Final WHO SARS-CoV-2 serology test kit evaluation results

21 July 2022

Final WHO SARS-CoV-2 serology test kit evaluation results

Work conducted for WHO by the National Serology Reference Laboratory (NRL), Australia, a WHO Collaborating Centre and authorized WHO IVD Prequalification Evaluation Laboratory.

Contents

Key points

- The concordance to RNA positivity for Rapid Diagnostic Tests (RDTs) testing for IgG and IgM ranged from 77.4 to 100% and 55.8 to 99.5% respectively, whereas the EIA reported concordance ranging from 60.1 to 100%.
- Specificity of RDT IgG and IgM ranged from 81.0 to 99.0% and 34.3 to 99.0% and EIAs reported specificities of 74.0 to 100%.
- Precision of Enzyme Immunoassays (EIA) ranged from 3.7 to 15.5 % coefficient of variation (this was not measured for RDTs as the readout is visual).
- Cross reactivity and interference were detected in most assays with some having an unacceptable percentage of false reactions.
- This study demonstrates a wide range of performance characteristic of SARS-CoV-2 assays and highlights the need for comprehensive, comparative testing of assay performance in an emergency setting.
- **DISCLAIMER**: This document and information made available in it, are provided for information only and do not constitute and should not be construed as a recommendation by the WHO for specific test kits, or any advice to order, buy or use any specific test kit. Any Member State, United Nations or other intergovernmental organization, non-profit organization and/or other third party that intends to use and/or uses test kits set forth herein should, prior to making any procurement decisions, conduct its own assessment with respect to: (i) the relevant test kits sought to be procured (including, but not limited to, as to their registration status in the countries of intended use); (ii) the relevant manufacturers (including, but not limited to, as to their good standing, reputation, financial stability and ability to supply the required quantities of the test kits within the required deadlines); and (iii) its own specific objectives, needs, financial situation, risk tolerance and other relevant criteria.

  The mention in this document of any test kit and/or manufacturer does not constitute or imply any approval, endorsement or certification by WHO of such test kit or manufacturer. Accordingly, the mention of any test kit or manufacturer in it, may not be used by any manufacturer or third party for any marketing, promotional or commercial purposes. In addition, the name, acronym or emblem of the World Health Organization may not be used by any manufacturer or third party for any purpose (including, without limitation, for any marketing, promotional or commercial purposes)".

Intended Audience

This report is intended for heads of public health laboratories, policy makers, public health professionals, and researchers.

Intended Use of Study

The findings of this study are intended to inform ministries of health, implementing partners, funding bodies and testing sites on the utility of serology for SARS-CoV-2 and to provide evidence for use in the selection of test kits. As the potential use of SARS-CoV-2 serology can vary between settings, the study does not seek to recommend any test kit for particular purposes, as the selection of test kits depends on the intended purpose.

Introduction

In November 2019, a novel acute respiratory disease (COVID-19) caused by a new coronavirus (SARS-CoV-2) was first recognized. Since this time, a major global pandemic has ensued, causing significant mortality, and morbidity and economic disruption. Due to the high level of concern regarding the spread of SARS-CoV-2 infection, regulators in many countries initially reduced their usual strict regulatory requirement that *In-vitro* diagnostic device (IVD) manufacturers demonstrate evidence of adequate performance, safety and quality. Most, if not all regulators allowed use of IVDs under emergency use conditions requiring limited pre-market evidence of performance. Within six months of the start of the pandemic, more than 700 IVDs used for the diagnosis of SARS-CoV-2 infection were available on the market. Over the same period of time the performance of these IVDs were assessed. Many of these early studies were poorly structured, had inappropriate interpretation of test results and assessed small numbers of test kits brands in each study and few findings were published in peer-reviewed journals (1-3). In April 2020, World Health Organization (WHO) published "Advice on the use of point-of-care immunodiagnostic tests for COVID-19" (4). At that time, WHO recommended that rapid serology tests be used as research only tool and "*They should not be used in any other setting, including for clinical decision-making, until evidence supporting use for specific indications is available*." The WHO also stated "*although research into their performance and potential diagnostic utility is highly encouraged*" (4).

In order to shape guidance and inform procurement of serological assays, the WHO commissioned the National Serology Reference Laboratory, Australia (NRL), a WHO Collaborating Centre and authorized WHO IVD Prequalification Evaluation Laboratory, to establish a comprehensive protocol and conduct a comparative evaluation of serology tests for

SARS-CoV-2. The protocol was designed to assess the performance characteristics of the IVDs using a large and diverse set of carefully selected samples collected from ambulatory individuals with a history of PCR-positive SARS-CoV-2 infection and from healthy individuals prior to the SARS-CoV-2 pandemic. The samples were acquired in sufficient volume to conduct head-to-head evaluations of up to 40 unique test kits. In November 2020, WHO issued an open call for expression of interest for manufacturers of SARS-CoV-2 antibody detecting assays, including rapid test devices (RDTs) and enzyme-immunoassays (EIAs) to participate in the evaluation scheme. One-hundred-and-two products from 71 manufacturers were submitted and 45 products from 44 manufacturers were accepted based on compliance with both entry and short listing criteria and after de-duplication of re-branders. Nine of the shortlisted manufacturers later withdrew their application for various reasons. A total of 35 products were evaluated. Starting in July 2021, NRL and WHO began publishing results in the form of individual product reports. This final report combines the findings from all 35 products, describes key observations and implications for use of these IVDs for clinical care and surveillance.

Aims

The primary aims of this study was to:

- Produce statistically significant and scientifically robust assessment of the performance of test kits in different formats, designed to detect antibodies to SARS-CoV-2;
- Compare performance data from a range of commercially available IVDs by testing each test on the same panel of samples;
- Create a database of IVD performance results to inform the selection of test kits used for screening, confirmation, and possibly, diagnosis of SARS-CoV-2 infection in particular for low-middle income countries;
- Collect testing data that may allow for the assessment of analytes such as IgA and IgM, which do not have reference or “gold standard” methods available.

This report summarizes the analysis of results when testing the IVD detailed below on the performance panel of samples.

Nature and Use of Report

This report details the results of a comprehensive, head-to-head study of selected SARS-CoV-2 serology assays identified as being suitable for use in low- and middle-income countries. The purpose of the study was to generate information that can be used for the selection of serology test kits for various purposes, including but not limited to, confirmation of post-exposure to

SARS-CoV-2 from infection or vaccination and for use in seroprevalence studies. The data generated may be used to determine whether serology, in particular IgM testing, could serve any useful purpose in the diagnosis of acute infection, particularly directly after the viral load becomes undetectable. Data summarized in the report and associated supplemental data can potentially be used to inform the selection of test kits in the establishment of testing algorithms to maximize sensitivity and specificity. Additionally, the evaluation identifies which test kits that report false reactivity when testing samples containing potentially cross-reacting and interfering substances, in particular antibodies from individuals infected with other Coronaviruses.

Expression of Interest and Product Selection

In November 2020, WHO issued an expression of interest along with the evaluation protocol. Manufacturers were provided several weeks to apply (deadline was set to 7 December 2020) (5). The following exclusion criteria were used to shortlist the initial list of 102 products from 71 manufacturers:

- exclusion of products targeting IgM or IgA only;
- exclusion of products needing proprietary platforms;
- exclusion of products for which IFU was not included in the application;
- exclusion of manufacturers without a free-sales certificate or ISO13485 accreditation;
- exclusion of products that had low accuracy in early evaluation performed by FIND (6), with (low accuracy defined as <80% sensitivity and <98% specificity);
- exclusion of rapid diagnostic tests targeting anti-N antibodies only;
- exclusion of multiple products from the one manufacturer, with the exception of ELISA targeting anti-N antibodies, which were accepted even if manufacturer already had another shortlisted product.

The latter two criteria were adopted considering the potential of future seroprevalence studies after mass vaccination campaigns with spike-based vaccines, assuming that anti-N antibodies would be helpful to distinguish vaccination from natural infection, and that this type testing would require the high accuracy of a product in ELISA format rather than one in lateral-flow format, which is consistently less accurate across products (2). The exclusion of rapid diagnostic tests targeting anti-N antibodies only was done based on the observations in the literature that anti-N antibody titers decayed more quickly than anti-S antibodies and therefore may be less sensitive to detect past infection (7). Manufacturers that failed to provide the required number of tests (for the amount of required tests (5), failed to commit to strict timelines for shipping or withdrew their application were also excluded. Of the included test kits selected, 26 were RDT

using lateral flow technologies, and eight were ELISAs. One ELISA (Omnipath) was added at a later stage to generate head-to-head comparison data on this specific assay, as this was the commercialized version of the product used within the RECOVERY trail, which recommended use of serology to distinguish seropositive and seronegative and provide treatment with a monoclonal antibody cocktail (Regeneron) based on serology testing (8). The final complete list of test kits evaluated are summarized in **Table 1**. The selected manufacturers signed a confidentiality agreement, which identified the proposed method of communications of results, and gave the manufacturers 30 working days to review the final report for their test kit prior to the report being published on a public website. All selected test kits were provided to NRL free-of-charge. The manufacturer paid the costs of shipping and importation. NRL used valid import permits to obtain entry through Australian customs.

