as descriptive statistics by mean, 95% confidence interval (CI), and coefficient of variation (CV) as appropriate. For determination of the error rate for anti-SARS-CoV-2, antibody testing results reported for the respective antibody class were considered, and for the method specific error rate, results obtained by laboratories using that particular method were included.

Individual estimates of standard deviations of the log values based on three samples were corrected to be unbiased according to the correction for normal distributions before averaging. The power of this averaged standard deviation was then used as an estimated of the relative standard error.

Bootstrap 95% confidence intervals for the relative standard error were computed based on 10,000 bootstrap samples for all assays together and also separately for the different assays.

In order to identify triplicates with significantly deviating standard deviations, a single sample χ^2 -test was applied to compare the variance in each triplicate with overall estimate of the variance. Bonferroni correction for multiple testing was applied to the p-values.

All statistical analyses were carried out using R version 4.0.0 using the packages DescTools (R Foundation for Statistical Computing).

Results

Participation

Samples were sent to 16 laboratories. About 15/16 (93.8%) laboratories participated in this pilot EQA scheme by submitting their results online, and 14/16 (87.5%) additionally provided their raw data obtained by replicate testing. For evaluation within this manuscript, only results from those 14 participants which submitted their raw data were considered. All laboratories reported in detail on anti-SARS-CoV-2 IgG testing; 10/14 (71.4%) reported in detail on IgA and 5/14 (35.7%) on IgM. Total results submitted for each of the three analytes (anti-SARS-CoV-2 IgG, IgA, and IgM) and those considered for data evaluation might differ from the absolute number of participants as laboratories were allowed to submit results for different test systems. Figure 1 provides an overview of EQA design and data evaluation.

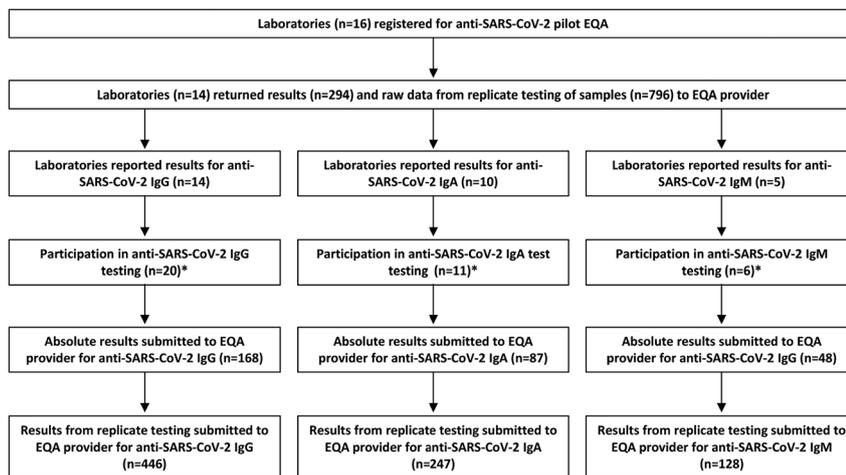


Figure 1: External quality assessment (EQA) design and data evaluation. The flow diagram displays the EQA design and the strategy for data evaluation. Fourteen laboratories returned a total of 194 interpreted results and 796 raw data results from replicate testing of the eight different EQA samples provided. The laboratories could choose to participate in anti-SARS-CoV-2 IgG, IgA, and/or IgM detection. The number of participating laboratories, the number of participation based on reported results per test system, the number of absolute results returned to the EQA provider, and the number of results obtained by replicate testing of samples in the participating laboratories is displayed for all three analytes.

* Laboratories were allowed to submit results for different test systems, accordingly participants were counted for each test system used

Scope and methods

A total of 294 interpreted results were reported from the eight different EQA samples provided, and 796 results from replicate testing. For anti-SARS-CoV-2 IgG, a total of 168 interpreted/446 replicate testing results were submitted, for anti-SARS-CoV-2 IgA, 87 interpreted/247 replicate testing results, and for anti-SARS-CoV-2 IgM 39 interpreted/103 replicate testing results were returned. For analysis of anti-SARS-CoV-2 IgG, 4/14 laboratories reported results for two different test systems and 1/14 laboratory for three different assays resulting in 20 overall results. About 1/10 and 1/5 laboratories reported results for two different assay providers resulting in 11 and six overall results for anti-SARS-CoV-2 IgA and IgM testing, respectively.

All laboratories used commercially available test systems for anti-SARS-CoV-2 antibody detection. IgG and IgA assays from Euroimmun were applied most frequently (10/20 for IgG and 10/11 for IgA), while for IgM testing the Novel Coronavirus COVID-19 IgM ELISA from Epitepe was used by the majority of laboratories (Figure 2). Detailed information about the different kits used is provided in Supplementary Table 1.

