



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Validating Real-Time Polymerase Chain Reaction (PCR) Assays

Melvyn Smith, Viapath Analytics, Specialist Virology Centre, King's College NHS Foundation Trust, London, United Kingdom

© 2021 Published by Elsevier Ltd.

Abbreviations

CLIA United States clinical laboratory improvement amendments. This is a regulatory body or program to ensure quality laboratory testing.

Ct Crossing threshold, the cycle number at which the PCR enters the exponential phase.

FDA United States food and drug administration. The FDA is a Federal agency responsible for protecting the public health through the control and supervision of medical devices, the FDA are one of the agencies responsible for developing and implementing CLIA.

IVD In-vitro diagnostic products. These are the reagents or instruments used to diagnose a disease, in this case the PCR assay.

LDT Laboratory developed test.

LOD Limit of detection, the lowest amount of analyte detected in an assay.

rt PCR Real-time polymerase chain reaction.

RT-PCR Reverse transcription PCR.

ΔRn The maximum difference in fluorescence between the start and end of the PCR.

Introduction

For many years, the development of assays took place in the laboratories where the test was required, the so-called in-house or LDT. Although more and more commercially-developed tests have become available, novel assays continue to be developed in academic hospital laboratories. These laboratories are more able to respond quickly to new and re-emerging infections and crucially have the samples necessary to develop the test. This was the case with the new human coronavirus causing clusters of pneumonia epidemiologically linked to a seafood market in Wuhan, China in December 2019. Researchers from Shanghai, Wuhan, Beijing and Sydney sequenced a sample from a patient who had worked at the market; the sequence was deposited in the publicly available GenBank database on 10th January. Further sequencing showed the virus shared 85% identity with a SARS-like coronavirus found in bats. The first LDT was published on 23rd January 2020, followed a month later by the first of the commercial tests.

Although often more expensive than LDTs, commercial kits enable the rapid introduction of new tests, with the further advantage of being CE marked or FDA approved, providing some form of reassurance to the laboratory. However, CE marking is only a declaration of compliance with European legislative requirements; it does not necessarily guarantee the rigorous validation of the assay, or give any details of how and when a test should be used. Furthermore, commercial assays, by definition, have to be commercially attractive. Therefore, specialist, small-scale tests for rarely occurring infectious pathogens will not be cost-effective. Thus, there will continue to be a need for laboratory-developed assays.

The Need for Validation

In-House Assays

Over the years many laboratories have established methodologies for validating their assays. However, the literature continues to show a lack of detail in some critical areas, e.g., optimization of extraction techniques, methods used in primer and probe design, no evidence of amplicon sequencing to confirm specificity, imprecise estimates of sensitivity and specificity and assays that do not include internal or extraction controls. Such lack of detailed experimental information makes assessing the clinical utility of the assay difficult. Some of these problems lie with the scientific literatures' approach to publishing, where space limitations restrict the amount of detail permitted in a paper. These difficulties have led to a number of papers calling for an improvement in the standards of reporting diagnostic assays, including the STARD initiative (standards for reporting of diagnostic accuracy) and the MIQE guidelines (the minimum information for publication of quantitative real-time PCR experiments; [Bustin et al., 2009](#)). In addition to the fundamental requirements of good scientific practice, there are a number of regulatory bodies that require assays to be validated to certain standards, such as the FDA and CLIA requirements in the USA and the IVD Regulations (EU) 2017/7462017 in Europe. There is also an obligation on health institutions to be accredited according to the ISO 15189 standard. There has been an ongoing debate in Europe regarding CE marking of LDTs, similarly in the USA the FDA announced its intention to shift from a policy of enforcement discretion to exercising regulatory oversight at a future date on LDTs. However, the imposition of rigorous controls on the development of assays must not stifle innovation and the ability of front-line laboratories to respond quickly to new and emerging threats. Under these circumstances it makes sense for laboratories developing and publishing assays to ensure