Table 1. Final list of test kits included in the WHO SARS-CoV-2 serology evaluations, including abbreviations used in the report.

Abbreviation	Manufacturer	Product Name	Product Code	IFU[1] Version	Test Type	Target
AllTest G	Hangzhou AllTest Biotech Co., Ltd.	COVID-19 IgG Rapid Test	INCPG-402	146253502 (2020/11/07)	RDT[2]	IgG
AllTest G/M	Hangzhou AllTest Biotech Co., Ltd.	2019-nCoV IgG/IgM Rapid Test Cassette	INCP-402B	146347400	RDT	IgG/IgM
Artron	Artron Laboratories Inc.	Artron COVID-19 IgM/IgG Antibody Test	A03-51-322	A03-51-322 Ver.09 (01/2021)	RDT	IgG/IgM
Biocan	Biocan Diagnostics Inc	Biocan Novel Coronavirus (COVID-19) IgG/IgM Antibody Test	B251C	B251C (07/2020)	RDT	IgG/IgM
Biogenix	Biogenix Inc. Pvt. Ltd.	SARS CoV-2 IgM/IgG Rapid Test	Not provided	BIPL-IFU-081	RDT	IgG/IgM
Biohit	Biohit Healthcare (Hefei) Co., Ltd.	Biohit SARS-CoV-2 IgM/IgG Antibody Test Kit	207.01.25.02	Ver 03 (2020/06/16)	RDT	IgG/IgM
bioLytical	bioLytical Laboratories, Inc.	Insti COVID-19 Antibody Test	90-1098	51-1311F (11/03/2021)	RDT	Total Ab
BioMedomics	BioMedomics, Inc.	COVID-19 IgM-IgG Rapid Test	51-002	51-PI-002.CE (Rev 04) (2020/12/16)	RDT	IgG/IgM
BioRad	Bio-Rad	Platelia SARS-CoV-2 Total Ab	12013798	16008267 (2020/06)	EIA[3]	Total Antibody
Boson	Xiamen Boson Biotech Co., Ltd.	Rapid 2019-nCoV IgG/IgM Combo Test Card	1N38C2	081985 /200612	RDT	IgG/IgM
BTNX	BTNX Inc.	Rapid Response COVID-19 IgG/IgM Rapid Test Device	COV-13C25	1110032621 Rev:5.2 (2021-02-11)	RDT	IgG/IgM
Core Tech	Core Technology Co., Ltd.	Coretests™ COVID-19 IgM/IgG Ab Test	Not provided	COVID IgM/IgG/05-C Ver, 1.3 (09/2020)	RDT	IgG/IgM
Deepblue	Anhui Deepblue Medical Technology Co., Ltd.	COVID-19 (SARS-CoV-2) IgG/IgM Antibody Test Kit (Colloidal Gold)	Not provided	COVID IgG/IgM-01, Ver.1.8	RDT	IgG/IgM

Abbreviation	Manufacturer	Product Name	Product Code	IFU[1] Version	Test Type	Target
Dynamiker	Dynamiker Biotechnology (Tianjin) Co., Ltd.	2019-nCoV IgG/IgM Rapid Test	DNK-1419-1	CE-SYSM-008 1.1 (21/01)	RDT	IgG/IgM
Epitope	Epitope Diagnostics, Inc.	EDI™ Novel Coronavirus COVID-19 IgG ELISA	KT-1032	Ver 13 (2021-05)	EIA	IgG
Getein	Getein Biotech, Inc.	One Step Test for Novel Coronavirus (2019-nCoV) IgM/IgG antibody (Colloidal Gold)	CG2057	WCG76-SIN-DX-S-01	RDT	IgG/IgM
Healgen	Healgen Scientific Limited Liability Company	COVID-19 IgG/IgM Rapid Test Cassette	GCCOV-402a	B21901-01 (2020-06-05)	RDT	IgG/IgM
Joysbio	JOYSBIO (Tianjin) Biotechnology Co., Ltd.	JOYSBIO COVID-19 IgG/IgM Rapid Test Kit	Not provided	V.1.0 (20/08)	RDT	IgG/IgM
Lysun	Hangzhou Lysun Biotechnology CO., LTD.	2019-nCoV IgG/IgM Rapid Test Device (Colloidal Gold)	COV-102	01 (2020-05)	RDT	IgG/IgM
MDGen	MicroDigital Co., Ltd.	MDGen AB96-COVID-19 IgG	Not provided	1.0.3	EIA	IgG
MPBio	MP Biomedicals Asia Pacific Pte. Ltd.	VivaDiag SARS-CoV-2 IgM/IgG Rapid Test	43140-020	MDC0011-ENG-3 (2020/11)	RDT	IgG/IgM
Omega	Genesis Diagnostics Ltd (subsidiary of Omega Diagnostics Group PLC)	Omega Diagnostics COVID-19 IgG ELISA Kit	ODL150/10	263-ODL 150/10 Ver 6.0 (02 / 2021)	EIA	IgG
OmniPath	Thermo Fisher Diagnostics	OmniPATH™ 96 Combi SARS-CoV-2 IgG ELISA Kit	R250120	IFU X9487A, Revised December 2020	EIA	IgG (Quantitative)
RightSign	Hangzhou Biotest Biotech Co., Ltd.	RightSign COVID-19 IgG/IgM Rapid Test Cassette	INGM-MC42	RP5381700 (2021/03/05)	RDT	IgG/IgM
Sensing	Sensing Self Pte. Ltd	COVID-19 Rapid IgG/IgM Combined Antibody Assay Pre-Screening Test Kit	ERCSS05310	Ver 1.1 (2021/02/19)	RDT	IgG/IgM
Serion	Institut Virion\Serion GmbH	SERION ELISA *agile* SARS-C0V-2 IgG	ESR400G	V a400AG-1 (20/09)	EIA	IgG

Abbreviation	Manufacturer	Product Name	Product Code	IFU[1] Version	Test Type	Target
Singclean	Hangzhou Singclean Medical Products Co., Ltd.	COVID-19 IgG/IgM Test kit (Colloidal Gold Method)	Not provided	A/0 (2020/08/03)	RDT	IgG/IgM
Standard Q	SD Biosensor, Inc.	Standard Q COVID-19 IgM/IgG Plus Test	Q-NCOV-02C	L23COV7DMENRO (2021/02)	RDT	IgG/IgM
Sugentech	Sugentech Inc	SGTi-flex COVID-19 IgM/IgG	COVT025E	IS209E-05 (2020.09.01)	RDT	IgG/IgM
Sure Status	Premier Medical Corporation Private Limited	Sure Status COVID-19 IgG/IgM Card Test	SS02P25	SS02-INS-001 (2020/11/10)	RDT	IgG/IgM
Vazyme	Nanjing Vazyme Medical Technology Co., Ltd.	Anti-SARS-CoV-2 Neutralizing Antibody ELISA	C8909C	February 1, 2021	EIA	Nt Ab[4]
Vircell	Vircell, S.L	COVID-19 ELISA IgG	G1032	L-G1032-EN-05 (2021/02/23)	EIA	IgG
VivaDiag	VivaChek Biotech (Hangzhou) Co., Ltd.	VivaDiag SARS-CoV-2 IgM/IgG Rapid Test	VID35-08-011	1604003703 (2020/04/20)	RDT	IgG/IgM
Wantai	Beijing Wantai Biological Pharmacy Enterprise Co., Ltd.	Wantai SARS-CoV-2 Ab ELISA	WS-1096	V.2020-02 (2020/06/22)	EIA	Total Ab
Wondfo	Guangzhou Wondfo Biotech Co., Ltd.	2019-nCoV Antibody Test (Lateral Flow Method)	W195P0004	A2 (2020/12/08)	RDT	Total Ab

1. **IFU - Instructions for use**
2. **RDT - Rapid Diagnostic Device**
3. **EIA - Enzyme immunoassay**
4. **Nt Ab - Neutralizing antibodies**

Methods

The protocol assesses the performance of tests according to several key criteria summarized below.

The performance characteristics evaluated depended on the class(es) of antibodies being detected and the method of reporting of results. All test kits were evaluated for:

- Sensitivity (concordance with SARS-CoV-2 RNA positivity)
- Specificity
- Analytical Sensitivity
- Quantification
- Lot-to-lot Variation
- Seroconversion
- Cross-reactivity
- Interference

Test kits reporting quantitative results (e.g sample to cut-off values [S/Co]) were evaluated for:

- Repeatability
- Reproducibility

It should be noted that some samples in this panel contained varying anticoagulants including sodium citrate or citrate dextrose solution. The anticoagulant of some other samples was not known. Some test kits evaluated specify the use of certain anticoagulants in the IFU. False reactivity due to particular anticoagulants used in the panel cannot be discounted, and results should be interpreted accordingly.

