UK Health Security Agency

UK Standards for Microbiology Investigations

Validation and verification of diagnostic tests

Issued by the Standards Unit, UK Standards for Microbiology Investigations, UKHSA

Quality guidance | Q 1 | Issue number: dza+ | Issue date: dd.mm.yy |

Acknowledgments

UK Standards for Microbiology Investigations (UK SMIs) are developed under the auspices of UKHSA working in partnership with the partner organisations whose logos are displayed below and listed on the UK SMI website. UK SMIs are developed, reviewed and revised by various working groups which are overseen by the UK SMI steering committee.

The contributions of many individuals in clinical, specialist and reference laboratories who have provided information and comments during the development of this document are acknowledged. We are grateful to the editors for editing the medical and clinical content.

UK SMIs are produced in association with:

Displayed logos correct as of August 2026

Contents

Amendment table

Each UK SMI document has an individual record of amendments. The amendments are listed on this page. The amendment history is available from standards@ukhsa.gov.uk.

Any alterations to this document should be controlled in accordance with the local document control process.

Amendment number/date	x/dd.mm.yy
Issue number discarded	
Insert issue number	
Anticipated next review date*	dd.mm.yy (usually issue date+3 years)
Section(s) involved	**Amendment**
Title	Title changed from ‘Evaluations, validations and verifications of diagnostic tests’ to ‘Validation and verification of diagnostic tests’.

*Reviews can be extended up to 5 years where appropriate

1 General information

View general information related to UK SMIs.

2 Scientific information

View scientific information related to UK SMIs.

3 Scope of document

This UK Standards for Microbiology Investigations (UK SMI) document describes the principles and processes required for the validation and verification of diagnostic tests used within microbiology and infection science laboratories.

It applies to a broad range of diagnostic methods, including:

- Commercial in vitro diagnostic (IVD) devices
- In-house (laboratory-developed) methods and reagents
- Modified commercial assays (including off-label use)
- Combinations of reagents used to create in-house methods
- Point of care tests (POCT).

The document also reflects the evolving regulatory landscape. Laboratories should be aware of the requirements of the UK Medical Devices Regulations 2002 (UK MDR 2002) as amended, and the ongoing programme of UK medical devices regulatory reform. Northern Ireland continues to implement the EU *In Vitro Diagnostic Regulation (IVDR; EU 2017/746)* (1). Whilst UK MDR 2002 remains in force, laboratories should ensure that validation and verification practices are aligned with current regulatory requirements where appropriate. In addition, the Health Institution Exemption (HIE), developed in the UK by the Medicines and Healthcare products Regulatory Agency (MHRA), provides guidance on in-house In-Vitro Diagnostic (IVD) Medical Devices Health Institution Exemption for general medical devices - GOV.UK.

In addition to validation and verification, diagnostic laboratories may also undertake evaluations of novel or significantly modified diagnostic tests, devices, reagents, software, or workflows as part of routine service development and quality improvement activities. These evaluations are typically conducted to assess operational suitability, usability, workflow impact, cost-effectiveness, or fitness for implementation within a local service. Such evaluations are distinct from the validation and verification activities described in this document. Where an evaluation supports the proposed introduction, modification or adoption of a diagnostic method, the method must subsequently undergo appropriate validation or verification, as applicable, before routine clinical use. Although many principles described here apply, evaluation of diagnostic tests is not in the scope of this document. Evaluation nevertheless remains an important component of service development and may

support decisions regarding the introduction, optimisation, or replacement of diagnostic methods. Laboratories should ensure that evaluation activities are conducted in accordance with local governance requirements and documented appropriately within their quality management systems. However, evaluation alone does not demonstrate fitness for purpose and does not replace the requirement for validation or verification before implementation into routine clinical service.

For information on CE marking, refer to Regulating medical devices in the UK - GOV.UK and on quality assurance, refer to UK SMI Q 2 - Quality assurance in the diagnostic infection sciences laboratory.

UK SMIs should be used in conjunction with other relevant UK SMIs.

4 Introduction

A key role of the laboratory service is to decide which tests should be offered and to select the most appropriate kits and methods for these. The performance specifications of any new or modified laboratory method are integral to providing a high-quality service.

Validation and verification are integral requirements for the accreditation of laboratories according to ISO 15189 (2) and ISO 17025 (3). These standards require that all examination procedures be validated or verified (as appropriate) for their intended use before adoption, with the methods and results obtained recorded (2).

Validation and verification are distinct processes: validation demonstrates that a method is fit for its intended purpose, whereas verification confirms that the expected performance can be achieved in the local laboratory setting (2).

5 Validation

According to ISO 15189:2022 (2), validation is defined as “confirmation of plausibility for a specific intended use or application through the provision of objective evidence that specified requirements have been fulfilled”. In other words, validation is an evidence-based assessment that a method is suitable for its intended purpose (2). Objective evidence can be obtained through observation, measurement, examination or other means.

Specified requirements of an examination method may include the following:

- performance specifications
- measurement trueness
- measurement precision including measurement repeatability and intermediate precision
- analytical specificity including interference from other substances
- detection limit and quantitation limit

- measuring interval
- clinical relevance
- diagnostic specificity
- diagnostic sensitivity.

To document this ability, each laboratory should produce a ‘Validation File’ for each method or system that has not been formally validated. The file should include a range of information and will have a different emphasis depending on whether the laboratory is using a modification of commercial systems or has developed an in-house system. Typically, the file will include sections such as validation data, tests on known samples, risk assessments, standard operation procedures or workbooks, and relevant publications (2,3).

Validation should be performed in the following scenarios before introduction into routine use:

- Non-standard methods
- Laboratory designed/developed methods
- Standard methods used outside their intended scope
- Validated methods that have been significantly modified.

Additionally, validation is required when:

- the conditions under which the original validation was done change (for example, use of an instrument with different characteristics or samples within a different carrier matrix)
- the performance of an existing method has been shown to be unsatisfactory, and changes are made that may affect the validated performance characteristics of the method
- the method is changed or modified beyond the original specification (for example, the use of a different sample type or a commercial kit modified for a clinical purpose not designated by the manufacturer).

Where examination, transportation or storage conditions, or durations differ from those specified in the manufacturer's instructions for use (IFU), laboratories should assess the impact of the variation on the examination performance. Appropriate validation should be undertaken to demonstrate fitness for the intended purpose before implementation (2,4). Examples include extending incubation beyond the manufacturer-recommended period to improve recovery of clinically significant organisms, using a culture medium for organisms not included within the manufacturer's intended use, applying a commercial assay to a specimen type not specified in the IFU, or using specimens transported or stored outside the conditions specified by the manufacturer. Please note, this is not an exhaustive list.

The examples above should be considered significant changes that require revalidation with adequate evidence for equivalent performance before implementation routinely (4,5).

Validated test methods or equipment do not require further validation unless significant changes are made that may affect performance (2,3). Ongoing fitness for purpose is monitored through the laboratory's quality assurance processes, including appropriate monitoring of the reagent (including new batches), equipment (including new or repaired instruments) and operator performance using suitable control materials. Further examples of validation scenarios are provided in Table 1.

Validation is also performed on in-house diagnostic tests to assess analytical and, where applicable, clinical performance characteristics. These may include analytical sensitivity, analytical specificity, precision and linearity (where applicable), in addition to diagnostic sensitivity and diagnostic specificity where relevant (6,7).

The purpose of validation is to provide documented evidence that a diagnostic test performs within the required specifications and is fit for its intended purpose. This involves generating and evaluating experimental data to determine its analytical performance; including accuracy, sensitivity (i.e. the lower limit of detection), specificity, reliability, repeatability, reproducibility, and measurement uncertainty. The scope of validation may vary considerably. For example, it may be extensive when validating a newly developed in-house method, or more limited when validating a commercial method that is already in use and has undergone only minor modifications. The validation approach should be appropriately powered and proportionate to the intended diagnostic use of the test, providing a robust understanding of its performance characteristics and fitness for intended use (7).

For methods already in use that do not have a formal validation record, it is important to provide documented evidence to support their continued use (4). It is usually sufficient to compile a validation file based on historical evidence such as results from comparisons or other studies undertaken, copies of published papers, External Quality Assurance (EQA), Internal Quality Assurance (IQA), and Internal Quality Control (IQC) results. Where appropriate, workbook records may be cross-referenced within the validation report (2). Refer to Appendix 1 for a summary of the content that should be included in a validation report.

6 Verification

ISO 15189:2022 (2) defines verification as "the confirmation of truthfulness, through the provision of objective evidence that specified requirements have been fulfilled". In other words, verification is the process of confirming whether a method performs as expected within the specific user laboratory environment.

Verification is the process by which the laboratory confirms that the product (for example a commercial kit or a piece of equipment) has established performance claims of a measuring system such as trueness, precision, and reportable range can

be replicated in the laboratory (6,7). This must be fulfilled before adoption for clinical use. Examples include confirmation that performance specifications of a measuring system are achieved or confirmation that a target measurement uncertainty can be met in a laboratory newly implementing a specific test that has previously been validated by another centre.

Verification should be repeated following significant changes that may affect examination performance (2).

