The Royal College of Pathologists
Pathology: the science behind the cure

Verification and validation methodology and sample sets for evaluation of assays for SARS-CoV-2 (COVID-19)

26 May 2020

Author: Professor William Egner, Chair of the RCPath Immunology Specialty Advisory Committee

This guidance is produced to assist NHS and research laboratories to evaluate immunoassays for clinical use and help to meet ISO accreditation standards for clinical use. It is not based on a review of the literature, but is an attempt to address a pressing need in the clinical community. It has been widely circulated for comment and improvement in the clinical COVID-19 testing community.

It is not intended to be exhaustive and suggestions for improvements are welcome for future iterations (please contact us at covid-19@rcpath.org should you have any comments). The sera-set compositions, experimental methodologies, website resources and data presentation suggestions are provided for guidance and will require local adaptation to be fit for your purposes. There may well be alternative compositions, methodologies and designs required for the purposes for which your results will be used.

The aim of harmonised verification and validation (V&V) datasets is to ensure comparable and adequately powered studies to evaluate assays. Such an approach also ensures that all centres attempt harmonised evaluations that address important performance metrics that are essential for quality assurance and patient safety.

Performance specifications for commercial assays are also published on the Medicines and Healthcare products Regulatory Agency (MHRA) website and may change with time: www.gov.uk/government/collections/mhra-guidance-on-coronavirus-covid-19#medical-devices

Principles of validation and verification

There are at least eight important principles common to V&V exercises, in addition to the specific performance characteristics of reproducibility, robustness, interferences and accuracy that are estimated in V&V. They are as follows.

1. The assay must be confirmed to be able to perform the clinical task required, in the population on which it will be used. This is achieved by specifying the clinical and technical acceptance criteria that are desired, before evaluation of the assay.

The Royal College of Pathologists
6 Alie Street
London E1 8QT
T: 020 7451 6700
F: 020 7451 6701
www.rcpath.org

Unique document reference number: G222-3

2. Similar sera-set compositions are essential for head-to-head comparisons of method performance where a single set of materials cannot be used on both assays. A single shared sera-set is likely to be impossible for validation of multiple immunoassays in situ in pathology

   and research laboratories UK-wide, and much of the panel will be locally sourced to a common specification. That commonality would enable shared V&V exercises to increase the translatability and generalisability of the observations across different laboratories.

3. Confidence intervals of the estimated performance should be stated in all evaluations. The 95% confidence intervals for methodologies may overlap considerably, meaning that there is unlikely to be a statistically significant difference between them, and the intervals may be very wide or skewed by the sera-set composition. The confidence intervals of sensitivity estimates widen markedly for lower prevalence in smaller datasets. Specificity tends to be more robust. The challenge is to identify the minimum dataset that will give 95% confidence interval limits that are small enough to meet performance acceptance criterion or to reliably differentiate performance levels of several tests.

4. Truly effective head-to-head comparisons of assay performance will require a very large sera-set for sufficient power to determine small differences between method sensitivities at relevant pre-test prevalences. Without very large datasets, few assays would be able to demonstrate 98% sensitivity, with the lower tail of the distribution greater than 95%. These large sera-sets may be very hard to source and a pragmatic approach to essential minimum datasets may be required, with rarer disease control sera being added retrospectively once the assay is in service. This document is not produced to dissuade anyone from performing the largest evaluation possible, but to illustrate a pragmatic approach where samples are limited, to encourage local laboratories to consider additional samples to examine issues that may not have been addressed, and to collaborate in assessing methodologies that have not undergone central Public Health England (PHE) evaluation like the large automated platforms.

5. The sera-set used should emulate the expected pre-test prevalence of expected positives. If possible, it should also replicate the expected prevalence of confounders/interfering substances/disease controls. This may not always be possible in initial assessment prior to assay acceptance and may require some additional retrospective evaluation as above. In low pre-test prevalence situations, a large negative sera-set is required to effectively assess the effect of false-positivity on predictive values. Estimates of pre-test prevalence of COVID-19 antibodies in different geographical locations and by different methods vary widely. An appropriate pre-test prevalence for your region and cohort is necessary.

6. Any perceived gaps in published evaluations can and should be examined as part of assay verification, even if full validation is not necessary.

7. Assay V&V should be performed consistently and to an agreed protocol. The essential components should be covered, although there are often a variety of valid experimental approaches to providing a particular component.