Test kits reporting results in international units per milliliter (IU/mL) or other quantitative units were tested for the comparison of quantification against a dilution series of the WHO international standard.

Sensitivity: A total of 199 serum or plasma specimens obtained from individuals with a recent history of clinical infection with ancestral SARS-CoV-2, confirmed by nucleic acid testing (NAT), were used as the sensitivity panel. As samples were collected from individuals infected with SARS-CoV-2 between January to April 2020, it is assumed that no samples were obtained from individuals infected with Delta or Omicron variants in this study. These samples were collected from time periods that ranged from 14 days to 71 days post onset of symptoms or post NAT positive result. Some samples may not have anti-SARS-CoV-2 IgG if they were collected prior to

developing an immune response. If this was the case, all test kits were subjected to the same challenge. Also note, there are no reference tests for IgG, IgM or IgA specific antibodies to SARS-CoV-2 and the immune reaction to these antibody classes are poorly elucidated. Therefore "sensitivity" of assays specifically for IgG, IgM, IgA and combined IgG/IgM should be interpreted with these caveats in mind. For this reason, the results of the positive panel of samples are reported as "Concordance with Recent Infection". Approximately half of the sensitivity panel samples were tested on one reagent lot and the other half tested on a separate reagent lot which were provided by the manufacturer.

Specificity: A total of 300 plasma specimens obtained from NRL's sample bank, having been collected prior to November 2019, were used as the specificity panel. These samples were plasma samples from healthy blood donors collected from different countries. All samples were screened negative for blood-borne infections by serology and NAT. Given the emergence of SARS-CoV-2 cases in late 2019, these specimens are assumed negative for SARS-CoV-2 antibodies and no further confirmation testing was performed. Approximately half the panel of samples were tested on one reagent lot and the other half tested on a separate reagent lot.

Analytical Sensitivity/Lot-to-lot Variation: Three of the 199 samples included in the clinical sensitivity panel each had 10 doubling dilutions, from 1:2 to 1:1024, prepared in human plasma negative for SARS-CoV-2 antibodies using samples from the Specificity panel detailed above. All dilutions were tested on two reagent lots and the reactivity compared.

Quantification: Ten doubling dilutions (commencing at 1:5 dilution) of the WHO International Standard- First WHO International Standard for anti-SARS-CoV-2 immunoglobulin (human) NIBSC code: 20/136 (NIBSC, Potters Bar, UK) were tested on one reagent lot to compare the assay's reported antibody concentration against the predicted binding antibody units (BAU)/mL of each dilution.

Cross-reactivity: A total of 55 plasma or serum samples known to contain potentially cross-reacting common analytes were tested in a single reagent lot, that was provided by the manufacturer (Table 2). A second set of 31 potentially cross-reacting specimens confirmed positive by PCR for Severe Acute Respiratory Syndrome coronavirus (SARS-CoV-1), Middle East Respiratory Syndrome coronavirus (MERS-CoV) and seasonal human coronavirus (HCoV-229E, HCoV-NL63 or HCoV-OC43) were tested in a single reagent lot, provided by the manufacturer. A summary of the analytes included in the cross-reaction panel is presented in **Table 2**. Samples

indicated as positive for an analyte indicates plasma samples from individuals with a past evidence of infection with the organism indicated, unless specifically indicated by IgM reactivity.

Interfering substances: A total of 35 plasma samples, known to contain potentially interfering substances, were tested in a single reagent lot. The interfering substance panel consisted of:

- Five visibly icteric samples;
- Five visibly haemolysed samples;
- Seven samples with visibly high levels of bilirubin;
- Five lipaemic samples;
- Five samples with anti-nuclear antibodies;
- Three samples positive for antibodies to double-stranded DNA (Lupus);
- Five samples positive for rheumatoid factor.

In addition, two positive samples, obtained from the sensitivity panel, were each diluted 1:2 in icteric, lipaemic and haemolysed samples (derived from the 35 interfering samples above). These six spiked samples were tested to detect interference to reactivity caused by interfering substances.

Reverse seroconversion panels: The reverse seroconversion panel was comprised of 47 plasma samples, taken at varying intervals commencing 18 days or later after onset of symptoms, from 10 different individuals. The purpose of this panel is to demonstrate the decline in antibody titre over time, in particular the IgM response.

Table 2: Cross-reacting panel comprised of 55 samples containing common cross-reacting analytes and a subset of 31 samples containing SARS-CoV-1, MERS-CoV or seasonal hCoV antibodies.

Analyte	Number of Samples
CMV IgM positive	4
EBV VCA IgM positive	2
Influenza A positive	3
Influenza B positive	3
Hepatitis A IgM positive	1
Hepatitis B e antigen positive	3
Hepatitis B surface antigen	5
Hepatitis B surface antigen/Hepatitis B c IgM positive	1
Hepatitis B surface antigen/Hepatitis B c IgM/Hepatitis B e antigen positive	1
Hepatitis C virus antibody positive	4
HIV antibody positive	8
Malaria antibody positive	5
Mycoplasma IgM positive	1
Parainfluenza positive	1
Parvovirus IgM positive	2
Psittacosis positive	1
Rubella IgM positive	1
Syphilis positive	6
Toxoplasma IgM positive	3
Severe Acute Respiratory Syndrome coronavirus	18
Middle East Respiratory Syndrome coronavirus	4
Human seasonal coronavirus (HCoV-229E, HCoV-NL63 or HCoV-OC43)	9

Seroconversion panels: Consisted of a total of 60 plasma samples, collected from five different SARS-CoV-2 NAT positive individuals at regular intervals from early infection to approximately 8 weeks post symptoms. Results of testing were used to determine the number of days post onset of symptoms the test kit first detected reactivity.

Repeatability: For repeatability studies (within run precision), a selected antibody positive clinical sample or commercial anti-SARS-CoV-2 quality control (QC) sample was tested 30 times

in the same test run and results presented as the percentage coefficient of variation (%CV). The sample was selected as having a low-level reactivity on the assay under evaluation.

Reproducibility: For reproducibility studies, the same sample used in the repeatability study was tested 30 times across no fewer than five different runs, and results presented as %CV.

Testing

Rapid Device Tests: Testing was performed according to the test kit instructions for use (IFU) by one operator. The results were read by that operator and one additional reader. The second reader was blinded to the results reported by the first reader. The intensity of the test and control lines were graded according to the reading scale in **Table 3**. When consensus for the sample reading was not met a third, independent reader recorded their result and the eventual consensus (2 of 3 readings being the same) was used as the final result. A grading of 0, was considered a negative result. Any test result that was invalid according to the IFU was recorded. For RDTs invalid results were when the control line did not appear or invalid due to obviously defective test device or associated kit components. For EIAs, invalid test runs were when the kit controls failed the manufacturer's validation criteria.

Table 3. Result legend for subjectively read assays

Scoring index	Intensity reading scale
0	Non-reactive
1	Very weak
2	Weak
3	Medium to strong reactivity

Enzyme Immunoassays: Testing was performed according to the test kit IFU by the same operator. All samples were tested in singlicate. Valid test results were determined according to the validation criteria contained in the test kit IFU.

All results were recorded on hard copy results spreadsheet at the time of reading and manually transcribed into MicroSoft Excel. All transcriptions and calculations were double-checked by a second, independent person daily

Results

Invalid Test Rate:

The number and percentage of all invalid test results was recorded, as per the IFU for each test kit and lot number, summarised in **Table 4** and **5**.

The invalid rate ranged from 0.00% to 1.40%. A total of 13 of the 23 RDT that reported IgG and IgM had no invalid test results. Only Biogenix reported an invalid rate of greater than 1.00%. Of the 12 tests that reported a single IgG, total antibody or neutralizing antibody result, seven had no invalid results. MDGen had no defined invalid criteria in their IFU. BioRad experienced two invalid results traced to a particular plate washer, so subsequent testing was performed on a second plate washer and no further invalid results were detected. It is noted that there were no obvious defects or issues identified with the initial plate washer. No further invalid results were detected for this test. The other three test kits had less than 1.00% invalid test results.

Concordance with recent infection:

The results of testing samples positive for anti-SARS-CoV-2 are expressed as “concordance with recent infection” as the true reference result for IgM could not be accurately assessed. The concordance with recent infection and specificity are expressed with 95% confidence intervals range (95% CI). The results of testing the panel of samples on test kits that reported IgG and IgM are summarized in Table 4. These test kits were all RDTs. The results were evaluated for reactivity to IgG, IgM, or either IgG and/or IgM to SARS-CoV2.