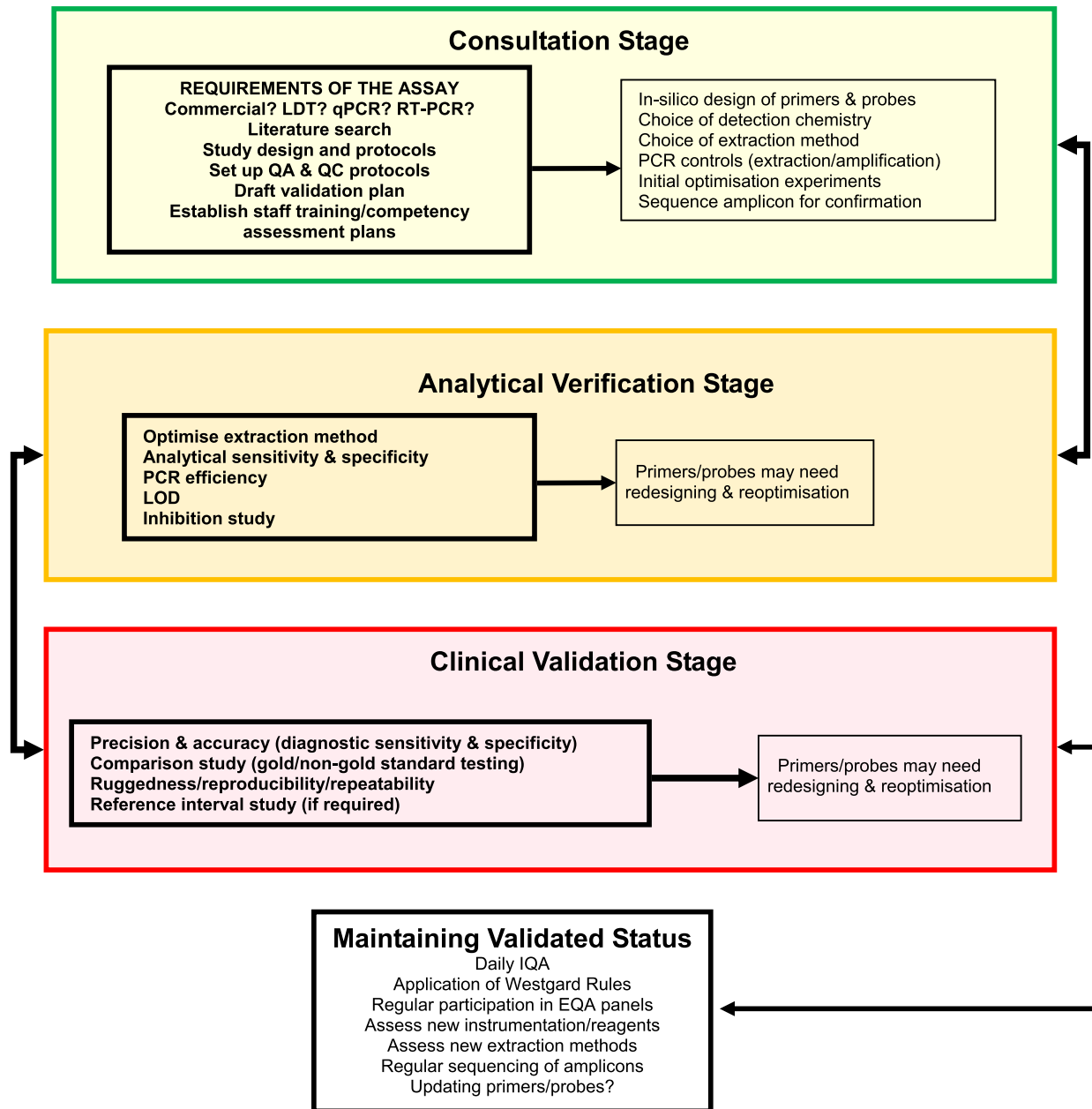


Fig. 1 Assay validation: Key concepts.

that they produce rigorously documented experimental proof to support any potential CE-marking/IVDD requirements that may be necessary in the future.