The selection of test sample size should comprehensively consider the complexity and precision requirements of the established methods, statistical protocol for data analysis and acceptable statistical confidence level. The samples used for verification should be clinically relevant and reflect the spectrum of expected results (e.g. positive, both true and false, negative, both true and false, and colonisation). The number of specimens included should be justified based on the intended use of the test, clinical risk, performance characteristics being assessed, and the confidence required in the performance estimates (4,8,9). See Section 9 for further guidance on sample size selection.

Verification may be sufficient to implement a new IVD device under circumstances where the examination (a set of operations having the objective of determining the numerical value, text value or characteristics of a property) is performed and used in the manner as directed in the kit insert (7). Examples of verification scenarios can be seen in Table 1.

The purpose of verification is to confirm whether a product (for example a validated method, commercial kit system or equipment) complies with a regulation, requirement, specification, or imposed condition in an individual laboratory. The minimum tested attributes in verification may include accuracy, precision, and linearity but are not limited to these. Refer to Appendix 3 for a summary of the content that should be included in a verification report.

Note: Validation of commercial assay kits and the equipment used should be performed by the manufacturer to ensure that they achieve the stated performance. The user should obtain this information from the manufacturer. However, the user should perform verification to confirm through review (published and unpublished evaluations, EQA data, etc.) and testing that the equipment and the commercial kits meet the written specification requirements (6,10).

7 Common terminologies used in diagnostic methods

This section defines key terminology used in diagnostic microbiology to support consistent understanding and application of validation and verification principles (5).

7.1 In-house diagnostic tests

In-house tests or laboratory developed tests are designed and developed within a single laboratory and are not distributed or supplied to other laboratories. These tests, including molecular diagnostic assays, should generally be considered an exception to the use of commercially available assays and are typically implemented only where no suitable CE/UKCA marked test is available, such as for rare or emerging pathogens (1).

This document does not encourage the routine development or use of in-house tests, where appropriate commercial assays exist. However, it acknowledges that in certain circumstances, in-house test development is necessary to meet specific clinical or public health needs (e.g., emergence of a novel pathogen). In such cases, laboratories must ensure that these assays undergo a comprehensive and robust validation process to demonstrate that they are fit for their intended purpose prior to routine clinical use.

In-house IVDs are defined as follows: devices developed, manufactured, and used within the same health institution without CE/UKCA marking; modifications of CE/UKCA marked IVDs; or products labelled for research use only (RUO) or without CE/UKCA marking that may subsequently be used as IVDs.

7.2 Off-label diagnostic tests

These are commercial diagnostic tests that comply with the UK MDR 2002 (that is CE/UKCA marked) but are used in ways that do not comply with the manufacturer's instructions. This could involve replacing a reagent, changing a procedure or using a different sample type. It also includes commercial kits modified for clinical purposes not designated by the manufacturer to suit a laboratory's needs. Off-label commercial tests are used outside the scope of their CE/UKCA marking and should therefore be treated as in-house tests, requiring validation to demonstrate fitness for purpose before routine use (1,7,11).

Other examples include commercial tests not CE/UKCA marked but sold for RUO where results are used to support clinical decisions or disease management. 'Research use only' tests typically carry manufacturer's limitations and must be used according to the guidelines described by the manufacturer.

7.3 Commercial diagnostic tests

These are diagnostic tests (i.e. IVDs) that have been developed and validated by the manufacturer and regulated by CE/UKCA marking. All test results reported are validated by the manufacturer in compliance with the regulations surrounding the award of CE/UKCA mark. A new or updated commercial kit, which will be used according to the CE/UKCA mark IFU, must undergo verification.

For more information, on commercially produced kits, please refer to Regulating medical devices in the UK and Regulation of medical devices in Northern Ireland.

7.4 Health Institution Exemption

Health Institution Exemption (HIE) allows laboratories to use in-house or modified IVDs without CE/UKCA mark where specific conditions are met (1,11).

HIE applies when:

- the test is developed and used within the same organisation,
- it is not supplied to other organisations, and
- no suitable CE/UKCA marked alternative exists, or available options do not meet clinical need.

Use of HIE must be justified and documented, including evidence of clinical need and review of available alternatives.

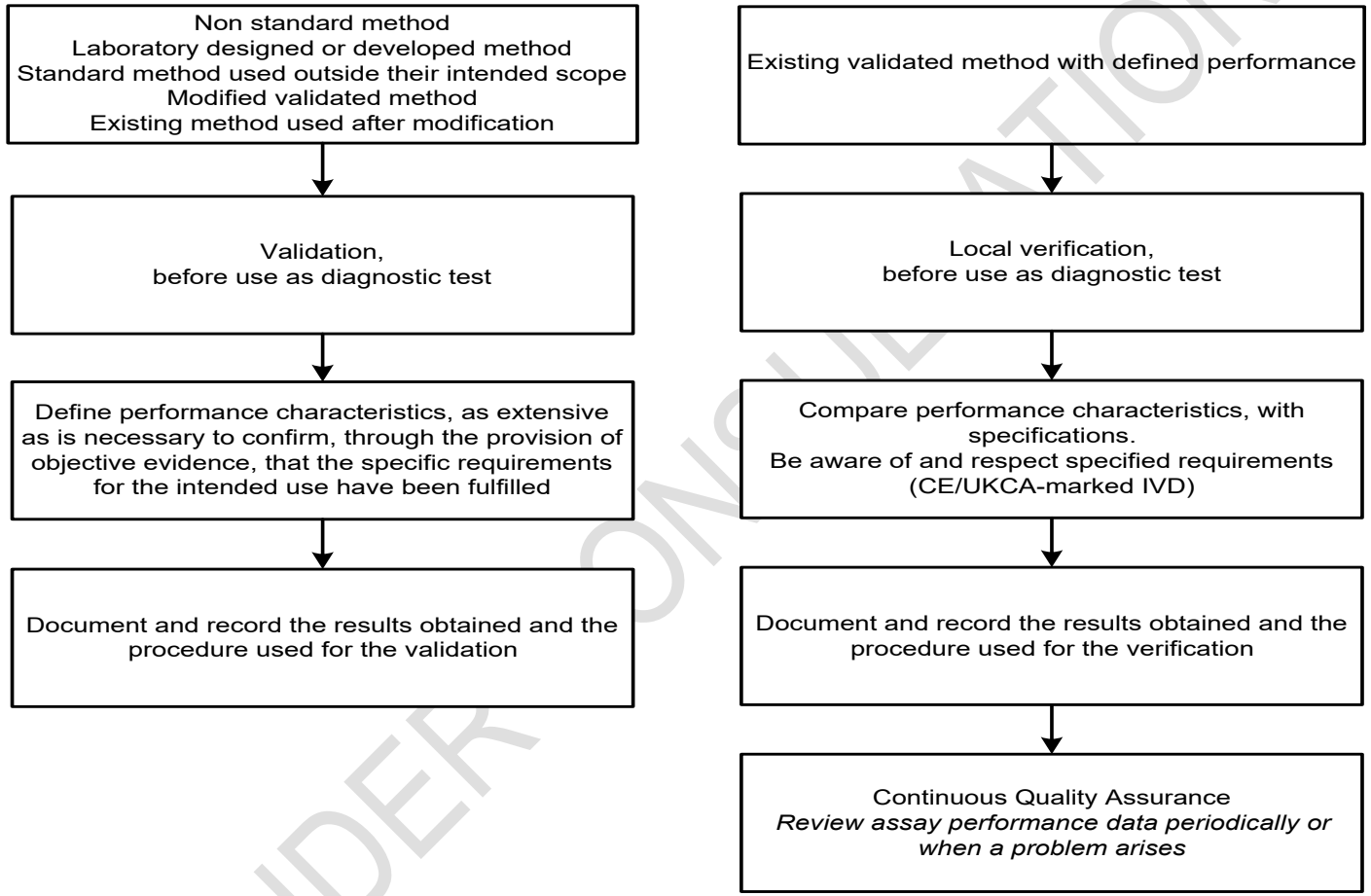
Laboratories must ensure that HIE assays:

- are appropriately validated for intended use,
- meet safety and performance requirements,
- are supported by risk assessment, and
- operate within a quality management system (e.g. ISO 15189).

Ongoing assurance should include quality control, (EQA) participation where available, and periodic performance review.

For information on Health Institution Exemption (HIE), refer to Health Institution Exemption for general medical devices - GOV.UK

Diagram 1: Validation and verification of diagnostic tests flowchart (10)



This flowchart is for guidance only.

Table 1: Examples of validation and verification

The following table provides illustrative examples of when validation and verification are required within diagnostic laboratories.