Of the 23 RDTs that reported results for both IgG and IgM, AllTest G/M and Healgen reported 100% concordance with recent infection for IgG. A further three tests (Biocan, Deepblue and Singclean) reported less than 90.0% concordance with a recent infection for IgG. All other tests reported between 90.0 to 99.9% concordance with recent infection for IgG. None of the 23 RDT kits reported 100% concordance with recent infection for IgM, with 9/23 (39.1%) having between 90.0 to 99.9% concordance for IgM and the remaining 14/23 tests (60.9%) having less than 90.0% concordance with recent infection for IgM. Two kits, Lysun (55.8%) and Singclean (57.3%) reported the lowest concordance with recent infection for IgM.

Table 4. Invalid rate, concordance with recent infection (n=199) and specificity (n=300) of rapid test devices testing for both IgG and IgM. Concordance and specificity results are presented as a heat map, with shades of green representing results greater than 90%, shades of yellow representing results between 60 to 90% and orange representing results less than 60%.

Abbreviation	Invalid Rate	Percentage Concordance with Recent Infection [95% CI range]			Specificity [95% CI range]		
		IgG	IgM	IgG and/or IgM	IgG	IgM	IgG and IgM
AllTest G/M	0.00%	**100.0** [97.6 - 100]	**90.0** [84.7 - 93.6]	**100.0** [97.6 - 100]	**94.0** [90.5 - 96.3]	**84.0** [79.2 - 87.9]	**80.7** [75.6 - 84.9]
Artron	0.00%	**98.5** [95.3 - 99.6]	**97.0** [93.3 - 98.8]	**99.0** [96.0 - 99.8]	**96.7** [93.8 - 98.3]	**88.0** [83.7 - 91.3]	**85.7** [81.1 - 89.3]
Biocan	0.63%	**89.4** [84.1 - 3.2]	**74.4** [67.6 - 80.2]	**89.4** [84.1 - 93.2]	**96.7** [93.8 - 98.3]	**96.0** [92.9 - 97.8]	**94.0** [90.5 - 96.3]
Biogenix	1.40%	**94.5** [90.1 - 97.1]	**82.4** [76.2 - 87.3]	**96.0** [91.9 - 98.1]	**96.3** [93.3 - 98.1]	**82.3** [77.4 - 86.4]	**80.3** [75.3 - 84.6]
Biohit	0.00%	**92.0** [87.0 - 95.2]	**77.4** [70.8 - 82.9]	**96.0** [91.9 - 98.1]	**99.0** [96.9 - 99.7]	**93.3** [89.7 - 95.8]	**92.7** [89.0 -95.2]
BioMedomics	0.13%	**93.5** [88.8 - 96.3]	**79.4** [73.0 - 84.7]	**96.5** [92.6 - 98.5]	**97.3** [94.6 - 98.8]	**89.7** [85.5 - 92.8]	**87.7** [83.3 - 91.1]
Boson	0.13%	**99.5** [96.8 -100]	**83.9** [77.9 - 88.6]	**100.0** [97.6 -100]	**97.7** [95.0 - 99.0]	**77.0** [71.7 - 81.6]	**75.7** [70.3-80.3]
BTNX	0.00%	**94.0** [89.5 - 96.7]	**84.9** [79.0 - 89.4]	**97.5** [93.9 - 99.1]	**99.0** [96.9 - 99.7]	**97.3** [94.6 - 98.8]	**96.3** [93.3 - 98.1]
Core Tech	0.00%	**99.5** [96.8 - 100]	**80.9** [74.6 - 86.0]	**99.5** [96.8 - 100]	**93.7** [90.1 - 96.0]	**90.3** [86.3 - 93.3]	**85.3** [80.7 - 89.0]
Deepblue	0.13%	**77.4** [70.8 - 82.9]	**94.0** [89.5 - 96.7]	**96.5** [92.6 - 98.5]	**95.3** [92.1 - 97.3]	**70.0** [64.4 - 75.1]	**69.3** [63.7 - 74.4]
Dynamiker	0.25%	**99.0** [96.0 - 99.8]	**99.0** [96.0 - 99.8]	**99.5** [96.8 - 100]	**96.0** [92.9 - 97.8]	**34.3** [29.0 - 40.0]	**34.0** [28.7 - 39.7]
Getein	0.00%	**99.0** [96.0 - 99.8]	**79.4** [73.0 - 84.7]	**99.5** [96.8 - 100]	**81.0** [76.0 - 85.2]	**97.0** [94.2 - 98.5]	**79.3** [74.2 - 83.7]
Healgen	0.00%	**100.0** [97.6 - 100]	**95.5** [91.3 - 97.8]	**100.0** [97.6 - 100]	**93.7** [90.1 - 96.0]	**90.3** [86.3 - 93.3]	**88.7** [84.4 - 91.9]
Joysbio	0.00%	**99.0** [96.0 - 99.8]	**97.0** [93.2 - 98.8]	**99.5** [96.8 - 100]	**94.7** [91.3 - 96.8]	**96.0** [92.9 - 97.8]	**94.0** [90.5 - 96.3]
Lysun	0.00%	**99.5** [96.8 - 100]	**55.8** [48.6 - 62.7]	**99.5** [96.8 - 100]	**89.3** [85.1 - 92.5]	**71.3** [65.8 - 76.3]	**65.0** [59.3 - 70.3]
MPBio	0.76%	**99.5** [96.8 - 100]	**76.0** [69.2 - 81.5]	**99.5** [96.8 - 100]	**90.7** [86.7 - 93.6]	**83.7** [78.9 - 87.6]	**79.0** [73.9 - 83.4]
RightSign	0.00%	**95.0** [90.7 - 97.4]	**80.4** [74.1 - 85.5]	**97.0** [93.2 - 98.8]	**98.3** [95.9 - 99.4]	**99.0** [96.9 - 99.7]	**97.7** [95.0 - 99.0]
Sensing	0.13%	**99.5** [96.8 - 100]	**99.5** [96.8 - 100]	**100.0** [97.6 - 100]	**56.0** [50.2 - 61.7	**48.0** [42.2 - 53.8]	**32.0** [26.8 - 37.6]
Singclean	0.13%	**85.4** [79.6 - 89.9]	**57.3** [50.1 - 64.2]	**86.9** [81.3 - 91.1]	**99.0** [96.9 - 99.7]	**79.0** [73.9 - 83.4]	**78.7** [73.5 - 83.1]
Standard Q	0.00%	**97.5** [93.9 - 99.1]	**75.4** [68.7 - 81.1]	**98.0** [94.6 - 99.4]	**98.3** [95.9 - 99.4]	**98.7** [96.4 - 99.6]	**97.7** [95.0 - 99.0]
Sugentech	0.00%	**98.5** [95.3 - 99.6]	**97.0** [93.2 - 98.8]	**99.5** [96.8 - 100]	**97.7** [95.0 - 99.0]	**68.3** [62.7 - 73.5]	**66.7** [61.0 - 71.9]
Sure Status	0.13%	**96.5** [92.6 - 98.5]	**72.9** [66.0 - 78.8]	**96.5** [92.6 - 98.5]	**98.3** [95.9 - 99.4]	**95.7** 92.5 - 97.6]	**94.0** [90.5 - 96.3]
VivaDiag	0.00%	**95.5** [91.3 - 97.8]	**92.0** [87.0 - 95.2]	**98.5** [95.3 - 99.6]	**98.7** [96.4 - 99.6]	**92.0** [88.2 - 94.7]	**91.0** [87.0 - 93.9]

Specificity:

Twenty of the 23 RDTs (87.0%) that reported both IgG and IgM had a specificity of greater than 90.0% and three had less than 90.0%, for IgG testing, specifically Getein (81.0%), Lysun (89.3%) and Sensing (56.0%). Eleven of the 23 RDTs (47.8%) had a specificity of greater than 90.0% for IgM; nine of 23 (39.1%) had a specificity between 70.0 to 90.0% for IgM and three, Dynamiker (34.3%), Sensing (48.0%) and Sugentech (68.3%) had a specificity of less than 70.0%.

The concordance with recent infection and specificity of 12 test kits that reported single results for IgG only, total antibodies or neutralizing antibodies are presented in **Table 5**. AllTest G, OmniPath and Wantai reported 100% concordance with recent infection for IgG and total antibodies respectively. Three of the 12 tests (25.0%) reported less than 80% concordance with recent infection, being MDGen (60.1%), Omega (64.3%) and Serion (78.9%), all testing for IgG. Of the 12 tests reporting a single result for IgG or total antibodies, only MDGen reported 100% specificity. Epitope (74.0%) had the lowest specificity of the 12 test kits.

Table 5. Invalid rate, concordance with recent infection and specificity of rapid test devices testing that report a single result for IgG only, total antibodies or neutralizing antibodies presented as a heat map, with shades of green representing results greater than 90%, shades of yellow representing results between 60 to 90% and orange representing results less than 60%.