Error rate

The overall proficiency was evaluated based on the following two criteria: (1) participants had to report results for all samples correctly and (2) borderline results for anti-SARS-CoV-2 IgG and IgA were considered inappropriate and for anti-SARS-CoV-2 IgM as conditionally correct if the target value of the respective sample was positive. Detailed information about the respective patient's medical history corresponding to each EQA sample is provided in Supplementary Data, results of sample characterization by the scheme organizer are summarized in Table 1.

Table 2 provides an overview of the interpreted results reported for each sample by the participants as well as the results submitted from replicate testing. About 18/294 results were considered inaccurate resulting in an overall error rate of 6.12%. Evaluation of results from replicate testing demonstrated an overall error rate of 5.78% (46/796). In total, 8/14 (57.1%) participants fulfilled above mentioned requirements for successful participation. Noteworthy, only one participant succeeded for all three analytes, three participants for IgG and IgA testing.

For anti-SARS-CoV-2 IgG, a success rate of 80.0% (16/20) was noted with 11/14 (78.6%) laboratories succeeding. Evaluation revealed an error rate of 2.98% (5/168) for interpreted results vs. 2.69% (12/446) for replicate testing.

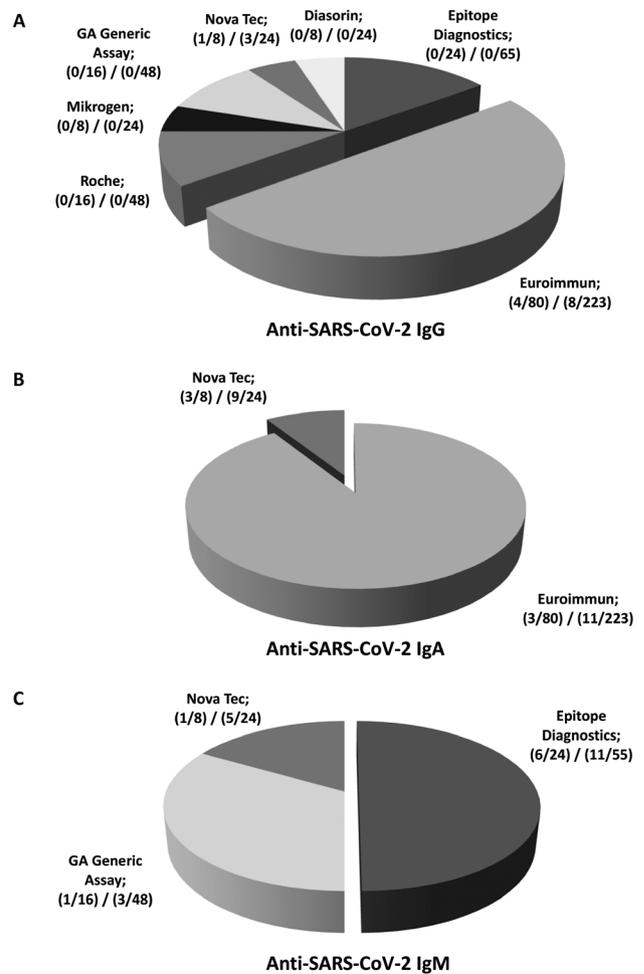


Figure 2: Method-specific error rate for anti-SARS-CoV-2 antibody detection. The different commercially available test systems used by the participants are displayed in the pie chart for (A) anti-SARS-CoV-2 IgG, (B) anti-SARS-CoV-2 IgA, and (C) anti-SARS-CoV-2 IgM. For each method, the number of wrong results compared to the total number of results submitted is indicated. Additionally, separated by a backslash, the number of wrong determinations compared to the total number of results submitted from the replicate testing is depicted for each method.

In detail, an error rate of 0% (0/84) was evaluated for the four negative samples and of 5.95% (5/84) for the four positive samples. Here, false-negative results were reported for sample 2 (VNT 1:20) and 3 (VNT 1:80). Depending on the test system used, 4/80 false results were noted for anti-SARS-CoV-2 IgG ELISA (Euroimmun) and 1/8 for NovaLisa SARS-CoV-2 IgG (Nova Tec) (Figure 2). The CV of laboratories using the Euroimmun test succeeding was lower compared to the CV determined for participants that failed. When assessing the overall and assay-specific variability for anti-SARS-CoV-2 IgG, no differences between the various methods used became obvious (Figure 3, Table 3). Although it was not possible to calculate the intra-

Table 2: Results of anti-SARS-CoV-2 antibody detection.