Scenario	Validation (Required when method is new or significantly changed)	Verification (Required for confirming expected performance locally)
Example 1: Modification of existing assays	Updating a molecular assay (e.g. SARS-CoV-2 or respiratory) to detect emerging variants through modification of primers or probes. Revalidation of a real time PCR following modification of primers and probe (e.g., introduction of a wobble base) to cover a new circulating strain.	Introduction of a CE/UKCA marked SARS-CoV-2 PCR assay. Verification of manufacturer performance claims using local clinical samples. Change/update of curated database for sequence-based identification, with assessment of the impact on performance.
Example 2: Use outside intended scope or method adjustment	Using MALDI-TOF for identification of isolates from a selective agar plate. Applying assays to non-validated specimen types (e.g. respiratory NAAT used on CSF or tissue). Expanding the use of a sequencing platform beyond manufacturer verified patient cohorts and sample types.	Replacement of automated platforms (e.g. VITEK to alternative system). Change in NAAT instrumentation (e.g. replacement with an equivalent thermal cycler) with comparison to the previous system.
Example 3: New methods and equipment	Introduction of a new in-house assay for rare pathogen e.g *Coxiella burnetii*. Development of in-house molecular assays for emerging or rare pathogens (e.g. mpox, Candida auris, antimicrobial resistance genes).	Installation of a CE/UKCA-marked sequencing platform with verification that manufacturer performance claims are achieved in the local setting. Introduction of automated liquid handling or digital NAAT systems• Confirmation that equipment meets manufacturer specifications in the local setting.
Example 4: AST	Modification of AST breakpoints, interpretive rules, or methodology.	Introduction of a CE/UKCA-marked AST platform according to manufacturer's instructions.

8 Performance characteristics of the diagnostic test

The table below summarises the key differences between analytical and diagnostic performance characteristics in diagnostic microbiology.

Aspect	Analytical Performance	Diagnostic Performance
Focus	Test measurement ability	Test clinical decision ability
Level	Laboratory / assay	Patient / population
Sample type	Controlled or artificial samples	Real patient samples
Outcome	Detecting the analyte	Diagnosing/excluding/monitoring disease
Example	Detecting viral RNA at low concentration	Identifying infected patients

The parameters that should be taken into account during validation or verification of diagnostic methods are as follows (12):

8.1 Analytical performance characteristics

Determining analytical performance characteristics is usually achieved by testing contrived and/or non-clinical samples, including the relevant micro-organisms.

8.1.1 Analytical sensitivity

Analytical sensitivity refers to the ability of an assay or test to detect minimal quantities of a target, such as a pathogen, biomarker, nucleic acid or chemical compound in a sample (4). It is a crucial parameter in diagnostic testing, as it determines the lowest concentration of the analyte that can be reliably (>95%) measured/detected by the assay (i.e., the limit of detection (LOD)).

High analytical sensitivity is essential for early detection of diseases, especially when the target substance is present in very low amounts, such as in the early stages of infection or in trace environmental contaminants. Very high analytical sensitivity may increase detection of low-level target material of uncertain clinical significance and may, in some circumstances, increase susceptibility to contamination or detection of non-clinically relevant findings. Analytical sensitivity can be expressed as a true

positivity rate across all expected positive samples or for individual expected positives according to varying burden.

8.1.2 Analytical specificity

Analytical specificity refers to the ability of an assay or test to accurately identify the target analyte in the presence of other substances or organisms that might be present in the sample (4). This parameter is crucial for ensuring that the rest results are not influenced by cross-reactivity with non-target analyte from other similar compounds, pathogens, or biomarkers thereby providing reliable and accurate results. This is particularly important in complex biological samples (particularly non-sterile sites) where multiple substances/organisms may coexist, as it ensures that the test preferentially detects the intended target with minimal interference from other substances or organisms. The number of target analytes that are detected by an assay is considered its range of detection. It is important to consider that analytical sensitivity/limit of detection may vary between target analytes with the range of detection of any assay. Analytical specificity can be expressed as false positivity rate or true negativity rate across samples expected to be negative.

8.1.3 Cross-reactivity to non-target organisms

Target specificity and cross-reactivity must be examined by screening a panel of analytes (e.g. DNA/RNAs or antigens) from other organisms in real-world samples. The cross-reactivity panel must include related agents, likely cross-reactive agents identified through in-silico analysis, as well as other agents likely to be encountered in a clinical sample.

The panel must include samples that are:

- Negative for the analyte of interest (target) with no other organisms present and
- Negative for the analyte of interest (target) but positive for other non-target organisms.

8.1.4 Limit of Detection

Limit of detection (LoD) is a measure of the analytical sensitivity of an assay and an important characteristic that must be determined for both quantitative and qualitative tests.

It is the lowest actual concentration of an analyte in a specimen that can be consistently detected (e.g. in 95% of specimens tested) with acceptable precision, but not necessarily quantified, under routine laboratory conditions and in a defined type of specimen. This is particularly important for tests where detection of an analyte, even in very low concentrations, is critical for diagnosis or monitoring.

8.1.5 Limit of Quantification

The Limit of Quantification (LoQ) is the smallest amount of an analyte that can be quantitatively measured with an acceptable level of accuracy and precision. Test results may not be reliable or consistent below this threshold.

The LoQ is crucial in analytical testing because it defines the lower boundary of the test's effective range, ensuring that the measurements are both accurate (close to the true value) and precise (consistent across repeated tests). The LoQ is determined through rigorous testing and validation processes, which involve analysing samples with known concentrations of the analyte and assessing the variability and accuracy of the results.

8.1.6 Analytical accuracy

Accuracy of measurement is the closeness of a measured value to the true or accepted value. It indicates how near your result is to the actual quantity being measured, often expressed as an absolute error or a percentage deviation from a known standard.

8.1.7 Precision (12)

Precision is defined as the 'closeness of agreement between results of replicate measurements', it can also be described as the level of concordance of individual test results within a single run (intra-assay) and from one run to another (inter-assay).

Precision is usually characterised by the standard deviation of the measurements and the relative standard variation (coefficient of variation). It is often expressed as the percent coefficient of the variation (%CV), where:

%CV = (standard deviation of measurements/mean) x 100

8.1.8 Bias (2,12)

Bias is the systematic difference between the result produced by a method and the true value or an accepted reference value and is influenced by the precision of the test. Unlike random error, bias consistently shifts results in one direction and may affect clinical interpretation, particularly when results fall close to clinical decision limits or interpretation breakpoints.

Assessment of bias should be considered during validation and verification by comparison with reference methods, peer-group data, manufacturer claims, EQA/PT performance, or previously validated methods (2,12).

For quantitative assays used to monitor disease progression or treatment response (for example viral load assays), systematic differences between an existing and a newly implemented method may affect comparison of results over time. Where a clinically significant bias is identified following implementation of a new method, laboratories should assess the impact on interpretation of results and determine

whether additional mitigation measures, such as parallel testing, clinician communication, or revised reporting guidance, are required.

8.1.9 Repeatability and reproducibility

Repeatability refers to the ability to obtain consistent results when the same experiment is conducted multiple times under the same conditions by the same operator (4).

It is a measure of precision within a single laboratory setting and indicates the reliability of the experimental procedure. It must be calculated using the dilution series replicates generated on one run. Use a minimum of 3 replicate measurements from a single run for SD & CV. State whether the level of variation is acceptable given the clinical range of the agents and the lower limit of detection (LLOD).

Reproducibility refers to the ability to achieve consistent results when an experiment is conducted under varying conditions, such as different operators, equipment, and dates (4).

This measure assesses the robustness of the experimental findings across different settings and ensures that the results are not specific to a particular set of conditions. It must be calculated using the dilution series replicates generated over several days. Use a minimum of 3 replicate measurements obtained over multiple days for SD & CV. State whether the level of variation is acceptable given the clinical range of the agents and the LLOD.

Note: The higher the number of variables, the more robust a test must be to achieve reproducibility.

8.1.10 Linearity and Efficiency

Linearity is the determination of the linear range of quantification for a test or test system.

Using laboratory equipment as an example, linearity defines how well the device's actual performance, across a specified operating range, approximates a straight line. In terms of a test, linearity is achieved when measured results are directly proportional to the concentration of the analyte (microorganisms or nucleic acid) in the test sample, within a given range.

Linearity is usually measured in terms of deviation, or non-linearity, from an ideal straight line and is typically expressed as a percentage of full scale or in parts per million (ppm) of full scale. Linearity testing challenges the entire equipment calibration range, including the extremes, and can detect problems such as reagent or equipment deterioration earlier than quality control or proficiency testing failures.

Linearity is the ability of a measurement or analytical method to produce results that are directly proportional to the concentration of the analyte within a given range and is critical to assays providing quantification. It must be calculated using a dilution series.

It should also be noted that it is good laboratory practice to periodically demonstrate linearity to detect reagent deterioration, monitor equipment performance, or re-confirm linearity after major service of equipment. Refer to Appendix 3 for a worked example of linearity.

For molecular amplification assays, efficiency describes how effectively the target nucleic acid is amplified during each reaction cycle and is typically assessed using a dilution series and standard curve analysis. Amplification efficiency provides an indication of assay performance and robustness and can be used to identify suboptimal assay design or reaction conditions.

For molecular assays, amplification efficiency is calculated using the equation:

$$E = 10^{[-1/\text{slope}]} - 1$$

Acceptable amplification efficiency is typically 90-110% (0.9-1.1), corresponding to a standard curve slope of approximately -3.6 to -3.1 (4).

8.1.11 Measurement uncertainty

ISO 15189 and ISO 17025 define measurement uncertainty as “a parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand (the parameter being measured)”(2,3). It is essential for the correct interpretation of a result and is particularly important when results are close to a specified limit or clinical threshold.

The ISO 17025:2017 standard, which covers the accreditation of calibration and testing laboratories, outlines specific requirements for laboratories to evaluate and report uncertainty of measurement. This requirement has also been incorporated into the ISO 15189 standard, and forms part of the assessment process used by UKAS.