Commercial Assays

Although commercial assays are typically CE marked or FDA approved this does not necessarily mean the assay has been validated for all applications. Furthermore, although the assay may perform acceptably in the commercial developer's laboratory, a number of factors can affect the assay's performance elsewhere, e.g., staff competences and workflow systems, where not all laboratories may have separate working areas recommended to reduce contamination. There may also be differences in equipment maintenance schedules, including anything from freezers and pipettes to thermal cyclers, which can fundamentally affect the assay. Therefore, it is necessary for laboratories to verify the stated criteria. Although there are no formal requirements in the UK for laboratories to evaluate newly-introduced commercial assays, in the USA, FDA-approved, FDA-modified (i.e., tests modified from the manufacturer's

instructions but with performance characteristics determined by the user laboratory, consistent with CLIA requirements) and LDTs require verification before introducing the test into the clinical laboratory. CLIA stipulates that prior to implementation of an FDA-cleared test laboratories must verify the manufacturer's performance specifications. Where a FDA-modified test or LDT is to be introduced, CLIA stipulates that, additionally, analytical sensitivity and specificity must be established.

In an attempt to clarify the situation, it is the intention in this article to provide a set of basic working guidelines for the validation process. These guidelines can be applied throughout the continuous process of maintaining the validated status of an assay. The key concepts in the process of validating an assay are illustrated in [Fig. 1](#).

The Validation Process: Consultation Stage

The process begins with the development of a validation plan and involves decisions based on the clinical need for the assay, e.g., epidemiological studies, infection control or screening. As discussed previously, there are differences between the levels of validation laboratories perform when introducing commercial or LDTs. As a minimum, the laboratory introducing a commercial test must establish that the manufacturer's performance claims can be reproduced. Once the performance characteristics of an assay have been met, whether commercial or LDT, the validation exercise must continue on a daily basis (see [Fig. 1](#)). This involves continually monitoring the levels of internal and external positive controls to ensure the validation status of the assay is maintained.

Preliminary Considerations

The initial step is to define the purpose of the assay; all the subsequent steps in the validation process are guided by this decision. The following three variables that can affect an assay's performance must be considered at this stage:

- (1) The sample-type and the host/pathogen interactions that determine whether a qualitative or quantitative assay is required.
- (2) The assay system: the biological, technical and operator-related factors that affect the assay's ability to detect the target in the specific sample-type. All assays are considered multiplex, since they must include a co-amplified extraction control.
- (3) The result: will the result accurately predict the status of an individual or population in regard to the analyte detected?

The sample-type (e.g., tissue, whole blood, CSF) may contain inhibitors that affect the activity of the polymerase in the PCR. Therefore, the extraction process needs to be evaluated and necessary alternatives considered. Quantification assays are useful in the management of patients, where pathogen load can be monitored during therapy. Some viral pathogens, such as EBV and CMV can establish a primary low-level latent infection and subsequently become reactivated, switching to a productive lytic infection. Is more than one pathogen to be identified in a multiplex assay? Another significant factor is the availability of sufficient numbers of well-characterized positive control samples to enable the validation. In order to maintain objectivity, the method chosen to resolve discrepant results must be established as part of the validation plan before testing begins.

A quality assurance plan must be established for the assay. Consideration must also be given to the availability of external QA reagents. This can be a problem with new assays targeting rare pathogens, where QA panels are unlikely to be available; the laboratory may need to consider working with the providers to produce suitable reagents. Once the requirement for a test has been established, the next decision is whether to use a commercial assay or develop a LDT. In the USA the CLIA specify that laboratories using FDA-cleared assays must verify that the manufacturer's performance specifications for accuracy, precision, reportable range and reference intervals can be replicated. For modified commercial assays the following should also be tested:

- (1) Analytical Sensitivity (LOD).
- (2) Analytical Specificity to include inhibitory substances.
- (3) Any other performance characteristics required for test performance.