IgG	Target value	Results submitted to RfB			Results of replicate testing		
		Positive	Borderline	Negative	Positive	Borderline	Negative
Sample 1	Negative	0	0	21	0	0	56
Sample 2	Positive	17	3	1	48	4	4
Sample 3	Positive	20	1	0	53	3	0
Sample 4	Negative	0	0	21	0	0	56
Sample 5	Positive	21	0	0	56	0	0
Sample 6	Negative	0	0	21	0	0	56
Sample 7	Positive	21	0	0	56	0	0
Sample 8	Negative	0	0	21	0	1	53
IgA	Target value	Results submitted to RfB			Results of replicate testing		
		Positive	Borderline	Negative	Positive	Borderline	Negative
Sample 1	Negative	0	1	10	0	2	29
Sample 2	Positive	8	2	1	21	6	4
Sample 3	Negative	0	0	11	0	0	31
Sample 4	Negative	0	0	11	0	0	31
Sample 5	Positive	10	0	1	27	1	3
Sample 6	Negative	0	0	11	0	0	31
Sample 7	Positive	10	0	1	27	1	3
Sample 8	Negative	0	0	10	0	0	30
IgM	Target value	Results submitted to RfB			Results of replicate testing		
		Positive	Borderline	Negative	Positive	Borderline	Negative
Sample 1	Negative	0	0	6	0	0	16
Sample 2	Positive	1	2	3	3	5	7
Sample 3	Negative	0	0	6	0	0	16
Sample 4	Negative	0	0	6	0	0	16
Sample 5	Positive	2	2	2	7	5	4
Sample 6	Negative	0	1	5	1	3	13
Sample 7	Positive	2	2	2	7	4	5
Sample 8	Negative	0	0	6	0	0	15

RfB, Reference Institute for Bioanalytics.

and inter-laboratory variability due to the limited number of participants, comparing the standard deviation of single samples with the overall variance revealed statistically significant differences. Here, the variability of one laboratory (participant #7) was significantly higher for all positive samples compared to the peer ($p < 0.05$).

For anti-SARS-CoV-2 IgA, a success rate of 72.7% (8/11) was obtained with 8/10 (80.0%) laboratories passing. In detail, an error rate of 6.90% (6/87) for absolute results vs. 8.1% (20/247) for replicate testing was seen. 1.9% (1/53) false-positive results and 15.2% (5/33) false-negative results were observed. 4/5 false-negative results were reported for sample 2. Importantly, the Nova Lisa SARS-CoV-2 (COVID-19) IgA (Nova Tec) classified all three positive samples falsely as negative (Figure 2). Figure 3 demonstrates that the analytical variability of anti-SARS-CoV-2 IgA testing was high with some major outliers. Comparable to the results for IgG testing, participant #7 had a significant higher variability compared to the mean ($p < 0.05$).

For anti-SARS-CoV-2 IgM, a success rate of 16.7% (1/6) was noted with 1/5 (20.0%) laboratories succeeding. An error rate of 16.7% (8/48) for absolute results vs. 15.6% (20/128) for replicated testing was evaluated. Overall, 3.0% (1/30) results were false positive and 38.9% (7/18) false negative if borderline results were considered positive. For the assay from Epitope, results between participants diverged substantially with a total of 5/9 (55.6%) false-negative results reported. The GA Generic Assay yielded 1/4 (25%) and the kit from Nova Tec 1/3 (33.3%) false-negative results (Figure 2). The analytical variability of all assays was comparable (Figure 3).

Evaluation of test performance

Table 4 provides an overview of the allocation of the reported results to the target values for anti-SARS-CoV-2 IgG, IgA, and IgM testing for each of the different assays. Additionally, diagnostic sensitivity and specificity with the