Abbreviation	Invalid Rate	Concordance with Recent Infection [95% CI range]		Specificity [95% CI range]	
		IgG	Total or Nt	IgG	Total or Nt
AllTest G	0.00%	**100.0** [97.6 - 100]	NA	**87.7** [83.3 - 91.1]	NA
bioLytical	0.50%	NA	**98.0** [94.6 - 99.4]	NA	**88.3** [84.0 - 91.6]
BioRad	0.00%	NA	**86.4** [80.7 - 90.7]	NA	**97.7** [95.0 - 99.0]
Epitope	0.25%	**92.5** [87.6 - 95.6]	NA	**74.0** [68.6 - 78.8]	NA
MDGen	NA	**60.1** [53.6 - 67.6]	NA	**100.0** [98.4 - 100]	NA
Omega	0.00%	**64.3** [57.2 - 70.9]	NA	**85.3** [80.7 - 89.0]	NA
OmniPath	0.00%	**100.0** [97.6 - 100]	NA	**98.0** [95.5 - 99.20]	NA
Serion	0.00%	**78.9** [72.4 - 84.2]	NA	**99.3** [97.3 - 99.9]	NA
Vazyme	0.00%	NA	**89.4** [84.1 - 93.2]	NA	**98.3** [95.9 - 99.4].
Vircell	0.00%	**93.5** [88.8 - 96.3]	NA	**92.0** [88.2 - 94.7]	NA
Wantai	0.00%	NA	**100.0** [97.64 - 100]	NA	**99.7** [97.9 - 100]
Wondfo	0.40%	NA	**98.5** [95.3 - 99.6]	NA	**97.0** [94.2 - 98.5]

NA Not assessed
Total Total antibodies
Nt Neutralizing antibodies

Analytical Sensitivity/Lot-to-lot Variation:

Three samples (COVID461, COVID491 and COVID492) were doubling diluted from 1:2 to 1:1024. The serial doubling dilution samples were tested in two reagent lots of each of the 35 tests. Results of testing doubling dilution series are presented in supplemental Table S1. There was a large range of analytical sensitivity reported by the different test kits. In several test kits, a non-reactive result was followed by a reactive result on a higher dilution, making interpretation of results difficult. There were two test kits where an analytical sensitivity could not be determined. These were Singclean for IgM reactivity on COVID491 and COVID492 dilution series and Deepblue IgG and IgM reactivity for COVID491 when tested on Lot A only. The results of these two tests did not demonstrate a patten of reactivity that allowed a confident analysis of results, as it was impossible to determine if results were truly or falsely reactivity.

All 35 test kits detected COVID461 at 1:2 for both IgG and IgM, and COVID491 and COVID492 for IgG. However, some test kits did not detect IgM for COVID491 and/or COVID492 at 1:2. Dynamiker reported reactive IgM results for all 10 dilutions for all three dilutions series, except for COVID491 tested on lot A, where the 1:1024 dilution was non-reactive. The Wantai kit reported reactive total antibodies for all 10 dilutions for COVID461.

Four of the 35 tests reported a difference in reactivity of two or more doubling dilution when the same sample was tested in two different lots. Lysun reported a fourfold difference in reactivity for both IgG and IgM for COVID461 when tested on two different lots. Singclean reported an IgM analytical sensitivity of 1:8 for COVID461 when tested in lot A but 1:256 for lot B. Sugentech and Biocan also reported a different in IgM reactivity for COVID461 when tested on different lots. Deepblue reported a greater than twofold difference for both IgG and IgM for COVID 461 and 492 when tested on different lot numbers and was unable to be evaluated for COVID491.

Quantification:

Only the OmniPath test reported quantitative results in ug/mL units. A dilutions series of the First WHO International Standard for Anti-SARS-CoV-2 immunoglobulin (human) was tested on one reagent lot. The lowest dilution to give a reactive result was a 1:80 dilution of the standard, representing 12.5 BAU/mL or 1.36 ug/mL.

Cross-reactivity and Interference:

The summary results of testing 55 samples having potentially common cross-reacting analytes, 35 samples with interfering substances and 31 samples from individuals with known past infection with SARS-CoV-1, MERS-CoV and seasonal human coronavirus (HCoV-229E, HCoV-NL63 or HCoV-OC43) are presented Table 6, with the complete set of results in supplemental **Table S2**.

There was a broad range of the number of false reactive results for the cross reacting, interfering and non-SARS-CoV-2 coronavirus samples. Only five tests (MDGen, OnmiPath, Serion, Standard Q and Wantai) reported no false reactivity for the 55 samples with potentially cross-reacting substances. There were four, seven and two test kits that reported 10 or more false reactive results, respectively when testing for IgG, IgM or total antibody in the cross-reacting sample panels. The three samples with potentially cross-reacting analytes that had the highest number of false reacting results were samples from individuals with recent infection to cytomegalovirus, *C. psittaci* and parvovirus B19 with 17, 5 and 13 of the 35 tests reporting false IgM reactive results respectively.

Of the 35 samples with potential interfering substances, the five samples containing rheumatoid factor were falsely reactive on the most tests. For each of the five rheumatoid positive samples, 12, 10, 8, 11, and 13 tests reporting at least one reactive result. Of note, Dynamiker reported false IgM reactivity for 47 of 55 (85.5%) cross reacting samples and 26 of 35 (74.3%) samples having interfering substances. There were six tests (Bio Hit, MDGen, OmniPath, Serion, Sure Status and Wondfo) that reported no false reactive results for the 35 samples having interfering substances.

Table 6. The number of reactive results of testing samples from individuals having potentially cross-reacting, interfering or known past infection with SARS-CoV-1, MERS-CoV and seasonal human coronavirus (HCoV-229E, HCoV-NL63 or HCoV-OC43). A heat map presents test kits having fewer than 10 false reactive results for each population highlighted in shades of green. Those with between 10 and 30 false reactive results are highlighted in shades of yellow and those with great than 30 false reactive results highlighted in orange.

Abbreviation	Cross reactivity (n=55)			Interference (n=35)			Other Coronaviruses* (n=31)			SARS-CoV-1 (n=18)		
	IgG	IgM	Total	IgG	IgM	Total	IgG	IgM	Total	IgG	IgM	Total
Alltest G	4	NA	NA	3	NA	NA	14	NA	NA	12	NA	NA
AllTest G/M	3	5	NA	1	3	NA	2	3	NA	6	2	NA
Artron	1	7	NA	0	2	NA	10	1	NA	9	0	NA
Biocan	0	1	NA	0	1	NA	4	2	NA	3	1	NA
Biogenix	0	9	NA	3	7	NA	3	3	NA	2	1	NA
Biohit	0	2	NA	0	0	NA	14	3	NA	13	2	NA
bioLytical	NA	NA	11	NA	NA	5	NA	NA	12	NA	NA	10
BioMedomics	0	7	NA	2	9	NA	12	1	NA	12	0	NA
BioRad	NA	NA	2	NA	NA	1	NA	NA	15	NA	NA	14
Boson	1	8	NA	1	3	NA	16	3	NA	14	1	NA
BTNX	0	4	NA	3	7	NA	14	0	NA	14	0	NA
Core Tech	0	3	NA	2	0	NA	10	1	NA	9	0	NA
Deepblue	1	10	NA	7	8	NA	3	11	NA	1	6	NA
Dynamiker	3	47	NA	2	26	NA	11	17	NA	10	8	NA
Epitope	12	NA	NA	8	NA	NA	12	NA	NA	10	NA	NA

Abbreviation	Cross reactivity (n=55)			Interference (n=35)			Other Coronaviruses* (n=31)			SARS-CoV-1 (n=18)		
	IgG	IgM	Total	IgG	IgM	Total	IgG	IgM	Total	IgG	IgM	Total
Getein	10	2	NA	12	1	NA	18	0	NA	18	0	NA
Healgen	1	7	NA	2	3	NA	17	6	NA	17	5	NA
Joysbio	5	4	NA	10	9	NA	16	10	NA	13	8	NA
Lysun	10	17	NA	7	15	NA	15	7	NA	11	3	NA
MDGen	NA	NA	0	NA	NA	0	NA	NA	0	NA	NA	0
MPBio	6	12	NA	5	7	NA	17	3	NA	14	0	NA
Omega	NA	NA	13	NA	NA	3	NA	NA	10	NA	NA	7
OmniPath	0	NA	NA	0	NA	NA	17	NA	NA	NA	NA	16
RightSign	0	1	NA	1	0	NA	19	1	NA	15	0	NA
Sensing	25	27	NA	21	19	NA	19	13	NA	17	9	NA
Serion	0	NA	NA	0	NA	NA	17	NA	NA	16	NA	NA
Singclean	0	29	NA	2	17	NA	4	22	NA	4	15	NA
Standard Q	0	0	NA	0	5	NA	9	1	NA	8	0	NA
Sugentech	2	19	NA	3	10	NA	1	4	NA	5	3	NA
Sure Status	0	3	NA	0	0	NA	6	2	NA	6	2	NA
Vazyme	NA	NA	1	NA	NA	2	NA	NA	17	NA	NA	16
Vircell	7	NA	NA	3	NA	NA	17	NA	NA	10	NA	NA
VivaDiag	0	7	NA	0	1	NA	6	2	NA	6	1	NA
Wantai	NA	NA	0	NA	NA	1	NA	NA	0	NA	NA	18
Wondfo	NA	NA	1	NA	NA	0	NA	NA	14	NA	NA	9