For microbiology tests, the measurement uncertainty applies to measured results, such as zone sizes or minimum inhibitory concentrations (MICs) of antimicrobials, microscopy of fluids (including cerebrospinal fluids (CSF) and urines), and quantitative and semi-quantitative organism counts, (e.g. cultures in Continuous Ambulatory Peritoneal Dialysis (CAPD) fluids and urines). Other variables that could affect a result, which do not entail an actual measurement, include specimen collection, specimen preparation, transport time and conditions, media inoculation and incubation temperature. Non-numeric tests, such as subjective plate reading, are also important in assessing measurement uncertainty. For microbiology assays providing a quantitative result, measurement uncertainty may be estimated using appropriate measures of variability such as the coefficient of variation (see above).

The approach used to estimate and express measurement uncertainty should be appropriate for the nature of the result generated.

For ordinal data (not continuous numerical) results such as MICs reported in doubling dilutions, conventional calculations based on mean, standard deviation, or coefficient of variation would not be appropriate. In these situations, performance may be more

suitably assessed using measures such as repeatability, reproducibility, essential agreement, categorical agreement, and error rates.

Similarly, semi-quantitative culture results may require alternative approaches to assessing variability and fitness for purpose rather than formal measurement uncertainty calculations.

For more information on measurement uncertainty, see UK SMI Q 2 - Quality assurance in the diagnostic infection sciences laboratory.

8.2 Diagnostic performance characteristics

8.2.1 Diagnostic sensitivity (10,12)

Diagnostic sensitivity is the ability of a test (e.g. a lab assay) to detect a disease (or condition) in a patient (4).

It must be distinguished from analytical sensitivity, which is the ability of an assay to detect the presence of some target substance or organism (i.e. the analyte) in a sample.

When diagnosis is made based on testing a sample, diagnostic sensitivity is a measurement of the ability of the test to detect samples that are positive by the chosen reference standard, i.e. “true positive” samples. Diagnostic sensitivity can only be determined by analysing samples from several patients with the disease/condition, i.e., a case population.

Diagnostic sensitivity can then be calculated as the number of true positive samples (a) correctly identified by the assay under evaluation, divided by the total number of true positive patients (i.e. those positive by the reference assays) (a+c), expressed as a percentage, as indicated in the table below. It is expressed as:

Sensitivity = a/(a+c) x 100

8.2.2 Diagnostic specificity (10,12)

Diagnostic specificity is the ability of a test (e.g., a lab assay) to correctly identify a person without the disease (or condition) in question (4).

It should be distinguished from analytical specificity, which is the ability of an assay to exclusively detect the target substance or organism (i.e. the analyte) in a sample, rather than similar but different substances.

When diagnosis is made based on testing a sample, diagnostic specificity is a measurement of the ability of the test to not detect samples that are negative by the chosen reference standard, i.e. “true negative” or control samples.

Diagnostic specificity can only be determined by analysing samples from several patients, i.e. a control population. Diagnostic specificity can then be calculated as the number of true negative patients correctly identified by the assay under evaluation (d),

divided by the total number of true negative/control patients (i.e. those negative by the reference assays) (b+d), expressed as a percentage. It is expressed as:

Specificity = d/(b+d) x 100

Calculation of diagnostic sensitivity and specificity:

	True positive specimens	**True negative specimens**	**Total**
Positive	a True positives	b False positives	a+b
Negative	c False negatives	d True negatives	c+d
Total	a+c	b+d	a+b+c+d

Note: These parameters are highly population dependent and can be influenced by the prevalence of disease. Predictive values are always affected by prevalence, while the terms ‘sensitivity’ and ‘specificity’ are better considered as inherent to the assay and will only be affected if the population is qualitatively rather than quantitatively different.

For example, if two populations have the same prevalence but with a different proportion of people in the very early phase of infection (with a low microbial load), the sensitivity but not specificity of the assay will be different. In contrast in a situation where populations have different prevalences but similar microbial loads in the infected cases, the sensitivity and specificity are likely to be the same. Predictive values will be different in this case.

8.2.3 Limitations of reference standards in diagnostic validation

In some situations, particularly where novel or more sensitive technologies are being introduced, an existing comparator method may not represent a perfect reference standard. Examples may include metagenomic sequencing, highly sensitive nucleic acid amplification tests, multiplex syndromic panels, or other emerging diagnostic technologies being assessed and compared to conventional microbiological approaches (e.g., culture).

Where a new method is capable of detecting targets potentially missed by the comparator method, apparent false-positive results should not automatically be assumed to represent incorrect results particularly when analysing samples from non-sterile sites (13). Additional investigations and patient record review may be required to determine the true disease status of the patient and appropriateness of the false

positive label assigned to the test under investigation. Investigating discordant results with a third test potentially detecting the same target as the index test (e.g., using PCR to confirm the presence of nucleic acid detected by NGS methods where culture is discordant) may be useful.

In such circumstances, assessment of diagnostic performance should consider all available evidence, which may include clinical presentation, radiological findings, results from alternative laboratory investigations (including other tests with similar performance to the test under investigation for example, qPCR vs NGS), treatment response, expert clinical review, and follow-up information. Composite reference standards or final clinical diagnosis may provide a more appropriate comparator than a single laboratory method where no true gold standard exists (4,12,14). If a syndromic method (e.g.,multiplex PCR, NGS) is being investigated then comparison with the existing range of tests used to cover this scope will be required. While newer technologies may increase positivity rates over existing testing and could reflect increased sensitivity, the presence of positivity with an individual pathogen that was not previously associated with a significant burden of disease needs should not be immediately considered to represent increased sensitivity for the novel test and potential sources of false positivity need to be thoroughly investigated.

The limitations of the chosen reference standard should be documented and considered when interpreting estimates of diagnostic sensitivity, diagnostic specificity, predictive values, and other measures of test performance. For transparency, presenting diagnostic performance for the test under investigation with strict adherence to the reference method (despite its limitations) and also when utilizing a composite reference standard may be appropriate.

8.2.4 Incorporation bias

When evaluating diagnostic performance, the results generated by the test under investigation should not be used to define disease status or determine the reference standard. This may lead to incorporation bias and artificially inflate estimates of diagnostic sensitivity, specificity, and overall accuracy. Where composite reference standards are used, the contribution of the test under investigation should be clearly documented and justified, and potential sources of bias should be considered during interpretation of the results (4,12,14).

8.2.5 Positive predictive value (PPV) (12)

PPV is the probability that a person whose test result is positive truly has the disease (or condition) of interest. It is expressed as:

Positive predictive value (% PPV) = (a/(a+b)) x 100

PPV improves with strong diagnostic specificity but can be artificially high if case numbers exceed control numbers, as pre-test probabilities of the disease/condition will already exceed 50%.

8.2.6 Negative predictive value (NPV) (12)

NPV is the probability that a specimen with a negative test result truly does not contain the target. It is expressed as:

Negative predictive value (% NPV) = (d/(c+d)) x 100

NPV improves with strong diagnostic sensitivity but can be artificially high if case numbers are very low (e.g., <10%), as pre-test probabilities of patient not having the disease/condition will already exceed 90%. If the incidence of disease is <5% then the NPV is 95% prior to performing any diagnostic tests and even tests with poor sensitivity (50%) will improve NPV to 97%, despite missing 50% of cases.

8.2.7 Positive likelihood ratio (12)

With predictive values highly influenced by the incidence (pre-test probability) of disease within the population tested, likelihood ratios (LR) provide a direct estimate of how much a test result will change the odds of having a disease and is less affected by incidence as it incorporates both the diagnostic sensitivity and the diagnostic specificity of the test.

Typically, a LR+ of 10:1 (10) is considered diagnostically acceptable.

The likelihood ratio for a positive result (LR+) tells you how much the odds of the disease increase when a test is positive. It is estimated as the ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease:

- Positive diagnostic likelihood ratio = sensitivity÷(100-specificity)

8.2.8 Negative likelihood ratio (12)

The likelihood ratio for a negative result (LR-) tells you how much the odds of the disease decrease when a test is negative. It is estimated as the ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease:

- Negative diagnostic likelihood ratio = (100-sensitivity) ÷ specificity

Typically, a LR- of 1:10 (0.1) is considered acceptable for excluding disease.

8.2.9 Clinical accuracy (4)

Accuracy is the closeness of agreement between the value obtained from a large series of test results and an accepted reference value. In microbiology, the factors that determine the degree of accuracy are as follows:

- The uniformity of microbial load in the sample
- The accuracy of equipment
- The volume of sample/reagents used for testing

- Sample storage
- Sample selection
- The media used and the incubation conditions
- The reading and interpretation of results by staff performing the test
- Operator error

The last two factors both require training and competency of staff

Overall probability that a patient is correctly classified:

% Accuracy = ((a+d) ÷ (a+b+c+d)) x 100

8.2.10 Additional statistical tests for validation (12)

The use of Receiver Operating Characteristic (ROC) curves, the Youden Index, and the Diagnostic Odds Ratio (DOR) can be helpful when evaluating the diagnostic performance of a test and to determine optimal test cutoffs.