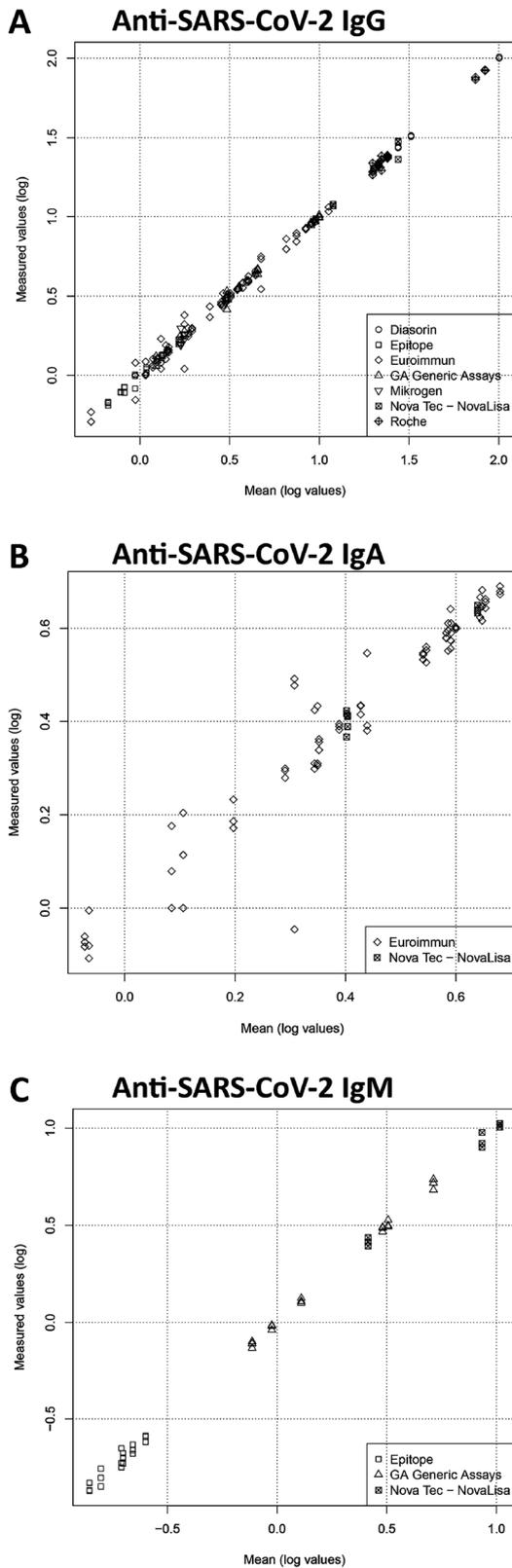


Figure 3: Analytical variability of anti-SARS-CoV-2 antibody detection. For each of the different test systems used by the participating laboratories, the log of the measured value (y-axis) is plotted against the mean of all measured values (x-axis) for each positive

sample. Results are displayed separately for (A) anti-SARS-CoV-2 IgG, (B) anti-SARS-CoV-2 IgA, and (C) anti-SARS-CoV-2-IgM.

Table 3: Evaluation of the method-specific variability of anti-SARS-CoV-2 antibody detection.

	Manufacturer	Relative SD	Lower limit 95% CI	Upper limit 95% CI	Number of samples, n	
IgG	TOTAL	1.11	1.07	1.15	68	
	DiaSorin	1.01	1.01	1.02	4	
	Epitope	1.06	1.03	1.09	8	
	Euroimmun	1.14	1.08	1.21	32	
	GA Generic Assays	1.07	1.03	1.11	8	
	Mikrogen	1.08	1.00	1.14	4	
	Nova Tec	1.09	1.02	1.16	4	
	Roche	1.06	1.02	1.09	8	
	IgA	TOTAL	1.21	1.09	1.34	27
		Euroimmun	1.22	1.09	1.36	24
Nova Tec		1.05	1.02	1.08	3	
IgM	TOTAL	1.07	1.05	1.09	15	
	Epitope	1.09	1.06	1.12	6	
	GA Generic Assays	1.05	1.04	1.06	6	
	Nova Tec – NovaLisa	1.07	1.02	1.10	3	

CI, confidence interval; SD, standard deviation.

respective 95% CI (confidence interval) were calculated. In total, IgG tests yielded a diagnostic specificity of 100% (95% CI 98–100%) and a sensitivity 98% (95% CI 95–99.5%). Although a specificity of 100% was achieved for all assays included, due to the limited number of results submitted the estimated 95% CI ranged from 74 to 100%. Anti-SARS-CoV-2 IgA revealed a comparable diagnostic specificity, while the sensitivity was lower compared to IgG analysis. Low diagnostic sensitivity (67%, 95% CI 48–84%) was revealed for anti-SARS-CoV-2 IgM detection with a moderate specificity (96%, 95% CI 89–98%).

Discussion

To make a significant contribution to the management and containment of SARS-CoV-2 pandemic, reproducibility and reliability of serologic test results are of particular importance. This can be achieved through quality assurance by means of internal and external quality control [26]. EQA schemes are particular important to assess the diagnostic performance of individual laboratories, to harmonize different test methods and to improve the overall quality [17, 19, 20]. Here, we present the results of the first pilot EQA

Table 4: Evaluation of test performance for anti-SARS-CoV-2 antibody detection.