* Includes samples from individuals having past MERS and seasonal coronavirus infections; excludes SARS-CoV-1 samples

NA = Not applicable

There were 18 samples obtained from individuals previously infected with SARS-CoV-1. In addition to the 18 SARS-CoV-1 samples, the panel included samples from four individuals previously infected with MERS-CoV and nine individuals previously infected with seasonal coronaviruses. Only MDGen did not report any false reactive results for the 31 samples obtained from individuals having past non-SARS-CoV-2 coronavirus infections, therefore being the only test kit that had no false reactivity across the cross reacting, interfering and non-SARS-CoV-2 panels. A total of 29 test kits (82.9%) reported more than 10 of the 31 non-SARS-CoV-2 panels samples as being reactive, indicating cross-reactivity with other coronavirus infections. There were many false reactive test results across a broad range of test kits for both IgG and IgM reactivity when testing the 18 SARS-CoV-1 samples. A total of 21 of the 35 tests that report either IgG or total antibodies reported 10 or more reactive results of the 18 SARS-CoV-1 samples. In contrast, only one test kit that reported IgM results (Singclean) that reported more than 10 of the 18 samples reactive.

Reverse seroconversion panels:

The reverse seroconversion panel, comprised of 47 plasma samples, from 10 different symptomatic individuals, taken at varying intervals commencing 18 days post onset of symptoms or later. These samples were tested on each test kit under evaluation, except for the seven IgG only tests. The majority of test kits reported reactive results for the analyte (IgG/IgM/Total/Nt) for all serial bleed samples (Supplemental Table S3). The purpose of the reverse seroconversion panels was to determine the analytical sensitivity of tests detecting IgM, as it was assumed that the level of IgM would decrease to undetectable levels over time. This was the case for several test kits. The IgM results from BTNX, DeepBlue, Dynamiker, Healgen, Lysun, Sensing, Standard Q, Sugentech and VivaDiag all reported one or more negative IgM results following reactive results from samples collected earlier. For one five-member series, four test kits (BioHit, Biomedomics, Getien and RightSign) did not detect IgM in any of the samples, whereas all other test kits reported a reactive result for one or more samples of that series.

Seroconversion panels:

A total of 60 plasma samples, collected from five different SARS-CoV-2 NAT positive individuals at regular intervals from early infection, were tested in each test kit. The results are summarized in Supplemental Table S4. The samples were taken from individuals exposed to SARS-CoV-2 and subsequently tested positive by nucleic acid testing. Samples were drawn from 8 days prior to up to 52 days post start of symptoms. The majority of test kits detected IgG and IgM antibodies to SARS-CoV-2 within the first three bleeds. Generally, the IgM response was earlier or at the same time as the IgG response. There were some notable exceptions. MDGen, testing for IgG only, failed to detect antibodies in one patient's series of nine samples and reported five negative, six equivocal and one reactive in a second patient's series of 14 samples. Serion, also testing for IgG only, reported negative results for the first five of a series of nine samples. The IgM response of both RightSign and Standard Q decreased to undetectable levels in the same two of five seroconversion panels.

Repeatability and Reproducibility:

Repeatability (within run precision) and reproducibility (between run precision) studies were conducted on six EIA test kits that reported quantitative results. The results were expressed as CV% and are summarized in **Table 7**. Repeatability ranged from 3.70% to 11.37% and reproducibility ranges from 3.52% to 13.42%.

Table 7. Repeatability and reproducibility results, expressed as percentage coefficient of variation (%CV) of enzyme immunoassays reporting a quantitative result.

Abbreviation	Repeatability (%CV)	Reproducibility (%CV)
BioRad	7.80	10.40
Epitope	5.98	6.45
MDGen	11.37	4.69
Omega	6.26	13.42
OmniPath	3.50	5.60
Serion	4.79	11.99
Vazyme	3.70	3.52
Vircell	7.38	15.29
Wantai	8.82	15.50

Discussion

Manufactures of diagnostic test kits responded quickly to the SARS-CoV-2 pandemic. Within six months, numerous tests were available for use, with rapid tests and EIAs for the detection of antibodies to SARS-CoV-2 being some of the first to become available, as they are relatively easy to design and manufacture. Whereas IVDs for infectious diseases usually take several years from inception to product registration and release, serology assays for SARS-CoV-2 were released months after the declaration of the pandemic. At that time, the pathogenesis and the immune response to infection, and an understanding of the utility of serology testing and what testing algorithms may be useful, were unknown. In response to the urgent need for diagnostic test kits, regulators world-wide registered novel tests as Emergency Use Authorization, requiring manufacturers to provide only limited evidence of test kit performance. This situation removed a significant barrier to manufacturers, resulting in many test kits becoming available. The situation also diminished the understanding of the performance and potential use of the different test kits available to laboratories.

At the same time, numerous studies comparing the performance of test kits were published (2, 3, 6). To facilitate rapid information transfer, medical journals published pre-print articles which were not subject to rigorous peer review (2). However, many of these studies were poorly designed, used remnant, un-characterized clinical samples, had limited statistical significance and inappropriate analysis and interpretation. Early in the pandemic, the utility of SARS-CoV-2 serology testing was unknown (9, 10). It was known that infected individuals produced an antibody response to both nucleocapsid and spike proteins, however the nature, strength and

duration of the immune response was poorly understood, nor was there sufficient clarity on how these novel tests could be used in series or parallel to maximize sensitivity (here reported as concordance with recent infection) and specificity. Furthermore, the characteristics of the populations being tested (e.g. severity of illness) and the timing of testing (eg. days since onset of symptoms) being tested were variable and/or not described in sufficient detail, leading inevitably to an inability to determine accurately performance and appropriate use of the tests. This led to conflicting and confusing information that could not easily be used to inform appropriate use of serological tests.

To address this situation, NRL developed a protocol and collected relevant samples to conduct a comprehensive, head-to-head comparison of multiple serological assays. The NRL, authorized WHO Pre-qualification Evaluation laboratory, was contracted by WHO to conduct performance evaluations on selected SARS-CoV-2 serology tests, using the developed protocol and sample set, to assess up to 40 unique serological tests that were suitable for use in LIMCs, including RDTs and EIAs. WHO released an expression of interest to manufacturers to participate in a comparative assessment of serological tests (5) and the selected test kits were evaluated by NRL using the protocol as outlines.

Samples were acquired in sufficient volumes from a range of reliable sources and in sufficient volume to evaluate. Samples from a population of individuals most likely to be targeted for serological testing, specifically those with mild-moderate SARS-CoV2 infection, confirmed by PCR. Plasma samples were drawn several weeks after onset of symptoms to ensure sufficient time for the development of an immune response. The study aimed to provide a better understanding the comparative performance and potential uses for these tests.

Utility of Serological Testing: Generally, serological testing for infectious diseases is used for several clinical purposes. Detection of pathogen-specific IgM in a single sample can assist in the diagnosis of acute disease, however this approach is not suitable for all pathogens (infections). The seroconversion of IgG from undetectable to detectable levels in paired acute and convalescent samples can diagnose acute infections retrospectively. The detection of organism-specific IgG can be used as an indicator of past contact with an organism, either through wild type infection or vaccination. Often in clinical serology, a testing algorithm is employed to maximize the sensitivity and specificity of diagnosis. For example, in HIV or HCV serology, reactivity in a screening assay is confirmed through supplemental and/or confirmatory testing. Some organisms elicit multiple antigens. The detection of an antibody response to each of these antigens can be used to determine the stage of disease. As an example, serology testing for anti-EBV VCA IgG and IgM, anti-EBNA IgG and occasionally anti-EBV EA IgG are used to diagnose acute

or convalescent EBV infection in a single bleed. Finally, as specified by the WHO, serology can be used for retrospectively determine the size of an outbreak or extent of infection in a population under study (11). Seroprevalence studies use serology to determine the proportion of a population that has evidence of immunity and to identify segments of the population requiring targeted vaccination.