- ROC curves graphically assess a test's ability to distinguish between diseased and non-diseased individuals by plotting sensitivity against 1 − specificity across a range of cutoff values. The Area Under the Curve (AUC) provides an overall measure of diagnostic accuracy, with values closer to 1.0 indicating better discrimination.
- The Youden Index (J = Sensitivity (in decimal) + Specificity (in decimal) − 1) is used to identify the optimal decision threshold. The cutoff that maximizes the Youden Index provides the best balance between sensitivity and specificity and is often selected during assay validation, but other threshold may be selected to optimise sensitivity or specificity.
- The Diagnostic Odds Ratio (DOR) combines sensitivity and specificity into a single measure of test effectiveness. It is calculated as the odds of a positive test result in diseased individuals relative to the odds of a positive result in non-diseased individuals. Higher DOR values indicate better discriminatory performance, while a DOR of 1 suggests no diagnostic value. While the value is a combination of sensitivity and specificity, it is not feasible to determine what parameter is optimal directly from the DOR and these should be derived from the likelihood ratios that are used to calculate the DOR (DOR: LR+/LR-).

Together, ROC curves evaluate overall diagnostic accuracy, the Youden Index helps determine the optimal cutoff value, and DOR summarizes the test's overall diagnostic power, making these metrics valuable tools for laboratory test validation and verification.

8.3 Additional performance parameters for antimicrobial susceptibility testing

For antimicrobial susceptibility testing (AST) methods, additional performance characteristics should be evaluated where applicable (15).

These may include:

- Essential Agreement (EA): agreement between MIC values generated by the test and the reference method, usually within ± one doubling dilution.
- Categorical Agreement (CA): agreement between interpretive categories (Susceptible, Intermediate/Susceptible-Increased Exposure, Resistant).
- Very Major Errors (VME): false susceptible results by the method being assessed.
- Major Errors (ME): false resistant results by the method being assessed.
- Minor Errors (mE): discrepancies involving intermediate or susceptible-increased exposure categories.

Where AST methods are validated or verified against a reference method, these measures should be assessed in accordance with relevant standards and guidance, including ISO 20776-2 where appropriate (15). Laboratories should also consider current guidance published by EUCAST and BSAC, available at www.eucast.org and www.bsac.org.uk, respectively.

8.4 Confidence intervals (4,12)

Estimates of diagnostic sensitivity, specificity, predictive values, likelihood ratios, accuracy, and other performance measures should, where appropriate, be accompanied by confidence intervals to indicate the precision of the estimate. The width of the confidence interval is influenced by factors such as sample size and the observed performance of the test. Their use across different tests when testing the same samples allows for direct performance comparison between tests, highlighting significant difference in performance when the intervals for a specific diagnostic parameter do not overlap between tests.

9 General considerations when carrying out validation or verification

Some basic guidance to describe core requirements for validation and verification for routine laboratories is stated below. If multiple laboratory sites are involved, the requirements may vary. For more information, refer to Appendix 5 which is generic, and may not apply to every organisation. Bearing in mind that studies/projects may vary in size or represent collaborations between different organisations, this guidance can be applied as it fits the individual organisations' projects.

The core requirements are as follows:

- **IVD Regulation:** The IVD Regulation mandates the use of CE/UKCA marked IVDs for clinical diagnosis (1). Justified use of in-house IVDs must meet the requirements for exemption from the requirement to CE/UKCA mark as stipulated by MHRA (11). There must be a justification for applying the exemption, namely that the target patient group's specific needs cannot be met or cannot be met at the appropriate level of performance by an equivalent CE/UKCA marked IVD available on the market. The justification should include evidence (e.g. market surveys, proficiency panel data, and literature reviews) for the availability on the market of equivalent CE/UKCA marked IVDs.
- **Post-Manufacture Surveillance:** There will be post-manufacture (in-service) surveillance scheme for actively and systematically gathering, recording, and analysing relevant data on the quality, performance, and safety of a device throughout an assay's entire lifetime. This includes drawing necessary conclusions (including ongoing suitability for clinical diagnostic use in line with the validation) and determining, implementing and monitoring any preventive and corrective actions.
- **Personnel Identification:** Identify the personnel that will be responsible for the project, including the project manager and key staff in the laboratory involved in the project. The responsibilities of all personnel involved should be defined in the protocol.
- **Computing Requirements:** Identify computing requirements including software and hardware requirements, transfer or manipulation of data, IT, networking/interfacing and LIMS requirements.
- **Project Design Plan:** Prepare a project design plan and define the purpose and objectives of the project. This plan should include any training requirements, risk assessments and COSHH assessments, standard or reference materials, case definition, satisfactory sample size and study design (e.g., retrospective/prospective, case/control or cohort) expected/acceptable performance, ethical and governance regulations.
- **Commercial Companies:** Identify any potential commercial companies whose products may be used in the project
- **Potential Problems**: Identify any potential problems that may be encountered in the project and addressing these.
- **Cost Measurement:** Measure and compare all aspects of costs associated with implementation of the test, including validation/verification costs.
- **Time Scale:** Determine a suitable time scale for the project in advance before it commences. This depends on several factors such as statistical sample size, time needed to optimise test procedures, and the time required for approval.

- **Statistical sample size:** Validation or verification studies should include enough samples to provide confidence in test performance. Sample size should be risk-based and fit for purpose, considering the test's intended use and the impact of incorrect results on patient care (4,7).

  The required number of specimens depends on factors such as the performance characteristics being assessed (e.g. sensitivity and specificity), the desired precision, the smallest clinically important difference to detect, and the availability and prevalence of suitable samples.

  As sample size requirements vary by test, intended use, and confidence level, no fixed minimum can be specified. Statistical software may be used to estimate appropriate sample sizes.

- **Data collection:** Data collection methods should be kept as simple as possible to improve efficiency and accuracy. Data should be recorded at the bench and entered into record sheets, preferably using tick-boxes or simple keystrokes, with appropriate data entry audits in place.

  Where practicable, examination results should be read and recorded independently using de-identified specimens to minimise interpretation bias and reduce the influence of previously reported results.

  Clinical samples used within validation and verification studies should, wherever practicable, be de-identified to remove direct patient identifiers while retaining sufficient information to permit comparison with routine laboratory results and relevant clinical data.

  Appropriate anonymisation also helps avoid ethical and governance concerns that may arise when investigation results generated during validation or verification differ significantly from routine clinical results but have not yet been established as suitable for patient management or clinical reporting.

  However, at the analyses stage, previous test results, disease classification, relevant clinical information, and other study-specific metadata may be required to assess test performance and interpret study findings.

  This applies to both validation and verification studies.

- **Analysis of Results:** The statistical analyses to be performed should be defined in the study protocol (4). Testing of samples with the assay under assessment should be performed without knowledge of the disease/condition classification, other diagnostic evidence nor personal identifiers. Results should be evaluated against predefined acceptance criteria, considering factors such as analytical performance, clinical validation (e.g. sensitivity and specificity), measurement uncertainty where relevant, user acceptability, clinical relevance, investigation of false results, workload, costs, health economic impact, and available product information.

The study should include appropriate arrangements for data management, ethical and governance compliance, staff training and competency, quality control, and contingency planning to ensure data integrity, reliability, and confidentiality.

Where methods demonstrate comparable performance, additional benefits such as lower false-positive or false-negative rates, reduced costs, or decreased labour requirements should be considered.

This applies to validation and verification studies.

10 Documentation for validation and verification

All validation and verification activities must be documented in line with the local Quality Management System (QMS) and UKAS-accredited requirements. Documentation should be appropriate to the service, proportionate to the method, and sufficient to demonstrate fitness for purpose.

A validation/verification record (or file) should be maintained for new methods, modifications, and, where required, existing methods. This may consist of original records or clear cross-references to data held elsewhere within the QMS. All records should be controlled, traceable, auditable, and retained in accordance with local document control and retention policies.

Evidence included should reflect the nature and complexity of the method, but may typically consist of:

- Quality control and assessment data (internal and/or external)
- Method performance evidence and comparison, where applicable
- Relevant manufacturer's information or literature
- Risk, safety, and compliance documentation
- Supporting study data or reports
- Relevant SOPs and version control records

For in-house methods, documentation must demonstrate scientific validity, performance characteristics, and compliance with applicable regulatory requirements, including justification for any exemption from CE/UKCA marking, in line with MHRA guidance.

Summary reports and checklists should be completed where required by the local QMS; however, duplication of information is not necessary where adequate cross-referencing is in place. Review and authorisation must be undertaken in accordance with local governance arrangements prior to implementation or continued use.

All associated SOPs must align with current validated practice and be managed through the document control process.

All validation and verification records must remain accessible, legible, and available for review throughout the defined retention period (7).

11 Interpretation of results

As required in ISO 15189:2022 (2), where applicable, reports should include interpretation of results and comments that consider the clinical context, pre-test probability, and limitations of test performance.

An overall conclusion should be drawn based on the totality of evidence, stating whether the diagnostic test meets the defined performance requirements and is suitable for its intended use. This conclusion should be supported by the data generated and should clearly state any conditions, restrictions, or ongoing monitoring requirements associated with the test.

The predefined performance requirements and acceptance criteria should be established before the study commences and should be appropriate for the intended use of the test, clinical risk, regulatory requirements, and relevant professional guidance.

Approval to proceed to routine use, clinical implementation, or further regulatory submission should only be granted where results demonstrate acceptable, consistent, and clinically meaningful performance against predefined acceptance criteria and are shown to be fit for their intended use.