IgG	True positive, n	False negative, n	True negative, n	False positive, n	Sensitivity (95% CI)	Specificity (95% CI)
Epitope Diagnostics	28	0	26	0	1.00 (0.88, 1.00)	1.00 (0.87, 1.00)
Euroimmun	104	4	111	0	0.96 (0.91, 0.99)	1.00 (0.97, 1.00)
Roche	24	0	24	0	1.00 (0.86, 1.00)	1.00 (0.86, 1.00)
Mikrogen	12	0	12	0	1.00 (0.74, 1.00)	1.00 (0.74, 1.00)
GA Generic Assay	24	0	24	0	1.00 (0.86, 1.00)	1.00 (0.86, 1.00)
Nova Tec	9	0	12	0	1.00 (0.66, 1.00)	1.00 (0.74, 1.00)
DiaSorin	12	0	12	0	1.00 (0.74, 1.00)	1.00 (0.74, 1.00)
Total	213	4	221	0	0.98 (0.95, 99.5)	1.00 (0.98, 1.00)
IgA	True positive, n	False negative, n	True negative, n	False positive, n	Sensitivity (95% CI)	Specificity (95% CI)
Euroimmun	75	1	137	0	0.99 (0.93, 1.00)	1.00 (0.97, 1.00)
Nova Tec	0	9	15	0	0.00 (0.00, 0.34)	1.00 (0.78, 1.00)
Total	75	10	152	0	0.88 (0.79, 0.94)	1.00 (0.97, 1.00)
IgM	True positive, n	False negative, n	True negative, n	False positive, n	Sensitivity (95% CI)	Specificity (95% CI)
Epitope Diagnostics	10	4	38	1	0.71 (0.42, 0.92)	0.97 (0.87, 1.00)
GA Generic Assay	9	0	33	3	1.00 (0.66, 1.00)	0.92 (0.78, 0.98)
Nova Tec	0	5	15	0	0.00 (0.00, 0.52)	1.00 (0.78, 1.00)
Total	19	9	86	4	0.67 (0.48, 0.84)	0.96 (0.89, 0.98)

scheme on SARS-CoV-2 serological testing. As of August 1st 2020, this is the first study publishing official EQA results on this topic.

Within this scheme, the 14 participating laboratories providing detailed information reported the use of immunoassays from seven different manufacturers. This indicates a heterogeneous landscape of anti-SARS-CoV-2 serological testing in clinical practice. Interestingly, two participants used the Roche Elecsys anti-SARS-CoV-2 test, which had not been officially launched on time of sample dispatch, but has since been externally validated in several studies [12–14]. It has to be mentioned that results reported for the Roche assay were only considered for IgG as the test does not differentiate between antibody classes [13, 25].

Results of all immunoassays used by the participating laboratories are reported qualitatively and even raw data was not reported as titer. Thus, tests in clinical routine seem to be not validated for evaluation of seroconversion defined as class-switch from IgM to IgG or a fourfold or more increased IgG titer. Such evaluation has been recommended by WHO for MERS-CoV and its suitability for the diagnosis of CoViD-19 is currently subject of research [27]. The presence of a specific seroconversion is rather accomplished by skipping the gray zone. Interestingly, two manufacturers (Roche, DiaSorin) do not define such gray area. However, no influence on the success rate was revealed in this EQA scheme.

In total, the results of this first pilot EQA already indicate a good diagnostic performance of the participating laboratories. This is especially true when considering the short time period since commercial serological tests for

SARS-CoV-2 diagnostics are available. However, detailed analysis points to further improvement potential.

For IgG antibody detection, the highest number of participants and different test systems was demonstrated. Overall, the results showed that the detection of IgG levels is robust in terms of both concordance rate and accuracy. Deviating from regular EQA schemes, all laboratories were asked to perform a triple determination facilitating an estimate of the inaccuracy of the assays. Triplicate measurements of all analyzed samples revealed an intra-assay variation (CV%) below 20% in 95% of all positive samples. The analytical variability of anti-SARS-CoV-2 IgG testing between different methods did not differ substantially. This indicates that harmonization of test results from different providers could be achieved by optimizing cut-offs, as recently supposed by Plebani et al. and others [14, 28–30]. Interestingly, some results reported by laboratories using the same commercial test system differed substantially. This could be explained by manual test processing, various applied spectrophotometers or by pre-analytical or analytical errors. The significant deviation of the variance of a single laboratory from the peer demonstrates a high intra-laboratory variability, which resulted in different qualitative results and thus a clinical performance below average. This illustrates the value of EQAs for a single laboratory, but also that the results provide valuable information for others as revealing the need for replicate testing of clinical specimens to reliably detect anti-SARS-CoV-2 antibodies.