When assessing a commercial assay, the extraction process also needs to be verified. In most cases the manufacturer's protocol includes details on the extraction process to use. However, the assay may need to be validated with a number of different extraction methods, depending on the type of equipment available.

The Validation plan

Verification: the process of establishing whether the individual components of an assay meet the analytical performance requirements established at the start of the development process.

Validation: the process of ensuring that the completed assay conforms to the users' needs, requirements, and/or specifications under defined operating conditions.

There are numerous aspects of an assay that need to be continuously monitored throughout its use in the laboratory. Micro-organisms, particularly viruses, often mutate. This means that the efficiency of the PCR needs to be monitored for potential false-negative results, as this may be the first sign that the primers and/or probe need to be updated and revalidated. Manufacturers continually develop new

Table 1 Theoretical number of samples from subjects of known infection status required for establishing diagnostic specificity and sensitivity estimates using likely estimated specificity/sensitivity value and desired error margin and confidence

Estimated specificity/sensitivity value	2% error in estimated specificity/sensitivity			5% error in estimated specificity/sensitivity		
	Confidence			Confidence		
	90%	95%	99%	90%	95%	99%
90%	610	864	1493	98	138	239
92%	466	707	1221	75	113	195
94%	382	542	935	61	87	150
95%	372	456	788	60	73	126
96%	260	369	637	42	59	102
97%	197	279	483	32	45	77
98%	133	188	325	21	30	52
99%	67	95	164	11	15	26

buffers, enzymes and extraction kits, these must be assessed and if adopted, re-verified, and the assay revalidated. New types of probe chemistry are constantly being introduced and these, too, need testing to maintain or improve the assay's performance.

Although each new assay may present its own particular challenges, the basic criteria for validation, such as specificity, sensitivity and reproducibility apply and any specific differences can be incorporated without difficulty. The first steps concern decisions based on the clinical need for the test, i.e., the scope, purpose and application of the assay and whether a commercial or a LDT will be required. For all in-house developments it is strongly recommended that the comprehensive MIQE guidelines are followed.

Analytical Verification Stage

Reference Materials and Sample Numbers

The first question that arises when developing a LDT, particularly for rare and emerging infectious disease pathogens, is the availability of samples. If sufficient samples are not available, are they obtainable elsewhere? If not, it may be necessary to construct test samples by spiking various concentrations of the analyte into a suitable matrix. Other sources of suitable samples may be other clinical/research laboratories or commercial standards, quality control materials or proficiency panels. Typically, 100 samples of 50–80 positive and 20–50 negative specimens are used. Consideration also needs to be given to potential inhibitory substances likely to be found in the specimens tested. Therefore, paired control specimens should be prepared by adding low concentrations of the analyte, with and without the known inhibitors, to suitable negative samples. However, such artificially constructed samples are unlikely to have the same properties as clinical samples. Therefore, when sufficient numbers of genuine samples become available, the assay and extraction methods will need to be re-assessed. Where samples are available, more than one specimen type may be required, e.g., respiratory pathogens can be extracted from nasopharyngeal aspirates, bronchoalveolar lavages or nose and throat swabs. A literature search should always be carried out to determine the type of specimens to be assessed and the most efficient extraction method for each of the sample type determined.

When sufficient samples numbers are available, the numbers need to be calculated. Ideally this should be determined by statistical analyzes, where the sample size required to detect a significant difference is determined by the standard deviation from the difference of the means of paired samples used in the test under development and the gold standard comparator. Alternatively, they can be estimated from [Table 1](#) below. This method allows for either a 2% or a 5% calculated error in diagnostic sensitivity and specificity. So, for example, if a 2% error is assumed, the number of samples required for an assay with 99% confidence and 99% estimated sensitivity/specificity is 164. However, if the same assay achieved a 95% estimated sensitivity/specificity then the samples required to achieve a high confidence (99%) would be 788.