Some important points to consider:

1) Sample quality and suitability can compromise the clinical value of examination results
2) Discrepancies when examinations are performed by different procedures (e.g. POCT) or in different locations
3) Possible risk of misinterpretation when different units of measurement are in use regionally or nationally
4) Interpretation of results should take account of trends and significant changes over time.

Where a new quantitative method replaces an existing assay, laboratories should consider the potential impact of analytical bias and systematic differences between methods. Changes in assay performance characteristics may affect longitudinal comparison of patient results. For tests used to monitor disease progression or treatment response, laboratories should assess whether patient re-baselining, parallel testing, or additional interpretive guidance is required during the transition period.

Interpretation of results must always be undertaken in conjunction with available clinical information, including patient symptoms, disease status and severity, epidemiological risk factors, travel history, immune status, and exposure history.

Detection of a closely related or less common pathogen does not necessarily imply causation of disease, particularly where colonisation, transient carriage, or subclinical infection is recognised (non-sterile site samples). The range of detection of the assay under assessment needs careful consideration.

11.1 Reporting of closely related and less common pathogens

The reporting of closely related and less common pathogens requires careful consideration to ensure that results are scientifically accurate, clinically meaningful, and do not lead to misinterpretation, inappropriate clinical management and/or mis reporting of national data.

Results involving closely related or uncommon pathogens should be escalated or discussed with clinicians, infection specialists, or public health teams where interpretation may influence patient management, antimicrobial use, or infection prevention and control measures, or statutory/national reporting requirements.

Laboratories should ensure that reporting practices are consistent, risk-based, and aligned with national guidance and professional standards.

12 Point of care testing

Point-of-care testing (POCT) refers to medical diagnostic testing performed at or near the location of the patient, rather than in a centralised laboratory. POCT in microbiology can provide rapid results to support timely clinical decision-making and improve patient management. However, the decentralised nature of POCT means that test performance, quality assurance, and clinical utility must be carefully considered.

Prior to introduction, all POCT assays should undergo appropriate validation or verification, as applicable. This should be undertaken by the relevant staff against reference laboratory methods at the site of intended use to ensure accuracy, reliability and clinical relevance. Local verification, training of staff, and inclusion within quality management systems are essential to maintain performance standards. Regular review of POCT processes and results should be undertaken to safeguard patient safety and ensure compliance with national guidance.

In addition, POCT implementation needs to consider operator safety. The appropriateness of processing potentially infectious sample types on a bench on a ward which in the laboratory would be processed in CL2 or CL3 (e.g. processing respiratory samples). Validation and verification need to be performed using appropriate PPE at the site of intended use.

ISO 15189:2022 (2) include specific requirements for point-of-care testing that are critical for maintaining and enhancing quality and competence in laboratory services. The acceptance testing requirements in the standard give scope for various ways of testing, providing no patient result is released before the test is verified. Whilst it is

commonplace that non-lab staff perform POCT testing, it is important that the laboratory owns and controls the acceptance testing process.

Guidance on the use of POCT devices is available from the Medicines and Healthcare Products Regulatory Agency (MHRA) in the UK.

13 Next-generation sequencing in diagnostic microbiology

Next-generation sequencing (NGS) methods offer powerful capabilities for pathogen identification, characterisation, typing, antimicrobial resistance (AMR) prediction, and detection of organisms directly from clinical specimens.

Due to their complexity, multi-stage workflows, and reliance on bioinformatics, NGS assays require rigorous evaluation, validation, and verification before clinical introduction. This ensures that results are accurate, reliable, reproducible, and clinically meaningful, and that the end-to-end process meets ISO 15189:2022 (2) and MHRA regulatory requirements.

The process begins with a clearly defined intended use, including the disease context, specimen type, sequencing scope (range of detection), and classes of genomic variants to be detected.

Validation and verification studies must also account for challenging genomic regions, low-frequency variants, and specimen-related artifacts, particularly in degraded samples such as formalin-fixed paraffin-embedded tissue, where relevant for the test.

The performance of NGS has two main components: sample processing and bioinformatics. It is critically important that both components are included in the validation and verification of NGS based tests.

The requirement of provision of an interpretation of results to the user applies to NGS. The pre-test probability (symptoms, disease severity, epidemiological risk factors, travel history, immune status, and exposure history) needs to be taken into consideration when reporting positive results for non-sterile site samples.

The user must also be provided with information of the pathogen specific performance of the NGS method compared to other standard methods, such as targeted PCR, to be able to decide whether a negative result can be used to rule out infection.

Validation of NGS methods should include comparison with the current standard diagnostic approaches used for the specimen type and clinical syndrome under investigation. Where NGS is intended to replace or supplement existing methods, validation should include clinically relevant pathogens and compare performance against those methods currently used for their detection in routine practice (for example culture, targeted PCR, antigen detection or other established investigations). This enables determination of pathogen-specific diagnostic sensitivity and specificity

and allows clinicians to understand any differences in performance between the new and existing methods.

The use of syndromic meta-genomic assays capable of detecting a broad array of potential pathogens requires careful consideration when undertaking validation, and the inclusion of multiple cases of infection by common pathogens is required to determine diagnostic performance for individual disease entities. Claims that a method is more sensitive than existing approaches do not remove the requirement for validation against current diagnostic pathways. Additional investigations and patient record review may be required to determine the true disease status of the patient and the true performance (true/false positive/negative) of NGS for that sample type.

For these approaches, it is critical to understand how frontend process (sample volume, sample fraction, NA extraction) could influence the data output.

Appropriate reference materials, characterised isolates, and validated reference datasets should be used where available to support validation and verification of assay performance (7).

13.1 NGS validation requirements differ according to assay type

1. Targeted/amplicon-based sequencing - e.g. 16S rRNA, AMR gene panels
2. Whole genome sequencing of isolates – e.g. organism identification
3. Metagenomics pathogen detection – syndromic detection

Pre analytical considerations e.g. acceptable specimen types (such as isolates, CSF, BALF, plasma):

- minimum volume
- sample fraction (e.g., BAL supernatant or pellet)
- storage and transport stability
- integrity expectations (e.g., DNA fragmentation)

Nucleic acid extraction validation - for the full spectrum of targets:

- extraction recovery
- reproducibility
- inhibition
- contamination
- extraction controls

Definition of minimum input quantity e.g DNA/RNA mass, Ct thresholds for viral sequencing, DNA quality

- Validation needs to cover wet lab sequencing and dry lab analysis

- Analytical sensitivity and limit of detection for WGS, mNGS
- Analytical specificity and cross reactivity e.g. false positives from related organisms, host background interference, reagent contamination
- Coverage and sequencing performance metrics – coverage breadth and depth, library complexity
- Precision and reproducibility – intra run repeatability, inter run reproducibility. Inter operator reproducibility
- Contamination monitoring and controls
- Interpretation and clinical reporting
- Ongoing quality assurance
- Documentation and change control

13.2 Bioinformatics validation

Bioinformatics are used throughout infection diagnostic pathways from sequencing analysis, interpretation, reporting and surveillance to outbreak response and quality assurance:

- read trimming and quality filtering
- human read depletion (mNGS)
- alignment tools and parameters
- variant calling algorithms
- AMR gene detection tools/databases
- taxonomic classification algorithms
- phylogenetic reconstruction methods

Because NGS diagnostics rely extensively on computational analysis; validation and verification of the bioinformatics pipeline is equally important. This includes assessment of sequence alignment, variant calling, annotation, and reporting workflows.

Key performance metrics include sequencing depth, coverage uniformity, variant concordance, and false positive and false negative rates. Software versions, parameter settings, and database resources/limitations must be documented and controlled, with revalidation required following major pipeline modifications.

14 Algorithm-based decision-making and artificial Intelligence (AI) in diagnostic microbiology

Algorithm-based decision-making and artificial intelligence (AI) are increasingly being incorporated into diagnostic microbiology workflows to support laboratory decision-making. These technologies may assist with activities such as antimicrobial susceptibility interpretation, automated image analysis, digital culture reading, organism identification and metagenomic analysis. Regardless of the technology used, laboratories remain responsible for ensuring that examination procedures are validated or verified as appropriate, are fit for their intended clinical purpose, and continue to perform safely and effectively throughout their operational lifecycle (16).

AI should be clearly distinguished from algorithm-based systems.

Algorithm-based systems apply predefined computational rules or decision trees to generate consistent outputs from the same input data. These systems do not adapt after deployment. A common example in diagnostic microbiology is the automated interpretation of antimicrobial susceptibility testing using predefined MIC breakpoints. Other examples include rule-based reporting algorithms or bioinformatic pipelines that use fixed analytical criteria.

In contrast, AI systems use data-driven models, such as machine learning, to identify patterns and generate predictions. Examples already being introduced into microbiology laboratories include AI-assisted faecal parasite screening, automated urine culture analysis, and digital image analysis systems that screen culture plates for microbial growth, identify negative cultures and prioritise specimens requiring review. Some metagenomic pathogen detection pipelines may also incorporate machine learning components alongside conventional algorithms. Because AI systems rely on trained and adapting models rather than fixed interpretive rules, their performance changes following software updates, model retraining or changes to the underlying reference data.

The distinction between algorithm-based systems and AI is important because it determines the level of validation, verification and ongoing monitoring required.