Including the remaining data for the antibody classes IgA and IgM, there is a significant decrease in overall

concordance rates from 97.3% (434/446) for IgG, to 91.9% (227/247) for IgA, and to 84.3% (108/128) for IgM. Noteworthy, the analytical variance diverged substantially for IgA. Hence, IgA testing might currently not be suitable for standard care. For determination of IgM, assays from three different providers were used. For the three positive EQA samples, an error rate of 7/18 (38.9%) was noticed. In contrast, only for one sample without neutralizing activity an inconclusive result with values within the gray zone was reported. Similar to SARS-CoV [31, 32], the immune response against SARS-CoV-2 does not seem to follow the general rule that IgM antibodies usually occur before IgG [2, 27, 33, 34]. Several studies have investigated the detection and time-dependent course of IgA, IgM, and IgG antibodies in SARS-CoV-2 infections [9, 29, 35–37]. Their results indicate that certain methodological peculiarities must be considered for IgM detection, which may also explain the comparatively low concordance rate in our EQA scheme.

The determination of imprecision within this study showed substantial differences between antibody classes with the lowest coefficient of variation for IgG. In particular, there was no obvious correlation between neutralizing antibody titers and measurement signals. For instance, sample 5 (VNT 1:40) yielded the highest signals in several IgG assays (Euroimmun, DiaSorin, GA Generic). All three assays use epitopes of the spike-antigen, which is considered the main target for antibody-mediated virus neutralization in coronaviruses [38, 39]. However, sample five provided the highest measurement signals in almost all other assays regardless of the antigen used. Also independent of the respective antigen used, the signal intensities for samples 3 (VNT 1:80) and 7 (VNT 1:20) were comparable. Even if these aspects go beyond the actual scope of our study, the question must be asked to what extent the currently available antibody tests indicate the presence of a protective immunity.

To estimate test performance, we determined diagnostic sensitivity and specificity for all three analytes included in this scheme. The results revealed a sensitivity of 98, 88, 67% and specificity of 100, 100, 96% for anti-SARS-CoV-2 IgG, IgM, and IgA detection. In principle, results on specificity are comparable between different kits and are consistent with the data provided by the manufacturers [6]. However, the number of samples provided, in particular samples prone to test interferences in immunoassays, is too small to adequately assess diagnostic specificity. In terms of diagnostic sensitivity, the EQA results are below manufacturer's specifications. This might be explained by the fact that samples from patients with oligosymptomatic or subclinical infection in whom the IgG

immune response is reported to be partly moderate were included [40]. For instance, the neutralizing antibody titer of 1:20 for sample 2 was low, which caused false-negative results for IgG. Thus, further schemes are warranted using serial dilutions to stress analytical test sensitivity.

A major limitation of this pilot EQA is the limited number of participants. Nevertheless, the results are already indicating a heterogeneous landscape of anti-SARS-CoV-2 serological testing, and revealing differences between detection of the different antibody classes, test systems and individual performance of laboratories. Thus, this pilot is an incentive for participation in upcoming inter-European or international EQA schemes.

Another limitation of this pilot EQA scheme could result from restriction of participation to CoViD-19 Task Force member institutions, which are either university hospitals or accredited laboratories ensuring high level of laboratory diagnostic quality. Thus, further open EQA schemes will be necessary to clarify whether comparable results can be achieved on a broader level.

In summary, 14 laboratories reported the use of commercially available test systems for anti-SARS-CoV-2 antibody detection from seven different providers based on three different methodological approaches. Manual test procedures further enhanced variations of test results, clearly demonstrating the lack of harmonization that is needed to ensure quality of serological testing [17, 20, 28]. These results are in line with those reported from former pilot schemes [41, 42]. Before anti-SARS-CoV-2 serological testing can be applied for clinical decision-making, for estimation of seroprevalence on population level or to adopt restriction regulations to current demands, an optimization and harmonization of tests is urgently required. For the first time, this pilot EQA scheme revealed objectifiable differences between testing modalities and Ig classes. This can help laboratories to become aware of intra-laboratory and assay-specific shortcomings as a prerequisite for an optimized diagnostic. From the point of view of an EQA provider, results obtained demonstrate the feasibility of the EQA design and will allow us to open this scheme on an international level.

Acknowledgments: We thank Laura Mirbach and Angelika Duda for organizing sample dispatch, and Jan-Hendrik Haselmann for critical reading of the manuscript.

Research funding: None declared.

Author contributions: VH, MKÖ, RW, and MN designed the study. CG and VH enrolled patients. GD performed neutralizing antibody tests. RW, VH, VC, and RE performed serological assays for sample characterization prior to shipment. WJG and VH were responsible for data

collection and management. FK, VA, and VH were responsible for biostatistics analyses. VH, MKÖ, and RW were responsible for interpretation of data. VH and RW prepared the tables and figures. VH, MKÖ, and RW drafted the manuscript. All authors contributed to revision of the manuscript and approved it for submission. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