Template, Primers and Probes: *In-Silico* Design

A thorough literature review is first carried out to identify a suitable locus to be amplified. It may be necessary to consider using degenerate primers and possibly probes to cover sequence variants that occur in different strains of the pathogen. Thorough sequence searches with BLAST (Basic Local Alignment Search Tool; See "Relevant Websites section") are required to ensure as many variants as possible are included in the design. A local database can be constructed and aligned using one (or more) of the sequence alignment program available, such as CLUSTALW (See "Relevant Websites section"). Optimal primer and probe function is critical to successful PCR. There a number of software packages available to assist in primer and probe design, such as Primer Express (Applied Biosystems) and OLIGO (See "Relevant Websites section"). If the assay is a multiplex, the sets of primers and probes must be checked for cross-reactivity, to eliminate any inhibition due to co-amplification issues. Multiplexing is a useful method of syndromic testing and reducing the cost diagnoses by targeting more than one pathogen in a sample. Furthermore, most, if not all assays will be multiplexed, due to the co-amplified EC.

Any matches with other sequences, particularly from other pathogens likely to be encountered in the sample types under investigation will necessitate redesign. It must be stressed that this preliminary specificity check against the databases is not conclusive evidence. The assay itself must be tested against the range of pathogens likely to be encountered in the sample types under investigation.

Choice of the Quantification Standards

There are two main options for quantifying by a real time PCR, either absolute or relative. In most diagnostic assays absolute quantification is chosen. A set of separately amplified, log-diluted previously quantified templates are used to form a standard curve, derived from Ct values plotted against the log concentrations of each standard. The copy number of any sample is estimated from its Ct intercept on the standard curve. The accuracy of the standards can be confirmed by testing against external QA controls. The National Institute for Biological Standards and Control in the UK produce over 300 WHO International standards. Several standards designed for nucleic acid based assay testing are also commercially available, including HIV, hepatitis, CMV, EBV, and B19 parvovirus and are reportable in international units/milliliter (IU/ml).

Reaction Controls

The LDT must include positive controls (PC), negative, no template controls (NTC) and extraction controls (EC). The Ct values of the positive control must be recorded as part of the routine quality assurance. Typically, a single positive control for each specific pathogen, a NTC after every eight samples, and a further NTC as the last sample on the run is used to monitor the assay. Placing a NTC at the start of the set-up could give a false impression of any contamination problems. The PC can be produced from extracted clinical samples or from commercial sources and should be diluted to be reproducibly amplified at the lowest detectable level (typically a Ct of 30). An internal PCR or amplification control (IC) is an absolute requirement for any diagnostic assay. Nucleic acid from a suitable non-target pathogen can be spiked into each extracted specimen prior to the PCR and amplified and detected with a separate set of primers and probe. Preferably, an EC, using a non-related virus or bacterium is used to monitor extraction efficiency. Using this approach, the EC is spiked into the sample before extraction. This has the advantage over using naked nucleic acid as an EC, because it controls for the all the steps in the extraction process. Phocine herpesvirus is a useful EC for assays detecting viral DNA pathogens and for RNA viruses, MS2 phage or mengovirus can be used. The use of an RNA template in RT-PCR controls for both the critical reverse transcription step and the PCR. Inhibition can be assessed by comparing the Ct values of the amplified EC in samples with the Ct value of the control extracted in a negative matrix, e.g., water or elution buffer. A difference, typically greater than three Cts (approximately equivalent to one log) indicates inhibition and a potentially false negative result. It is necessary to establish that the primers and probes and the amount of control added to the reaction does not interfere with the amplification of the pathogen target.