14.1 Validation and verification

Before implementation, laboratories should demonstrate that an AI-enabled examination procedure is fit for its intended clinical use within the local laboratory setting. Validation or verification should be proportionate to the intended use, complexity of the system and the potential clinical impact of incorrect or misleading results.

Evaluation should include, where appropriate:

- assessment of analytical and diagnostic performance against predefined acceptance criteria

- comparison with an established reference method or expert laboratory interpretation
- evaluation using representative local clinical specimens and clinically relevant subgroups
- assessment of performance throughout the intended diagnostic workflow rather than the AI component alone
- identification and documentation of known limitations, sources of bias and situations where performance may be reduced (17).

For rule-based algorithms, validation or verification is generally straightforward because the decision rules are fixed and transparent. However, where those rules are modified—for example, through changes to interpretive criteria or updates to bioinformatic algorithms—the laboratory should assess whether revalidation or reverification is required before implementation.

14.2 Change control and documentation

The general principles of change control, risk assessment, validation/verification, record keeping and document control described elsewhere in this guidance apply equally to algorithm-based and AI-enabled examination procedures (16). Laboratories should maintain appropriate records of system versions, performance monitoring activities, changes, assessments and supporting evidence necessary to demonstrate continued fitness for purpose and compliance with the laboratory quality management system.

14.3 Ongoing performance monitoring

Unlike conventional algorithms, AI systems are designed to evolve over time and therefore require continued assurance that they remain fit for purpose (16).

Laboratories should establish procedures for ongoing monitoring that include:

- routine quality assurance and audit of AI performance
- review of discordant or unexpected results
- monitoring of false positive and false negative rates where appropriate
- comparison with expert laboratory interpretation or alternative validated methods
- investigation of performance trends that may indicate deterioration or drift over time.

This approach is analogous to ongoing competency assessment of laboratory staff and should form part of the laboratory quality management system.

14.4 Professional oversight

AI should support, rather than replace, the judgement of appropriately trained laboratory professionals (17). Laboratories remain responsible for all reported results, irrespective of whether AI contributed to the examination process.

Laboratories should ensure that staff understand the intended use and limitations of AI-assisted outputs, that clinically significant or unexpected results are subject to appropriate review, and that mechanisms exist for investigating discrepancies between AI-generated outputs and expert interpretation.

15 Public health responsibilities of diagnostic laboratories

Diagnostic laboratories have public health responsibility as part of their duties. Amongst these are additional local testing, or referral to further characterise the organism as required, primarily for public health purposes e.g. routine cryptosporidium detection; serotyping or microbial subtyping; and a duty to refer appropriate specimens and isolates of public health importance to a reference laboratory.

Diagnostic laboratory outputs inform public health intervention, and surveillance data is required to develop policy and guidance forming an essential component of healthcare. It is recognised that additional testing and referral of samples may entail some costs that has to be borne by the laboratory but in certain jurisdictions these costs are covered centrally.

Diagnostic laboratories should be mindful of the impact of laboratory investigations on public health and consider requests from the reference laboratories for specimen referral or enhanced information.

Appendix 1: Validation / Verification report summary form

Document Control/Document Title:

Local QMS Reference:
Department:
Prepared by:

Method Details:

Type: □ New method □ Modified method □ Validation □ Verification
Equipment:
Sample Types:
Test:
Procedure:
Manufacturer/Product Code(s): _____
Regulatory status: □ CE marked □ Not CE marked □ N/A

Brief description of the method:

Health and safety precautions
Data storage and analysis
Project timescale
Staff involved
Objectives
Define the acceptance criteria

Project team

Role (e.g. statistician, laboratory scientist, biomedical scientist)	Name	Laboratory	Area of expertise (e.g. statistics, molecular diagnostics, HIV serology)
Project Manager			
List other key individuals			

Scope, purpose and background of the method, including the reason for its introduction:

Brief details of the validation/verification plan:

Include manufacturer details and product codes (e.g., brand of culture medium or specific version of a PCR kit where multiple versions exist).

Relevant SOPs

Number	Title

Relevant COSHH and other risk assessments

Number	Title

Cross-reference all documents relevant to this study, including:

EQA data
IQC data
IQA data
In-house R&D records
Results of testing:
- known positives
- known negatives
- low positives
- high positives
- problem samples

Published and unpublished papers and reports
Workbooks (especially applying to in-house testing)
Work carried out with collaborating laboratories
Comparisons with alternative methods
Comparisons with previously used test methods
Evaluation reports
Review meeting minutes
Manufacturer's instructions
Manufacturer's product specification

Diary (include dates of all important events, such as review meetings)

Event	Date
Project start	

Results, Performance Characteristics and Limitations

Analytical performance

Diagnostic performance

PPV

NPV

Positive likelihood ratio

Negative likelihood ratio

Clinical accuracy

Expected results in normal and affected populations

Interpretation of results

3) Have customers been informed of any significant changes in method performance (e.g. sensitivity, specificity, bias or turn-around times)? Yes ☐ No ☐

Summarise the known limitations of the verified test or procedure:

- For verification, this may be addressed by reference to the existing SOP/literature provided by the manufacturer or developer. State any additional limitations identified during verification that are not described elsewhere
- include measurement uncertainty where relevant

Assessment of analytical bias, where applicable, including comparison with reference methods, peer group data, manufacturer claims, EQA/PT results, or previously validated methods. State the magnitude and direction of any observed bias and whether it is considered clinically acceptable. For quantitative monitoring assays, describe any mitigation measures required, such as parallel testing, clinician communication, or revised reporting guidance.

Summary of findings

Overall comments, including whether the test/device is considered fit for purpose.
Consider whether acceptance criteria were met.
Have any specific issues been identified?
Has the validation/verification demonstrated that the method or procedure meets the predefined acceptance criteria?

Validation/Verification Authorisation

Outcome ☐ Approved for use ☐ Not approved for use

Signed (Project Leader) Date_____

Introduction of method authorised

Signed (Project Manager) Date_____

Appendix 2: Diagnostic kit or reagent validation checklist

1) Regulatory status

Is the assay/reagent intended for distribution outside the organisation that created it? Yes ☐ No ☐

If YES, is the reagent/diagnostic kit CE marked? Yes ☐ No ☐

2) Method validation

i. Has a comparison with existing methods (currently or previously in use) taken place? Yes ☐ No ☐

ii. Has the performance in EQA and/or IQC schemes been evaluated? Yes ☐ No ☐

iii. Has the method been assessed by an IQA scheme? Yes ☐ No ☐

iv. Has all or part of this work been published in a peer-reviewed journal? Yes ☐ No ☐

v. Are there any other reports related to the method available, for example, project reports, manufacturer literature? Yes ☐ No ☐

vi. Are test results available for samples which challenge the performance of the method? Yes ☐ No ☐

vii. Are workbooks cross referenced? Yes ☐ No ☐

viii. Has the test been validated by collaborating laboratories? Yes ☐ No ☐

ix. Are comparisons with alternative methods available? Yes ☐ No ☐

x. Are evaluation reports available for this test? Yes ☐ No ☐

xi. Has the assay been costed? Yes ☐ No ☐

3) Have the following performance and safety characteristics been evaluated?

i. Sensitivity Yes ☐ No ☐

ii. Specificity Yes ☐ No ☐

iii. Precision/Reproducibility Yes ☐ No ☐

iv. COSHH assessment completed Yes ☐ No ☐

v. Risk assessment completed for new procedures, equipment and materials Yes ☐ No ☐

4) Documentation

Has the User Manual, SOP and relevant website documentation been updated? Yes ☐ No ☐ N/A ☐

Comments:

Appendix 3: Linearity

An example of linearity is illustrated in the graph below for a laboratory-developed test. Seven concentrations of the organism were prepared by dilution of a high concentration standard and tested in triplicate. This worked example has been reproduced from Burd et al. (18).