References

- Haveri A, Smura T, Kuivanen S, Osterlund P, Hepojoki J, Ikonen N, et al. Serological and molecular findings during SARS-CoV-2 infection: the first case study in Finland, January to February 2020. *Euro Surveill* 2020;25. <https://doi.org/10.2807/1560-7917.es.2020.25.11.2000266>.
- Wolfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Muller MA, et al. Virological assessment of hospitalized patients with covid-2019. *Nature* 2020; 581: 465–9.
- Lee CY, Lin RTP, Renia L, Ng LFP. Serological approaches for covid-19: epidemiologic perspective on surveillance and control. *Front Immunol* 2020; 11: 879.
- Padron-Regalado E. Vaccines for SARS-CoV-2: lessons from other coronavirus strains. *Infect Dis Ther* 2020;1–20. <https://doi.org/10.1007/s40121-020-00300-x>. 32328406.
- Find. Covid-19 diagnostics resource center. Available from: <https://www.finddx.org/covid-19/sarscov2-eval-immuno/> [Accessed 21st of May 2020].
- Center for Health Security. Johns Hopkins Bloomberg School of Public Health. Available from: <https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html> [Accessed 21st of May 2020].
- FDA. Covid19 IVD. Available from: <https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations-covid19ivd> [Accessed 21st of May 2020].
- Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. *Diagnostics (Basel, Switzerland)* 2020;10. <https://doi.org/10.3390/diagnostics10050319>.
- Whitman J, Hiatt J, Mowery C, Shy B, Yu R, Yamamoto T, et al. Test performance evaluation of SARS-CoV-2 serological assays. preprint at: medRxiv 2020042520074856. <https://doi.org/10.1101/2020.04.25.20074856>.
- Tang MS, Hock KG, Logsdon NM, Hayes JE, Gronowski AM, Anderson NW, Farnsworth CW. Clinical performance of two SARS-CoV-2 serologic assays. *Clin Chem* 2020. <https://doi.org/10.1093/clinchem/hvaa120>. 32402061.
- Perera RA, Mok CK, Tsang OT, Lv H, Ko RL, Wu NC, et al. Serological assays for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), March 2020. *Euro Surveill* 2020;25. <https://doi.org/10.2807/1560-7917.es.2020.25.16.2000421>.
- Haselmann V, Kittel M, Gerhards C, Thiaucourt M, Eichner R, Costina V, Neumaier M. Comparison of test performance of commercial anti-SARS-CoV-2 immunoassays in serum and plasma samples. *Clin Chim Acta* 2020; 510: 73–8.
- Egger M, Bundschuh C, Wiesinger K, Gabriel C, Clodi M, Mueller T, Dieplinger B. Comparison of the elecsys(r) 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.
- Favresse J, Euchet C, Elsen M, Marie TH, Dogne JM, Douxfils J. Clinical performance of the elecsys electrochemiluminescent immunoassay for the detection of Sars-CoV-2 total antibodies. *Clin Chem* 2020. <https://doi.org/10.1111/bjh.16954>.
- Favresse J, Euchet C, Elsen M, Laffineur K, Dogne JM, Douxfils J. Response of anti-SARS-CoV-2 total antibodies to nucleocapsid antigen in covid-19 patients: a longitudinal study. *Clin Chem Lab Med* 2020. <https://doi.org/10.1111/bjh.16954>.
- Jaaskelainen AJ, Kuivanen S, Kekalainen E, Ahava MJ, Loginov R, Kallio-Kokko H, et al. Performance of six SARS-CoV-2 immunoassays in comparison with microneutralisation. *J Clin Virol* 2020; 129: 104512.
- Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem* 2011; 57: 1670–80.
- Haselmann V, Geilenkeuser WJ, Helfert S, Eichner R, Hetjens S, Neumaier M, Ahmad-Nejad P. Thirteen years of an international external quality assessment scheme for genotyping: results and recommendations. *Clin Chem* 2016; 62: 1084–95.
- Rumer L, Domingo C, Donoso Mantke O, Dobrydneva Y, Greiner M, Niedrig M. Statistical approach for optimization of external quality assurance (EQA) studies of molecular and serological viral diagnostics. *Clin Chem Lab Med* 2016; 54: 1589–98.
- Markey K, Douglas-Bardsley A, Asokanathan C, Fry NK, Barkoff AM, Bacci S, et al. Improvement in serological diagnosis of pertussis by external quality assessment. *J Med Microbiol* 2019; 68: 741–7.
- Zhang K, Wang L, Lin G, Sun Y, Zhang R, Xie J, Li J. Results of the national external quality assessment for toxoplasmosis serological testing in China. *PLoS One* 2015; 10: e0130003.
- (UMM) UCM. Immunitor. Available from: <https://www.umm.de/institut-fuer-klinische-chemie/immunitor-studie/> [Accessed 4 June 2020].
- ECDC. Laboratory support for covid-19 in the EU/EEA. Available from: <https://www.ecdc.europa.eu/en/novel-coronavirus/laboratory-support> [Accessed 30 07 2020].
- Euroimmun. Anti-SARS-CoV-2 ELISA (IgG) – instruction for use. <https://www.fda.gov/media/137609/download> [Accessed 30 07 2020].
- Roche. Roche. Elecsys® anti-SARS-CoV-2. Available from: <https://diagnostics.roche.com/us/en/products/params/elecsys-anti-sars-cov-2.html> [Accessed 30 07 2020].
- World Health Organization (WHO). Coronavirus disease (covid-19) pandemic. Geneva: WHO; 2020. Available from: <http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic> [Accessed 21 May 2020].
- Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-2 in patients with covid-19. *Nat Med* 2020. <https://doi.org/10.1038/s41591-020-0897-1>. 32350462.
- 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.