Reverse Transcription PCR

Reverse transcription PCR is used to detect RNA viruses, but the use of differential gene-expression assays is becoming more widespread to distinguish between infection and colonization in other infectious micro-organisms. Reverse transcription-PCR can be carried out as a two-step reaction, where the RT step is carried out separately and an aliquot of the reaction transferred to the PCR. More usually, in infectious disease assays, a one-step reaction is employed, where the reverse transcription and PCR occur in the same tube on the thermocycler. The one-step is quicker and more suited to high-throughput diagnostic laboratories, since the potential for contamination is lessened by number of steps that require opening tubes. The choice of priming the RT step needs to be considered; typically, the downstream, anti-sense primer is used. However, random hexamers may be more efficient. Often the choice is down to a sensitivity/efficiency balance between the hexamers ability to bind and hence copy all the RNA species in the sample and the sensitivity provided by the sequence-specific binding of the downstream primer.

Table 2 The Basic Westgard Rules

A	1_{2SD}	If one control measure exceeds the mean ± 2 SD, the control values from the previous run should be considered to rule out a trend.
B	2_{2SD}	This rule detects systematic error. The rule is violated when two consecutive control values exceed the same mean ± 2 SD limit.
C	4_{1SD}	This rule detects systematic error. The rule is violated when 4 consecutive control values exceed the same (mean $+1$ SD or mean -1 SD) limit. The run need not be rejected if this rule is violated but should trigger recalibration or equipment maintenance.
D	1_{3SD}	This rule detects random error. Violation of this rule may also point to systematic error. The assay run is rejected if one control value exceeds the mean ± 3 SD.
E	R_{4SD}	This is a range rule and it detects random error. The rule is violated when the difference in the SD between 2 control values exceeds 4 SD (i.e., -2 SD and $+2$ SD).
F	10_x	This rule detects systematic error. The rule is violated when the last 10 consecutive values are on the same side of the mean. Its violation often indicates the deterioration of assay reagents.

Westgard Rules

For each run the Ct values of controls are plotted on Levy-Jennings or Shewhart charts to monitor the performance of the assay; visual presentation of the control data is helpful in spotting variations in the operation of the assay (see [Table 2](#)). Westgard rules can also be applied to determine when a corrective action should be taken and give some indication of the nature of the analytical error. The use of Westgard rules and Shewhart charts must form part of the ongoing monitoring of the assay's validated status.

Experimental Optimization

Optimal primer and probe concentrations must be verified, together with the PCR program itself. A well characterized, stable positive control must be used throughout this verification process to provide an invariant baseline from which the effects of changes to the various components can be measured. The formation of primer dimers and the degree of non-specific amplification can be assessed by analytical agarose gel electrophoresis which also provides evidence of the size of the amplicon. To optimize primer and probe concentrations, ideally two log-dilutions of the positive control are used in a checkerboard layout of varying primer/probe combinations. The optimal concentrations and thermal cycling conditions are those that provide the lowest Ct values, the greatest ΔR_n and a consistent Ct difference of approximately three cycles between the two log-diluted samples. Once the basic parameters of the assay have been established it will be necessary to confirm the amplicon product by sequencing and BLAST analysis. Due to the short amplicon lengths in real-time PCR, the fragment may need to be cloned into a suitable plasmid vector.

Normalization

Expression analyzes with reverse transcription PCRs is not normally used in diagnostic PCR. However, the range of diseases and their associated pathogens being diagnosed by RT-PCR is increasing. In some cases, e.g., with fungal species, there is a need to distinguish between carriage, environmental contamination and infection. Therefore, assays capable of detecting and quantifying differentially expressed genes indicative of the disease condition are needed. This requires carefully selected reference genes to control for variations in extraction, reverse-transcription and amplification efficiencies, so that comparisons across different mRNA concentrations can be made and fold changes in expression levels determined. A detailed literature search for suitable reference gene mRNA targets must be made and the selected candidates experimentally tested for stable expression in both diseased and non-diseased samples.