8. Evaluation of sample types must be included. Stability studies to include transport and storage are essential for adoption of near-patient testing methods like dried blood spot and capillary samples. Whole blood test comparisons will require volunteer donors meeting the sera-set criterion, unless aliquots of EDTA blood or plasma may be used instead.

Acceptable clinical quality requirements should be defined for each use scenario

By defining the clinical quality requirements to match the intended clinical use, the selection of an appropriate test performance to achieve the desired clinical outcome can be assured. Some potential examples are provided here.

Please note: There will be other use scenarios that may arise nationally and locally and each will require evaluation of the suitability of the assays.

Scenario 1: As a companion diagnostic in COVID-19

In hospitalised cases with a high pre-test prevalence, you already think the patient probably has COVID-19, but wish to avoid nosocomial harm by cohorting a non-COVID-19 pneumonia in the COVID-19 wards.

You might therefore aim for high sensitivity but accept a slightly lower specificity to result in a appropriately low post-test likelihood that COVID-19 disease is present. It may be that an antibody test alone may not be able to give the performance required, but a combination of assays or serial testing may be able to do so. You might alter your performance expectations to fit a different use context. Alternatively, you may decide the test cannot deliver the required performance or that RT-PCR alone is sufficient.

Clinical patient pathways that involve sequential or combinatorial use of several test results (RT-PCR, Nucleocapsid, Spike/RBD antibody assays) will produce different post-testing likelihood ratios that will influence the acceptable performance parameters for each component

Scenario 2: Health care worker screening/patient cohorting for resumption of treatment

You may want to exclude nosocomial infection imported to a workplace or avoid treating a patient who is currently infected with immunosuppressive treatments or surgery.

You may then prioritise sensitivity or specificity in a low pre-test prevalence population depending on a risk assessment about the consequences of false positivity or false negativity. Since there is currently no proven link between antibody and protection then the purpose of a positive test would only be to demonstrate contact with the virus in the recent past and accompanying seroconversion (probably at least two weeks before the test). Alternatively you want to monitor seroprevalence and seroincidence in your target population.

You may choose a patient pathway with serial or concurrent testing including RT-PCR to exclude asymptomatic and active infection, but wish to combine this with serology to exclude past infection or seroconversion between hospital visits or to monitor acquisition of infection at work. You may consider serial or sequential testing and other non-test-based interventions like pre-procedure quarantining. Each use case requires you to define the acceptable performance characteristic of the component assays.

Scenario 3: Community/population sero-prevalence studies

In this scenario, you may simply wish to estimate seroprevalence and/or seroincidence in a population and to balance sensitivity and specificity at an acceptable level.

The required performance characteristics may vary from centre to centre and in different pre-test prevalences.

Should it become established that an assay correlates well with sero-protection from infection, then you may wish to optimise specificity over sensitivity, since you do not wish to imply that a false-positive patient has protection. As of May 2020, there is no consensus that antibody detection in any assay predicts protection and it is not to be used in this context.

In the future you may need to choose your assay(s) to match the vaccine, or distinguish infection and vaccine induced reponses.

Evaluating performance characteristics, cross-reactivity and interfering substances

The larger the dataset the better. Evaluations of commercial assays have been undertaken on very large datasets with 500–1,000 samples.

This guidance is intended to highlight additional samples that can add value to existing evaluations by locally sourcing hard-to-obtain materials (pre-COVID samples). It is up to the user to decide if extending the evaluation can be helpful in verification exercises in that context.

Evaluate optimum performance characteristics on samples thought to be at the peak of antibody production

Although day 14 and later may capture most of the seroconversion, it is clear that some patients will develop antibodies more slowly and this may depend on the sensitivity of your assay.

The performance assessment should be based on performance from day 21 to approximately day 40 (the likely peak of antibody production). Currently, it is expected that IgG antibody detection will peak sometime in the late first to early second month post-symptoms. IgM and IgA will have different kinetics.

Evaluate performance characteristics relevant to the intended use scenario

To demonstrate cumulative seroconversion with time, assay CV characteristics that would enable precise estimation of two- or four-fold antibody increases in the clinical acceptance criteria should be chosen.

To understand assay performance at earlier stages of infection use stratified sample cohorts within specified time periods and evaluate each separately, for example day 0–5, day 6–10 and so on. This information will assist in clinical interpretation of results in the context of time from symptom onset or exposure.