29. Favresse J, Bayart JL, Stoefs A, Gheldof D, Douxfils J, Dogne JM, Gruson D. Neutralization of biotin interference: preliminary evaluation of the VeraTest Biotin, VeraPrep Biotin and BioT-Filter(r). *Clin Chem Lab Med* 2020; 58: e130–3.
30. Tre-Hardy M, Wilmet A, Beukinga I, Favresse J, Dogne JM, Douxfils J, Blairon L. Analytical and clinical validation of an ELISA for specific SARS-CoV-2 IgG, IgA, and IgM antibodies. *J Med Virol* 2020. <https://doi.org/10.1002/jmv.26303>.
31. Li G, Chen X, Xu A. Profile of specific antibodies to the SARS-associated coronavirus. *N Engl J Med* 2003; 349: 508–9.
32. Woo PC, Lau SK, Wong BH, Chan KH, Chu CM, Tsoi HW, et al. Longitudinal profile of immunoglobulin g (IgG), IgM, and IgA antibodies against the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in patients with pneumonia due to the SARS coronavirus. *Clin Diagn Lab Immunol* 2004; 11: 665–8.
33. Lippi G, Salvagno GL, Pegoraro M, Militello V, Caloi C, Peretti A, et al. Assessment of immune response to SARS-CoV-2 with fully automated MAGLUMI 2019-nCoV IgG and IgM chemiluminescence immunoassays. *Clin Chem Lab Med* 2020. <https://doi.org/10.1515/cclm-2020-0473.32301750>.
34. Okba NMA, Muller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, et al. Severe acute respiratory syndrome coronavirus 2-specific antibody responses in coronavirus disease 2019 patients. *Emerg Infect Dis* 2020;26. <https://doi.org/10.3201/eid2607.200841>.
35. Adams ER, Ainsworth M, Anand R, Andersson MI, Auckland K, Baillie JK, Barnes E, et al. Antibody testing for COVID-19: a report from the national COVID scientific advisory panel. Preprint at medRxiv 2020. <https://doi.org/10.1101/2020.04.15.20066407>.
36. Grzelak LTS, Planchais C, Demeret C, Huon C, Guivel F, Staropoli I, et al. SARS-CoV-2 serological analysis of COVID-19 hospitalized patients, pauci-symptomatic individuals and blood donors. Preprint at medRxiv 2020. <https://doi.org/10.1101/2020.04.21.20068858>.
37. Van Elslande J, Houben E, Depypere M, Brackenier A, Desmet S, Andre E, et al. Diagnostic performance of seven rapid IgG/IgM antibody tests and the Euroimmun IgA/IgG ELISA in COVID-19 patients. *Clin Microbiol Infect* 2020; 26: 1082–7.
38. Nie Y, Wang G, Shi X, Zhang H, Qiu Y, He Z, et al. Neutralizing antibodies in patients with severe acute respiratory syndrome-associated coronavirus infection. *J Infect Dis* 2004; 190: 1119–26.
39. Zhou G, Zhao Q. Perspectives on therapeutic neutralizing antibodies against the novel coronavirus SARS-CoV-2. *Int J Biol Sci* 2020; 16: 1718–23.
40. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention. *JAMA* 2020. <https://doi.org/10.1001/jama.2020.2648>.
41. Haselmann V, Ahmad-Nejad P, Geilenkeuser WJ, Duda A, Gabor M, Eichner R, et al. Results of the first external quality assessment scheme (EQA) for isolation and analysis of circulating tumour DNA (ctDNA). *Clin Chem Lab Med* 2018; 56: 220–8.
42. Niedrig M, Leitmeyer K, Lim W, Peiris M, Mackenzie JS, Zambon M. First external quality assurance of antibody diagnostic for SARS-new coronavirus. *J Clin Virol* 2005; 34: 22–5.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2020-1183>).