Analytical Specificity and Sensitivity

Analytical specificity: the LDT's ability to detect the target it was designed for and not cross-react with other analytes in the sample.

Specificity is demonstrated either by spiking samples with a range of different pathogens prior to extraction, or adding the extracted nucleic acid from these pathogens to the extracted sample under investigation. All reactions should be analyzed by agarose gel electrophoresis to ensure that amplification has not occurred in any of the samples expected to be negative. In some cases, primers may amplify non-target templates, which may not be detected by the probe. This will compromise the sensitivity of the assay and must be eliminated, either by re-evaluating the PCR conditions (typically raising the annealing temperature), or preferably by redesigning the primers.

Analytical sensitivity: The LDT's ability to detect very low levels of a given analyte in a biological specimen.

This is synonymous with the assay's LOD, i.e., the lowest concentration of the analyte consistently detected in $\geq 95\%$ of samples tested with acceptable precision. The LOD is usually determined for diagnostic assays by using a set of log-diluted controls, such as patient samples, a suitable cell line, or proficiency panels. It is important to include a NTC in the LOD study to ensure that the PCR does not generate a signal that could interfere with true low-level positive signals from a sample.

There is no hard and fast rule as to how many samples to use, The Clinical and Laboratory Standards Institute suggests a minimum of 60 data-points (12 separate measurements from each of 5 samples) are required from a manufacturer to establish the LOD. This seems a reasonable figure to adopt as a standard approach for both LDTs and commercially produced assays and will provide the necessary intra-assay variation measurements. This measurement should also be carried out on a second thermocycler to establish machine-to-machine variation.

Estimation of the LOD is usually carried out by probit analysis, a type of regression used to analyze binomial response variables, where a sigmoid dose-response curve is transformed into a straight line that can be analyzed by regression either through least squares or maximum likelihood. For PCR assays the lowest concentration of analyte that can be detected with a stated probability can be determined by plotting the data from positive replicate results versus the analyte concentration.

Inhibition Study

The factors that inhibit or prevent the amplification of nucleic acids by PCR can be present in the extracts from a number of sources. Inhibitors typically act by: (1) interference with the cell lysis necessary for extraction of nucleic acids, (2) interference by degrading the

nucleic acid or inhibiting its capture, (3) inhibiting polymerase activity during amplification of the target DNA, resulting in false-negative results or inaccurate quantification. Different sample types present different problems due to the range of endogenous inhibitors found, for example in fecal samples, complex polysaccharides, breakdown products from hemoglobin, and bile acids can all be present and careful thought needs to be given to the method of extraction. Other potential inhibitors include metabolites resulting from pathological conditions such as diabetes mellitus and homeostatic hepatitis, or from medications used in treatment.

Manufacturers continue to develop their PCR reagents specifically to overcome many of the common endogenous inhibitors found in clinical specimens and it is recommended that a number of different commercial amplification kits are evaluated for their performance in this respect. As discussed previously, inhibitors can be detected using an IC. If the IC fails to amplify or exhibits suppression below the acceptable threshold, the amplification of the intended target sequence may also be inhibited. To avoid reducing the sensitivity of the assays, the IC should be used at the lowest concentrations that can be reproducibly amplified to minimize any competition between its amplification and that of the target pathogen.

The IC must demonstrate inhibition by the substances expected to be found in the sample, for which an inhibition study needs to be carried out. The study must first identify what types of inhibitors are likely to be present in the sample types used in the assay. The IC can be tested in negative samples with and without the interfering substance(s) in parallel and the effect on the C_t values monitored. This test must be repeated with the full assay (including the pathogen detection primers and probes) using paired samples, with and without the target analyte to fully validate the assay. Once interference is found, of either or both the IC and target, samples can be serially diluted to test the limit at which the target will amplify. In most cases dilution of inhibited samples provides a straightforward method of enabling amplification. The assay must then be tested against clinical samples; this is because artificially constructed inhibition control samples may not accurately reflect the assays performance in routine use.