Do not mix pre-21-day and post-21-day sample calculations of overall performance, or mix asymptomatic, mild and hospitalised cases. Those represent different use scenarios and the overall performance may not be helpful to understand the potential differences in assay performance in each context. Thus, in order to evaluate performance in mild disease, use only mild cases in the true positive sera-set of your evaluation and calculate performance characteristics on this cohort.

Evaluate cross-reactivity and interfering substances

It is very important to evaluate the specificity for COVID-19 antibodies versus common causes of similar winter illnesses,. This will be very important for the next winter season when all will be circulating (see Appendix 2). It may be necessary to evaluate this in the early phases of the next season if historical samples are not currently available. Consider prospective sample collection.

It is also important to evaluate relevant and extensive panels of interfering substances. Interferences should preferably be used at the expected prevalence and levels likely to be encountered in the tested patient cohort.

The potential for cross-reactivity in nucleocapsid and spike proteins in human coronaviruses (hCoV) is of particular concern, but appears relatively uncommon. Seroprevalence of SARS-CoV-1 and MERS is likely to be low in most regions of the UK. Nevertheless, it will be important to evaluate cross-reactivity with other hCoV in circulation. SARS-CoV-1 is not in active circulation globally, so is not likely to be a problem for most services. Furthermore, since antibody responses to other hCoV may be short lived or weak, sera needs to be obtained from patients contemporaneously with a proven hCoV infection by respiratory virus PCR. However, these are in short supply and may be hard to find in sufficient numbers from the pre-COVID-19 era to evaluate this robustly. Prospective collection of samples for future evaluation may be needed.

Other disease controls should also be evaluated, such as:

- high acute phase responses
- prothrombotic states with high d-dimers or fibrinogen
- high immunoglobulins (polyclonal and monoclonal >25g/L)
- rheumatoid factor positives
- high titre ANA positives (strictly these should be from untreated autoimmune diseases with a penchant for polyclonal antibody production and assay interferences, like SLE, High anti-double stranded DNA signal is a reasonable surrogate for that)
- patients on dialysis or in acute renal failure.

Proposed sera-set for validation or verification

This guidance is intended to assist users to verifiy their own lateral flow or other immuoassays that have not undergone PHE evaluation.

It is not intended to replace or substitute for PHE evaluations which have been conducted on large sera-sets with many positive RT-PCR samples and therefore can evaluate senisitivity and specificity with a high degree of confidence and very tight confidence intervals. So far, none of these evaluations have consistently included large numbers of pre-COVID sera.

A sera-set of 200–250 is probably the minimum sample set numbers for useful validation in routine practice, and half that number would probably be sufficient for verification of specificity, albeit with larger confidence intervals. However, neither will be sufficient to robustly evaluate sensitivity in low prevalence scenarios with sufficiently tight confidence intervals to distinguish between assays. Sourcing 1,000 samples for evaluation would improve the data.

Suggested validation sera-set for full validation of a quantitative in-house immunoassay

Several collaborating centres using the same methodology may be able to combine data to enhance the robustness of evalution.

If evaluation data is available from elsewhere using large numbers of positives (approximately 100), then limited verification plus an extended focus on missing interferences or cross-reactivity sera may be all that is required to meet clinical acceptance criteria. If verifying an in-house immunoassay, the full set may be needed.

Total sample-set minimum ~200–250 (verification may only require half this number in similar ratios)

1. TRUE positives (e.g. minimum ~13 samples for a ~5% pre-test prevalence)

1.1. Use locally derived COVID-19 positive samples at day 21 onwards, at lowest reasonable pre-test prevalence of disease. These samples should all be either hospitalised COVID-19 patients or mild or asymptomatic disease, as appropriate to the clinical use scenario. Sensitivity and specificity for each use scenario can be calculated using the same negative and confounder data. If you choose to use higher numbers you will improve the confidence intervals for specificity estimates, but may reduce your ability to estimate false-positive rates in the negative sera-set unless large sample numbers.

1.2. Any additional information on time from symptoms should be reported separately stratified in five-day inclusive periods (e.g. day 1–5, day 6–10), but should be not be combined in the evaluation for overall sensitivity and specificity. This should focus on the peak period for seropositivity for the desired clinical use scenario.