SARS-CoV-2 IgM Response: Early in the pandemic, SARS-CoV-2 serology assays that detected IgM were used to diagnose recent infection (12, 13). This evaluation assessed whether the IgM reactivity of the test kits were due to infection or cross reaction from other biological substances. The concordance of IgM assays compared with NAT-confirmed recent infection, ranged from 55.8 to 99.5%, with 14 of the 23 (60.9 %) test kits having concordance with recent infection of less than 90.0%. Specificity of IgM assays ranged from 34.3 to 99.0%, with the 12 of 23 (52.2 %) test kits having a specificity of less than 90.0%. Five of the 23 test kits (27.3%) that reported IgM results reported more than 30 of the 90 cross reacting and interfering substance samples as falsely reactive. Of note, one test (Dynamiker) reported 47/55 (85.5%) and 26/35 (74.3%) reactive results for cross-reacting and interfering substance samples, respectively. The IgM assays reported reactive results for seroconversion panel samples ranging from the initial sample generally obtained within the first week of symptoms, to sample three taken within the second week. Results of the reverse seroconversion panels, comprised of serial bleeds collected several weeks after initial infection, identified some decrease in IgM reactivity. However, this decrease was not universally seen in the samples collected, indicating that the decline in IgM response varies from individual to individual. Given there was a broad range of results across the concordance with recent infection panel and the seroconversion and reverse seroconversion panels, no definitive general conclusion can be made regarding SARS-CoV-2 IgM testing except that an IgM response was universal detected across all infected individuals and there is some evidence that the IgM response decreases over time to undetectable levels.

The utility of IgM testing for SARS-CoV-2 is debatable. Assuming the test kits demonstrate a high level of accuracy, and low levels of false reactivity due to cross reacting and interfering substances, a reactive IgM response may be used to ascertain a recent infection in an individual where the NAT or rapid antigen test were unable to be collected during the viral excretion phase. However, the use of IgM testing as a method of primary diagnosis is not recommended as the reactivity is delayed compared with NAT or RAT and does not identify if the individual is infectious.

SARS-CoV-2 IgG Response. All tests evaluated detected SARS-CoV-2 IgG either independently (IgG only), in association with the detection of IgM (IgG/IgM) or as a total antibody (IgG, IgM, IgA) or

neutralizing antibody test. Of the 23 test kits that reported both IgG and IgM, the concordance with recent infection of the IgG response ranged from 77.4 to 100% and the specificity ranged from 56.0% to 99.0%. Only three tests (Biocan, Deepblue and Singclean) reported concordance with recent infection of less than 90.0% and three tests (Getein, Lysun and Singsen) reporting a specificity of less than 90.0%, indicating a relatively high level of concordance with recent infection and specificity of IgG testing for these IgG and IgM tests. Of the 12 tests that reported either IgG only, total antibodies or neutralizing antibodies, the concordance with recent infection and the specificity ranges were 60.1 to 100% and 74.0 to 100% respectively. Three tests, AllTest G (IgG only), OmniPath (IgG only) and Wantai (Total antibodies) reported 100% concordance with recent infection. MDGen (IgG only) reported 100% specificity. A total of eight of 23 IgG/IgM tests reported both concordance with recent infection and specificity IgG results of greater than 95%. However, most of these tests had lower relative concordance with recent infection and specificity for IgM results compared with other similar tests, underlining the importance of selecting tests depending on the intended use. Three of the 12 IgG only, total antibodies and neutralizing antibody tests (OmniPath, Wantai and Wondfo) reported both concordance with recent infection and specificity of greater than 95%.

Use in Seroprevalence Studies: Seroprevalence studies can be useful to assess the extent of exposure to a virus in a defined population. Following the detection of SARS-CoV-2, the WHO has quickly published standardized protocols for early outbreak investigations under the banner of the "Unity Studies". Among the Unity Studies protocols, a global sero-epidemiological standardization protocol was published, which aimed at increasing the evidence-based knowledge for action (14). Within the protocol, it is stated that measurement of IgG or total antibodies using EIA – demonstrated to have on average higher accuracy than RDTs in a systematic review (2) - are the preferred method to assess seroprevalence (15). The evaluation of serology tests reported here included four assays that reported total antibodies (bioLytica, BioRad, Wantai and Wondfo) and 6 that reported IgG (Epitope Diagnostics, MDGen, Omega, Omnipath, Serion, Vircell). The Wantai assay reported a concordance with recent infection of 100% and a specificity of 99.7% indicating it would be a suitable assay for seroprevalence studies, meeting the desirable criteria for tests for prior infection for moderate, high volume needs (≥98% sensitivity and ≥99% specificity) outlined in the Target Product Profiles (TPP) for priority diagnostics to support response to the COVID-19 pandemic (16). Wondfo would also be suitable, as it met acceptable criteria outlined in the TPP (≥95% sensitivity and ≥97% specificity) reporting high concordance with recent infection (98.5%) and specificity (97.0%). The other two assays measuring total antibodies reported lower performance (Table 5). Of the EIAs that reported IgG, only OmniPath would be suitable, as it reported a high concordance with infection

(100%) and specificity (98%), therefore meeting acceptable criteria outlined in the TPP. Of note, this does not constitute a recommendation by the WHO for these test kits. The selection of test kits must be assessed by the procurer to ensure the performance meets the intended use.

When testing samples with potential interfering and cross-reacting substances, Wantai reported only one false reactive result from the 35 interfering substances and none from the 55 cross reacting samples, whereas Wondfo reported none and one false reactive result, respectively. bioLytica and BioRad reported 5/35 and 11/55 and 1/35 and 2/55 respectively. The results of this study indicate that the Wantai or Wondfo assays are suited for use in seroprevalence studies, however this does not constitute a recommendation by the WHO for these test kits. The selection of test kits must be assessed by the procurer to ensure the performance meets the intended use.

Precision: There were seven tests that reported results with a quantitative basis, therefore allowing the estimation of imprecision. The repeatability of testing of these tests ranged from 3.70 to 11.37 %CV, whereas the reproducibility ranged from 3.52 to 15.50 %CV. Vazyme reported the lowest repeatability and reproducibility results, however it must be noted that this test detects neutralizing antibodies as distinct from binding antibodies. The results should be interpreted accordingly. The imprecision of test kits that report a quantitative value should continually be monitored using a well-developed quality control program. Generally, the %CV reported in this study were in line with EIAs for other infectious diseases.

The results of this study indicate that the SARS-CoV-2 IgG response is detectable at the same time or one bleed after the detection of IgM. However, this information is of limited clinical use, as seroconversion is unlikely to be used as a means of confirming a recent infection. Most reverse seroconversion samples, obtained from individuals at least three weeks post onset of symptoms where reactive for IgG. The results of the seroconversion and reverse seroconversion panels indicate that there was little evidence that the IgG response became undetectable within seven weeks post infection.

This study demonstrated a wide range of performance characteristics associated with the test kits evaluated (Table 8). There were some tests that demonstrated unacceptably poor concordance with recent infection and specificity, others that reported unacceptable high levels of false reactivity when testing samples containing cross reactivity and interfering substances and antibodies from other coronaviruses. Generally, all tests detected IgM and/or IgG around the same time in seroconversion panels, being reactive within the first week after the onset of symptoms. Although a limited number of test kit lot numbers were evaluated, several tests

demonstrated a more than two-fold difference in analytical sensitivity between the two lot numbers tested. This information highlights the possibility of lot-to-lot variability, and therefore the necessity for monitoring the performance of test kits in the field using lot release testing and/or quality assurance programs such as external quality assessment schemes or quality control programs.

Table 8. Summary of study findings and the potential use of SARS-CoV-2 serological assays.

Study Panel	Study Findings
Sensitivity Panel	• Significant variation between tests for both IgG and IgM reactivity • All individuals produced IgG and IgM response • Poor sensitivity leads to increased false negative results
Specificity Panel	• Significant variation between tests for both IgG and IgM reactivity • Poor specificity leads to increased false reactive results
Dilution Series, Lot-to-lot	• Tests demonstrated variable analytical sensitivity for both IgG and IgM • Several tests demonstrated greater than two-fold dilution difference between lots, indicating lot to lot variation • Random false reactivity was detected in some tests
Cross Reaction Panel	• All tests reporting IgM demonstrated false reactivity with samples containing potentially cross-reacting substances. • Most tests reporting IgG demonstrated cross reactivity • Some tests had false reactivity with a high proportion of cross-reacting samples • Knowledge of cross reactivity is an important performance characteristic to understand prior to test kit section
Interference Panel	• Almost all test kits demonstrated false reactivity with samples having interfering substances • Many tests reported IgG and/or IgM false reactivity for samples containing rheumatoid factor • Various interfering analytes caused false reactivity
Seroconversion Panel	• IgM reactivity was detected within the first bleeds in all five individuals • Some tests demonstrated delayed IgM reactivity indicting IgM testing is not appropriate for diagnosis, but useful to demonstrate recent infection (if unable to confirm using NAT or rapid antigen testing) • All individuals produced an IgG response, indicating IgG testing may be useful in determining exposure to infection or vaccination
Reverse Seroconversion Panel	• All individual demonstrated both IgG and IgM response • Some serial bleeds demonstrated a decline in IgM reactivity over time • Most tests were reactive for IgG and IgM for all serial bleeds
Precision	• Tests with a quantitative value demonstrated precision similar to other EIAs. • Precision of assays should be monitored over time using a quality control program

When selecting test kits, the purpose of the test kits must be determined, and relevant performance characteristics evaluated prior to use. Each component of the test kits (IgG or IgM) should be evaluated for sensitivity and specificity, noting that generally, highly sensitive tests have poorer specificity and visa versa. Knowledge of potential cross reactivity and interference of assay is important when choosing tests for use. As demonstrated in this study, there was significant variability in these performance characteristics, which would otherwise gone unrecognized without a systematic, robust evaluation. Tests with poorer performance for certain characteristics may not be suitable for some purposes, or the results interpreted with this poor performance in mind. For example, an assay with lower sensitivity but higher specificity may be useful in a confirmatory algorithm. Selection of tests is also dependent on aspects other than performance, such as accessibility, cost, ease of use. The findings of this study support this selection.