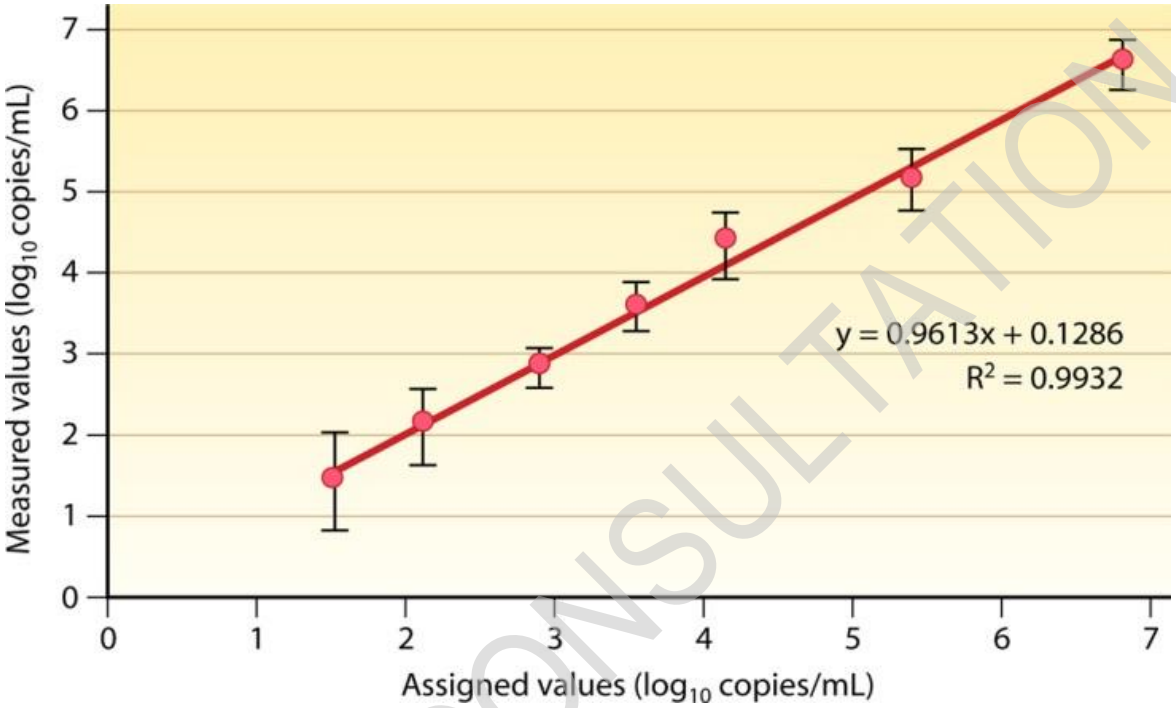


Figure 1: Plot of results from a linearity experiment to determine reportable range. Assigned values (converted to $\log_{10}$) were plotted on the x-axis versus measured values (converted to $\log_{10}$) on the y-axis using Microsoft Excel. The reportable range in this example translates to 30 copies/mL (lower limit of quantification) through 3,000,000 copies/mL (upper limit of quantification) (18).

Appendix 4: Principles for conducting validation and verification

Project manager

The Project manager is a designated individual who has overall responsibility for the development of a documented protocol, data analysis, preparation of the report, and formal sign-off. The Project manager should document any conclusions made based on the analysis of the results and sign a formal declaration that the method is suitable for diagnostic use. The report should have the support of stakeholders identified in the protocol design.

Project group

The project group should comprise competent and experienced personnel with clearly defined roles and responsibilities, ensuring all aspects of the method are adequately covered. The size and composition of the group should be appropriate to the complexity of the validation/verification, ranging from a single qualified individual to a multidisciplinary team. Where required, additional expertise (including external personnel or a statistician) should be included. Clear lines of accountability should be established, and for multi-site studies, a designated local coordinator at each site be appointed.

Computing requirements

Prior to validation or verification, the laboratory should assess information system capability and data handling processes, including LIMS availability, data transfer and entry requirements, and equipment interfacing. Consideration should also be given to software flexibility, system compatibility (particularly across sites), and the availability of appropriate statistical and data analysis tools.

Hardware requirements such as new instruments to run the test should be considered.

Project design (5,19)

The complexity of the project will depend on the method, with new or significantly modified methods requiring more extensive investigation than minor changes or verifications. Projects involving new in-house methods and/or considerable changes to a method already in use will require a more thorough investigation than minor changes to existing methods or performance verification.

During preparation of the protocol, it is important to ensure that assays / kits being evaluated are the same version as those currently marketed or about to be purchased. Assay kits changed at any time can invalidate results. Any modifications may not be clearly documented by manufacturers, so confirmation of version consistency is critical during protocol preparation.

The Project Manager should develop a documented plan that clearly defines the purpose and objectives of the validation/verification and ensures all personnel are appropriately trained and competent. Relevant risk and COSHH assessments must be completed.

Appropriate reference materials and a comparator method (“gold standard”) should be identified, with justification provided where none exists. The plan should specify the type and number of samples, ensuring they are representative (including positive, negative, and challenging samples) and suitable for the intended use.

Procedures for sample sourcing, handling, transport, and storage should be standardised, and ethical requirements addressed where applicable. The study should incorporate robust quality control measures, with clear arrangements for data collection, analysis (including statistical input where required), and documentation.

Progress should be reviewed at defined stages, and the study conducted in accordance with the approved protocol, with all results accurately recorded.

Involvement of commercial companies

Confidentiality agreements may be required when working with manufacturers, particularly for developmental products.

While scientific integrity must be maintained, manufacturers should be involved in study design to ensure correct use of the product and given the opportunity to review the protocol and comment on reports or publications.

Avoiding bias

Bias should be identified, minimised, and controlled at all stages of the validation/verification to ensure reliable results. This includes ensuring standardised procedures, independent interpretation of results, and appropriate, representative sample selection.

Additional measures include avoiding premature analysis or discussion of results, ensuring simultaneous testing in multi-site studies, and providing adequate training and competency assessment for all personnel prior to commencement.

Cost benefit analysis

All relevant costs can be assessed and compared with the specified gold standard, including consumables, equipment, maintenance, labour, and overheads. Consideration may also be given to the costs of testing, confirmation of results, and additional organism detection.

The overall impact on laboratory operations should be evaluated, including staff workload and time constraints. A cost–benefit assessment can then be undertaken, considering the clinical value of the test and its effect on turnaround times, with input from health economics where appropriate.

Time scale

An appropriate timeframe should be agreed in advance, taking into account sample size requirements, specimen availability, and approval processes. Studies should be conducted where sufficient samples are available, with consideration given to seasonal variation where relevant.

The study plan should include defined phases, such as initial familiarisation, extended testing with reference and routine samples, and final implementation in parallel with existing methods, to ensure a structured and efficient process.

References

An explanation of the reference assessment used is available in the scientific information section on the UK SMI website.

1. European Parliament and Council. In vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. pages 1–157. 2017. **++**

2. Institute BS. BS EN ISO 15189:2022+A11:2023 Medical laboratories. Requirements for quality and competence. BSI Standards Limited. 2023. **++**

3. Institute BS. BS EN ISO 17025:2017 General requirements for the competence of testing and calibration laboratories. BSI Standards Limited. 2017. **++**

4. Saunders N and others. Guidance on the development and validation of diagnostic tests that depend on nucleic acid amplification and detection. J Clin Virol 2013: volume 56, issue 3, pages 260–70. **+** https://pubmed.ncbi.nlm.nih.gov/23246357/

5. World Organisation for Animal Health (WOAH). Development and Optimisation of Nucleic acid detection assays. OIE Validation Guidelines; 2014. pages. 1–11. **−**

6. Rabenau HF and others. Verification and validation of diagnostic laboratory tests in clinical virology. JClinVirol 2007: volume 40, issue 2, pages 93–8. **+** https://pubmed.ncbi.nlm.nih.gov/17766174/

7. IBMS LabMed MHRA UKAS. New validation guidance published for medical laboratories. IBMS. 2026. **++** https://www.ibms.org/resource/new-validation-guidance-published-for-medical-laboratories.html

8. Institute BS. BS EN ISO 17822:2020 In vitro diagnostic test systems — Nucleic acid amplification-based examination procedures for detection and identification of microbial pathogens — Laboratory quality practice guide. BSI Standards Limited. 2020. **++**

9. Institute BS. ISO/TS 5798:2022 In vitro diagnostic test systems — Requirements and recommendations for detection of severe acute respiratory syndrome coronavirus 2 (SARSCoV-2) by nucleic acid amplification methods. BSI Standards Limited. 2022. **++**

10. Berwouts S and others. Approaches to quality management and accreditation in a genetic testing laboratory. EurJHumGenet 2010: volume 18 Suppl 1, pages S1–19. **−** https://pubmed.ncbi.nlm.nih.gov/20720559/

11. Medical Device Coordination Group (MDCG). MDCG 2023-1 Guidance on the health institution exemption under Article 5(5) of Regulation (EU) 2017/745 and

Regulation (EU) 2017/746 2023. **++** https://health.ec.europa.eu/medical-devices-sector/new-regulations/guidance-mdcg-endorsed-documents-and-other-guidance_en

12. National Association of Testing Authorities. Guidelines for the validation and verification of quantitative and qualitative test methods. Technical Note 17. pages 1–32. 01/10/2013. **+**

13. Zhang Z, Tian L. Validation of mNGS results using extensive lab and clinical data. BMC Microbiology 2025: volume 25, issue 1, pages 173. **2+** https://pmc.ncbi.nlm.nih.gov/articles/PMC11951646/

14. World Organisation for Animal Health (WOAH). Principles and methods of validation of diagnostic assays for infectious diseases. WOAH Terrestrial Manual; 2023. pages. 1–27. **++**

15. Institute BS. ISO 20776-2:2021 Clinical laboratory testing and in vitro diagnostic test systems. Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices. Part 2: Evaluation of performance of antimicrobial susceptibility test devices against reference broth microdilution. BSI Standards Limited. 2021. **++**

16. International Organization for Standardization/International Electrotechnical Commission. ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. ISO/IEC. 2023. **++**

17. World Health Organization. Ethics and Governance of Artificial Intelligence for Health. WHO. pages 150. 2021. **++** https://www.who.int/publications/i/item/9789240029200

18. Burd EM. Validation of laboratory-developed molecular assays for infectious diseases. Clin Microbiol Rev 2010: volume 23, issue 3, pages 550–76.**+** https://www.ncbi.nlm.nih.gov/pubmed/20610823

19. World Organisation for Animal Health (WOAH). Principles and Methods of Validation of diagnostic assays for infectious diseases. OIE Validation Guidelines; 2019. pages. 1–18. **+**