2. Negatives minimum ~150–175 samples (ensure these represent true negatives)

2.1. Historic Pre-COVID-19 samples with low probability of SARS-CoV-1 exposure (these are likely to be in short supply) – aim for at least 20 if possible. If possible, these should comprise the majority of your true-negative sera-set because they cannot have had undiagnosed COVID-19.

2.2. If true negative pre-COVID-19 materials are not available, then you may have to include current COVID-19 era but SARS-CoV-2 PCR negative patients who are not thought to have COVID-19. This is not ideal; wherever possible use pre-COVID samples in preference. These samples may be difficult to assign true negative status too, but may be all that you have to work with. The results will give you some information on the relative efficacy of your current local RT-PCR diagnostic pathway but the results can be excluded from the sensitivity/specificity calculations. Some centres may prefer samples from patients with two serial negative PCR tests to reduce, but not eliminate, this problem. Samples should be timed at day 21 or later from the date of symptoms or PCR to capture peak possibility of seroconversion. There remains a possibility of missed COVID-19 here and any positives in this cohort may have to be subjected to inhibition experiments to determine whether seropositivity is real, should it occur.

3. Interfering substances (locally derived) minimum ~ 25 samples (use pre-COVID-19 samples where possible. Practical consideration may dictate that some of these may have to be from the COVID-19 era, and a decision will need to be made on how to deal with this possibility if historic samples cannot be obtained, if inhibition experiments to differentiate true and false positives are not available – utilising sample exchanges with other methods may assist)

3.1. IgG paraproteins (or IgM for IgM assays, etc.) at medium and high levels

3.1.1. Myeloma sera (isotype matched) x 2

3.1.2. IgM paraproteins (if possible) x 2

3.2. Hypergamma samples

3.2.1. IVIG preparation x 2 pooled from different countries of origin in pre-COVID-19 era batches (US versus European produced) – it is possible to get enough from the residue in the bottles after infusions in local immunology, haematology or neurology units. Use at 3 dilution spiked into a negative serum/plasma/whole blood matrix – for total IgG within normal range (10g/L), high (20g/L) and very high 30–40g/L)

3.3. Hyperlipidaemia x 2

3.4. Hyperbilirubinaemia x 2

3.5. Polyclonal hyper gammaglobulinaemia x 4 (2 non-HIV, 2 HIV)

3.6. Biotin at physiological doses (if the assay using biotin susceptible technology)

3.7. Heterophile/HAMA sera x 2 (if available)

3.8. Haemolysis x 2

3.9. Heparinised and anticoagulated sera (x 2) (particularly relevant to hospitalised patients)

3.10. Samples from disseminated intravascular coagulation cases (if available) x 2

3.11. Samples from any relevant confounding conditions like haemodialyis.

4. Disease controls minimum ~30–40 samples

4.1. Historic Pre-COVID-19 with known PCR positivity for hCoV, blood taken within three months after swabbing (preferably from a nationally or regionally produced sera-set, as likely to be in short supply). If a full viral PCR screen for respiratory viruses has been done, then this will provide some or all viral disease controls if positive for adenovirus, RSV, etc. ~20 samples.

4.2. Other severe pneumonias or syndromes, known to be non-COVID-19 if possible (these may have to be evaluated prospectively after introduction into use, as they become available), some centres use double-negative RT-PCR patients.

Comprising:

4.2.1. principally influenza or bacterial pneumonias x 2

4.2.2. pneumonias with high acute phase response (CRP>100) x 2

4.2.3. samples with non-COVID-19 high D-dimer/coagulopathy (x2)

4.2.4. samples with high Troponins (x2)

4.2.5. a range of likely cross-reactive infections or those predisposing to interfering antibody production e.g. CMV, EBV, influenza, parainfluenza. Rarer infections may have to be assessed prospective after method introduction if not available (see Appendix 2).

4.3. National Institute for Biological Standards and Control (NIBSC) working calibrant and quality control materials, and any EQA material back-samples available at the time.

Suggested verification sera-set

The same sera-set above would be ideal but smaller sample sets in similar proportions (probably 40–80 samples total). It may be necessary to focus on the common causes of cross-reactivity and interferences. It may not be practical to perform the rarer serological cross-reactivity and disease controls unless sources of materials are available.

With verification exercises, it is rare to have a sample-set large enough for very narrow confidence intervals. Appendix 3 demonstrates the confidence intervals that can be expected in similar sized datasets. The larger the sample size, the greater the confidence in estimates.