All testing should be quality assured. Tests should be performed as per the manufacturers' IFU, processes documented, staff trained and assessed as competent. Patient sample traceability must be ensured. Equipment must be well maintained and calibrated and records of all parts of the process documented and retained. All test sites should participate in an external quality assessment scheme and have a well-designed quality control programs. It is important that lot-to-lot variation is monitored, either through quality control and/or batch release testing. The accuracy of test results impact on patient's health and well-being and must be maintained at the highest level.

Limitations

This study was designed as a comprehensive head-to-head evaluation of SARS-CoV-2 serology test kits using the same panel of samples for all test kits. The numbers of samples used were statistically significant and the protocol developed using processes commonly used by regulators and WHO Prequalification group. However, several limitations of the study should be noted.

The study did not assess safety, usability, cost or test kit robustness. Although the IFU of each test kit was reviewed, a systematic review of the content against specific regulations was not undertaken. Each of these aspects are important when selecting the test kit for specific use. To ensure sufficient volume of sample, the study predominantly used citrated plasma collected using plasmapheresis. It is noted that although almost all test kit nominated plasma as a validated sample type, few, if any specified the anticoagulant used. Therefore, it is possible that

citrated plasma may interfere with the results reported. Some of the samples used in the cross reacting and interfering substance panels had limited clinical information.

The study used SARS-CoV-2 positive samples collected from induvial early in the pandemic. Therefore, all positive samples were from individuals infected with the ancestral variant. Further studies evaluating test kits using samples positive due to Delta and Omicron and other variants may be useful. Also, the study does not evaluate the test kits using samples obtained from individuals vaccinated with various vaccines or numbers of doses, with or without past history of infection.

The study was designed to evaluate the ability of the test kit to detect antibodies post infection. To this end, samples were collected from individuals at least 14 days after the onset of symptoms. Therefore, the only samples collected from infected individuals within 14 days of infection were in the seroconversion panel, so an assessment of the test kit testing samples during very early infection was limited.

Conclusion

The results in this report present a one-time independent assessment of the performance characteristics of products from two reagent lots. The assessment provides the manufacturer independent evidence of the test kit performance. Similar performance testing of other test kits, using the same panel of samples, protocol and analysis will be published, allowing the manufacturer and users to compare results.

It should be noted that when comparing results, the different formats and uses of the test kits should be taken into consideration. Each test kit will detect different antibody isotypes to different antigens and have varying sensitivity and specificity. Some tests are expected to detect antibody responses early whereas others may be reactive later in disease. Knowledge of the performance characteristics will allow users to determine whether the test kit is fit for their intended purpose. Therefore, differences in test kit performance can be beneficial attributes.

Acknowledgements

We thank scientific staff at the National Serology Reference Laboratory (Australia): Sadaf Mohiuddin, Jingjing Cai, Bethmi Liyanage and all technical support staff who took part in composing the panels and performing testing. In particular, we acknowledge Technopath Clinical Diagnostics (Ireland) for their generous donation of positive donor samples including the seroconversion samples. Other panel samples were obtained commercially from Boca Biolisitics (United States of America); BioMex (Germany); Medical Research Networx (United States) and

Seracare (United States). We also acknowledge Duke-NUS Medical School (Singapore), Erasmus Medical Center (Netherlands), Tan Tock Seng Hospital (Singapore) and International Vaccine Institute (IVI) for contributing samples from individuals previously infected with SARS-CoV-1, MERS-CoV and seasonal human coronaviruses.

Funding

This study was funded by WHO.

Declaration of interests

The authors declare no competing interests.

References

1. Andersson M, Low N, French N, Greenhalgh T, Jeffery K, Brent A, Ball J, Pollock A, McCoy D, Iturriza-Gomara M, Buchan I, Salisbury H, Pillay D, Irving W. 2020. Rapid roll out of SARS-CoV-2 antibody testing-a concern. BMJ 369:m2420.

2. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, Adriano A, Beese S, Dretzke J, Ferrante di Ruffano L, Harris IM, Price MJ, Dittrich S, Emperador D, Hooft L, Leeflang MM, Van den Bruel A, Cochrane C-DTAG. 2020. Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane Database Syst Rev 6:CD013652.

3. van den Beld MJC, Murk JL, Kluytmans J, Koopmans MPG, Reimerink J, van Loo IHM, Wegdam-Blans MCA, Zaaijer H, Serology Workgroup for S-C, GeurtsvanKessel C, Reusken C. 2021. Increasing the Efficiency of a National Laboratory Response to COVID-19: a Nationwide Multicenter Evaluation of 47 Commercial SARS-CoV-2 Immunoassays by 41 Laboratories. J Clin Microbiol 59:e0076721.

4. World Health Organization. 2020. Advice on the use of point-of-care immunodiagnostic tests for COVID-19. WHO Headquarters (HQ), WHO, Geneva, Switzerland.

5. World Health Organization. 2020. Expression of Interest - General Information and Timelines, on WHO. https://www.who.int/docs/default-source/coronaviruse/eoi_general-info_timelines.pdf?sfvrsn=215a8b28_5. Accessed 30 May 2022.

6. Foundation for New and Innovative Diagnostics (FIND). 2020. FIND Evaluation of SARS-COV-2 Antibody (AB) Detection Tests, on FIND. https://www.finddx.org/sarscov2-eval-antibody/. Accessed 5 May 2022.

7. Van Elslande J, Oyaert M, Lorent N, Vande Weygaerde Y, Van Pottelbergh G, Godderis L, Van Ranst M, Andre E, Padalko E, Lagrou K, Vandendriessche S, Vermeersch P. 2022. Lower persistence of anti-nucleocapsid compared to anti-spike antibodies up to one year after SARS-CoV-2 infection. Diagn Microbiol Infect Dis 103:115659.

8. Recovery Collaborative Group. 2022. Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet 399:665-676.

9. Infantino M, Damiani A, Gobbi FL, Grossi V, Lari B, Macchia D, Casprini P, Veneziani F, Villalta D, Bizzaro N, Cappelletti P, Fabris M, Quartuccio L, Benucci M, Manfredi M. 2020. Serological Assays for SARS-CoV-2 Infectious Disease: Benefits, Limitations and Perspectives. Isr Med Assoc J 22:203-210.

10. Post N, Eddy D, Huntley C, van Schalkwyk MCI, Shrotri M, Leeman D, Rigby S, Williams SV, Bermingham WH, Kellam P, Maher J, Shields AM, Amirthalingam G, Peacock SJ, Ismail SA. 2020. Antibody response to SARS-CoV-2 infection in humans: A systematic review. PLoS One 15:e0244126.

11. World Health Organization Emergencies Preparedness. 31 December 2020 2020. Coronavirus disease (COVID-19): Serology, antibodies and immunity, on WHO. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-serology. Accessed 20 June 2020.

12. Lou B, Li TD, Zheng SF, Su YY, Li ZY, Liu W, Yu F, Ge SX, Zou QD, Yuan Q, Lin S, Hong CM, Yao XY, Zhang XJ, Wu DH, Zhou GL, Hou WH, Li TT, Zhang YL, Zhang SY, Fan J, Zhang J, Xia NS, Chen Y. 2020. Serology characteristics of SARS-CoV-2 infection after exposure and post-symptom onset. Eur Respir J 56.

13. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, Wang X, Yuan J, Li T, Li J, Qian S, Hong C, Wang F, Liu Y, Wang Z, He Q, Li Z, He B, Zhang T, Fu Y, Ge S, Liu L, Zhang J, Xia N, Zhang Z. 2020. Antibody Responses to SARS-CoV-2 in Patients With Novel Coronavirus Disease 2019. Clin Infect Dis 71:2027-2034.

14. World Health Organization Emergencies Preparedness. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigation Protocols, on WHO. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations. Accessed 20 June 2022.

15. World Health Organization Emergencies Preparedness. 2020. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection (version 2). WHO, Geneva, Switzerland. https://apps.who.int/iris/rest/bitstreams/1278667/retrieve.

16. World Health Organization. COVID-19 Target product profiles for priority diagnostics to support response to the COVID-19 pandemic v.1.0 (who.int) WHO, Geneva, Switzerland.