Again, pre-test prevalence should reflect the clinical use scenario and local estimates of prevalence (which are currently estimated to be anywhere between 1% and 25% in different UK cohorts). Sensitivity confidence intervals become very large at pre-test prevalences below 5% in sera-sets of 80 or fewer (see Appendix 3).

An example of a possible verification plan for a lateral flow assay is given in appendix 1. This may be adapted for local use for planning and recording an EIA validation or verification exercise.

Record validation and verification data in a harmonised format

An example of a V&V record for accreditation purposes is provided in appendix 4.

Appendix 1: Example of a verification plan for a lateral flow assay

The user must specify the use scenario and select appropriate quality requirements for relevant assay type and clinical use scenario.

Clinical quality requirements

- For example, detect IgG or IgM to coronavirus to identify previous infection in ≥95% of positive cases.
- A qualitative output, easily and reproducibly read by users.
- Specificity 98%; sensitivity greater than 95% in hospitalised infections.
- Specificity 98%, sensitivity greater than 80% in mild infections.
- CE marked.
- Cost effective.
- Quick turnaround time
- No relevant common interferences.
- Suitability for NPT in hospital or in community by trained staff
- Differentiates other circulating winter viral pneumonias without interference (influenza, Parainfluenza etc.).

Number of measurements per sample

Samples will be run in singlet initially for the lateral flow and in duplicate or singlet according to clinical practice for the quantitative immunoassays.

Reading may be done by several people independently, but rapidly within the appropriate window of opportunity for each assay (generally <15 mins).

Risk assessment (needed for each laboratory)

- Generally CL2 only (local risk assessment required).
- No requirement for heat inactivation.
- Follow guidance for handling and processing laboratory specimens: www.gov.uk/government/publications/wuhan-novel-coronavirus-guidance-for-clinical-diagnostic-laboratories/wuhan-novel-coronavirus-handling-and-processing-of-laboratory-specimens

Analysis

- Inter- and intra-assay precision not possible for qualitative assays.
- PPV/NPV/Sens/Spec from 2x2 tables on day 21+ samples.
- Validate reproducibility of reading (multiple reader with a chart to harmonise intensity reading).

- Analyse results for use in the context/clinical outcome desired, in light of above uncertainties and variables.
- The intensity of signal should be compared to a local reference ELISA or well-validated immunoassay on the same sera-set.
- Consider combining data from other centres evaluating the same assay using similar methods to increase the sera-set size and statistical robustness of the data.

Appendix 2: Suggested analytical and clinical specificity matrix for SARS-CoV-2 IgG, IgM, IgA assays

All pre-pandemic samples **or** situations in a confirmed SARS-CoV-2 PCR negative patient.

a. Essential list (analytical specificity): at least two of each if possible (minimum 10–20 samples)

- other coronaviruses, hCoV 229E, OC43, HKU1, NL63 epitopes
- Human Metapneumovirus (hMPV)
- Parainfluenza virus 1-4
- Influenza A, B
- Enterovirus (e.g. EV68)
- Respiratory syncytial virus
- Rhinovirus

b. Desirable list (clinical specificity): at least one of each if possible (minimum 10 samples)

- Adenovirus (e.g. C1 Ad. 71)
- Chlamydia pneumoniae
- Haemophilus influenzae
- Legionella pneumophila
- Mycobacterium tuberculosis
- Streptococcus pneumoniae
- Streptococcus pyogenes
- Bordetella pertussis
- Mycoplasma pneumoniae
- Pneumocystis jirovecii (PJP)
- COVID-19 PCR negative sera from autoimmune antibody positive (RF, ANA) sera, HIV positive polyclonal sera
- For IgM assays: qualify non-specific cross-reactivity rates against other IgM assay in routine viral serology diagnostics

NPV (negative predictive value) matrix

- Day 28 convalescent sera in COVID-19 PCR positive patients
- Any control SARS-CoV-2 seropositive from NHSBT convalescent pool or NEQAS (whichever available).

Appendix 3: Examples of potential confidence intervals for similar results at different pre-test prevalences and sera-set numbers

Example 1 – VALIDATION: Statistics and confidence intervals for a validation,or extended verification dataset of 213 samples at a pre-test prevalence of 6% where 12 of 13 COVID-19 patients were positive (8% false negative) and 3 of 200 negative samples were 'false-positive' (1.5% FPR)

Table Analysed	Sensitivity and specificity		
P value and statistical significance			
Test	Fisher's exact test		
P value	<0.0001		
P value summary	****		
One- or two-sided	Two-sided		
Statistically significant (P < 0.05)?	Yes		
Effect size	Value	95% CI	
Sensitivity	0.9231	0.6669 to 0.9961	
Specificity	0.9850	0.9568 to 0.9959	
Positive Predictive Value	0.8000	0.5481 to 0.9295	
Negative Predictive Value	0.9949	0.9720 to 0.9997	
Likelihood Ratio	61.54		
Methods used to compute CIs			
Sensitivity, specificity, etc.	Wilson-Brown		
Data analyzed	COVID	NO COVID	Total
Spike CHOAb +	12	3	15
Spike CHOAb -	1	197	198
Total	13	200	213
Percentage of row total	COVID	NO COVID	
Spike CHOAb +	80.00%	20.00%	
Spike CHOAb -	0.51%	99.49%	
Percentage of column total	COVID	NO COVID	
Spike CHOAb +	92.31%	1.50%	
Spike CHOAb -	7.69%	98.50%	
Percentage of grand total	COVID	NO COVID	
Spike CHOAb +	5.63%	1.41%	
Spike CHOAb -	0.47%	92.49%	

Example 2 – VERIFICATION: Statistical analysis and confidence intervals for a verification dataset of 107 samples with a pre-test prevalence of 9%

P value and statistical significance			
Test	Fisher's exact test		
P value	<0.0001		
P value summary	****		
One- or two-sided	Two-sided		
Statistically significant (P < 0.05)?	Yes		
Effect size	Value	95% CI	
Sensitivity	0.9000	0.5958 to 0.9949	
Specificity	0.9794	0.9279 to 0.9963	
Positive Predictive Value	0.8182	0.5230 to 0.9677	
Negative Predictive Value	0.9896	0.9433 to 0.9995	
Likelihood Ratio	43.65		
Methods used to compute CIs			
Sensitivity, specificity, etc.	Wilson-Brown		
Data analyzed	COVID	NO COVID	Total
Spike CHOAb +	9	2	11
Spike CHOAb -	1	95	96
Total	10	97	107
Percentage of row total	COVID	NO COVID	
Spike CHOAb +	81.82%	18.18%	
Spike CHOAb -	1.04%	98.96%	
Percentage of column total	COVID	NO COVID	
Spike CHOAb +	90.00%	2.06%	
Spike CHOAb -	10.00%	97.94%	
Percentage of grand total	COVID	NO COVID	
Spike CHOAb +	8.41%	1.87%	
Spike CHOAb -	0.93%	88.79%	

Example 3 – VERIFICATION: Statistical analysis and confidence intervals for a verification dataset of 106 samples with a pre-test prevalence of 6%

Table Analyzed	Sensitivity and specificity		
P value and statistical significance			
Test	Fisher's exact test		
P value	<0.0001		
P value summary	****		
One- or two-sided	Two-sided		
Statistically significant (P < 0.05)?	Yes		
Effect size	Value	95% CI	
Sensitivity	0.8571	0.4869 to 0.9927	
Specificity	0.9899	0.9450 to 0.9995	
Positive Predictive Value	0.8571	0.4869 to 0.9927	
Negative Predictive Value	0.9899	0.9450 to 0.9995	
Likelihood Ratio	84.86		
Methods used to compute CIs			
Sensitivity, specificity, etc.	Wilson-Brown		
Data analyzed	COVID	NO COVID	Total
Spike CHOAb +	6	1	7
Spike CHOAb -	1	98	99
Total	7	99	106
Percentage of row total	COVID	NO COVID	
Spike CHOAb +	85.71%	14.29%	
Spike CHOAb -	1.01%	98.99%	
Percentage of column total	COVID	NO COVID	
Spike CHOAb +	85.71%	1.01%	
Spike CHOAb -	14.29%	98.99%	
Percentage of grand total	COVID	NO COVID	
Spike CHOAb +	5.66%	0.94%	
Spike CHOAb -	0.94%	92.45%	

Appendix 4: Example of a verification and validation record sheet

Adapt as necessary for COVID-19 antibodies or qualitative lateral flow evaluations

ASSAY NAME: Antibodies to SARS-C0V-2 by ELISA	
SOP NUMBER	16278674
EQUIPMENT AND TECHNIQUE	ELISA
DATE VERIFICATION/VALIDATION STARTED	June 2020
DATE VERIFICATION/VALIDATION COMPLETED	June 2020
VERIFICATION/VALIDATION PERFORMED BY	xxxxxx
DATE AUTHORISED AND BY WHOM	xxxxx
LOCATION OF DATA	I:/labs/validationdata/Tetanus
SUMMARY SHEET COMPLETED ON: 05/06/20	**BY:**

CLINICAL QUALITY REQUIREMENT

Specificity 98% or greater, sensitivity 95% or greater at day 21+.

Reliable estimation of antibody concentrations across the working range of the assay.

No significant interference that cannot be mitigated.

Inter-assay and intra-assay CV of <=15% at threshold of x and at y.

Good EQA performance in comparison with method group (with performance characteristics within the 2.5th and 95th centiles of ALTM distribution in any period) across x to y ran.

Calibration against standard/working calibrant – specify.

CONCLUSION AND SUMMARY OF RESULTS

Reliable estimation of antibody concentrations across the working range of the assay.

No more susceptible to interferences than comparable EIA's, all can be mitigated in reporting or sample acceptance criteria.

Satisfactory Inter-assay and intra-assay CV of <=15 % at threshold of x and y.

Good EQA performance in comparison with method group (with performance characteristics within the 2.5th and 95th centiles of ALTM distribution in any period) across measuring rang.

Calibration against reference standard satisfactory across reporting range of assay.

No sample lability issues.

VERIFICATION OF RESULTS

1. Precision (analytical uncertainty) or reproducibility

Inter-assay CV at relevant concentrations across the working range of the assay =

UoM (Uncertainty of Measurement) statement =

Inter-assay CV at at relevant concentrations across the working range of the assay =

UoM (Uncertainty of Measurement) statement =

2. Bias

3. Linearity/recovery

VALIDATION OF RESULTS (additional requirements)

1. Detection limits

LLOQ =

2. Interferences

List those checked

Uncertainty of measurement

For verification of qualitative tests only reproducibility and bias (EQA samples or comparison with alternate method) need to be assessed. There may be additional test dependent validation requirements.

Once in use, assay performance should be reviewed monthly and discussed as appropriate at departmental quality control meetings.

ISO15189 paragraph 5.5.1.4 states: 'The laboratory shall determine measurement uncertainty for each measurement procedure in the examination phase used to report measured quantity values on patients' samples.' The overall uncertainty of measurement for an assay is described above – under precision. The individual measurements (e.g. pipetting steps, plate-reader) within an assay that contribute to overall uncertainty of measurement are considered in the following table.

Table 1: Uncertainty of measurement examination steps and control measures

The text in this table is an example of one potential approach to content – please add your own.

Examination step	Consideration of UoM	Control measures
Reagents stored at 2–8ºC	Variance in Ab:Ag binding	ELISA processor holds reagent in a temperature controlled reagent bay at 2–8ºC. Out-of-range temperatures are flagged by the analyser. Fridge and lab temperatures checked each working day.see SOP xxxx. Reagents not used post expiry date. Controls and tests assayed within same run.
Off-line manual sample dilutions	Variance in pipette performance	See Pipette SOP xxxx, requires CV<2%, accuracy +/-2%. Pipette number(s) recorded on work sheet. Pipettes checked six-monthly; annually by external UKAS registered contractor.
On-board automated calibrator and sample dilution	Variance in calibrator and sample dilutions	Analyser component failure flags. Controls and tests assayed within same run. Calibrator or control failures.
Incubation	Variance in temperature and variance in incubation timings	ELISA processor incubates assays at room temperature (18–24ºC). Controls and tests assayed within same run. Instrument flags indicating out of range conditions. Note shifts in control results or calibration failures.
Reading spectrophotometer	Clarity of ELISA wells Variance in detection system	ELISA well blank reading checks. Controls and tests assayed within same run. Analysers monitor light source output and flags failures. Increase in CV% of the assay or shift in IQC not due to other changes.
Final result	Accumulated effects of the foregoing factors	See QC SOP xxxx: select third party IQC, where possible, to monitor assay performance at clinically relevant levels. Monitor on-going CV% and review IQC against multi rule Westgard rules. EQA, or sample exchange, to compare performance with peer group.