PCR Efficiency

The efficiency of the PCR can have a significant impact on the robustness and precision of the assay. The efficiency of the reaction is determined by a number of factors, including primer design, cycling conditions and the reagents used in the reaction mixture. PCR efficiency is particularly important in assays reporting fold changes of mRNA for target genes relative to those of reference genes, where both templates must be amplified with equal efficiencies. PCR amplification efficiency is typically established using calibration curves and it makes sense to use the dilution series (covering 5–6 orders of magnitude) from the LOD analyzes for this purpose. The equation of the linear regression line, together with Pearson's correlation coefficient (r) and the coefficient of determination (r^2) are used to determine amplification efficiency. Amplification efficiency itself is determined from the slope of the log-linear portion of the calibration curve and is given by Eq. (1):

$$E = 10^{(-1/\text{slope})}. \quad (1)$$

Ideally the amount of template will double during each round of exponential amplification, this translates to a reaction efficiency of 2, therefore, using equation 1: $2 = 10^{(-1/\text{slope})}$ which gives a slope for the standard curve of -3.32 . This ideal figure of 3.32 also represents the difference in cycle number for each log dilution in the series. The efficiency can also be expressed as a percentage of the template amplified in each cycle using Eq. (2):

$$\% \text{ Efficiency} = (E - 1) \times 100\% \quad (2)$$

In the ideal example given above; $\% \text{ Efficiency} = (2 - 1) \times 100\% = 100\%$.

The experimental measurement should be carried out in triplicate on at least two thermal cyclers and the information included in the validation documentation. In practice acceptable assays should achieve efficiencies of 90%–100%. Reaction efficiencies lower than this may be caused by suboptimal reaction conditions or poor primer design. Reaction efficiencies of more than 100% could be due to measurement errors e.g., in preparation of the dilution series, or co-amplification of primer-dimers. Improving specificity will improve the sensitivity and increase the dynamic range of the assay, by reducing competition between the specific and non-specific amplification products.

Linear Dynamic Range or Reportable Range

The linear dynamic range can be determined during the LOD analysis by extending the range of the dilution series. The dilution series is often set at the range one would expect the analyte to be found in clinical specimens, although this may not always be known, especially for a novel pathogen and so initial testing can be carried out from 1 up to 10^7 or higher copies/ml of target. If using a generic assay to detect a range of pathogens, e.g., a pan-fungal PCR, each of the species the assay is designed to detect must be tested. Similarly, for multiplex assays, the LOD must be determined individually for each target analyte in the assay and must include the individual targets in the multiplex at high and low concentrations to ensure competitive amplification does not prevent all targets being identified. The reportable range is defined as the lowest and highest results (in suitable units of concentration) reliably detected in the assay. The linearity of the range can be established by calibration curves, where an r^2 value (coefficient of determination) of 1.00 indicates a perfect fit of the data points. Generally, in qPCR assays an r^2 of not less than 0.99 is considered acceptable. The linearity and reportable range should be carried out on at least 5 log-dilutions of the target nucleic

acid extracted from an appropriate sample type (serum, urine, NPA etc.) in triplicate, ideally on two different thermocyclers. Quantitative results from the test can only be reported when they fall within the linear range of the assay. A reportable range is not applicable to qualitative assays, since the result is simply positive or negative (below the LOD). However, a log-diluted panel should be tested, because the information derived is useful in assessing both the efficiency and clinical utility of the test and the results (as C_t values) reported as part of the supporting documentation for the assay.

Clinical Validation Stage

Precision, Accuracy and Trueness

Precision: a measure of the closeness of agreement between independent test results obtained under defined conditions.

Accuracy: the level of agreement between the true value of an analyte and the value obtained by the new test.

Trueness: defines the level of agreement between the average value obtained from a large series, the test and the accepted reference value.

Precision experiments are designed to measure the random error of an assay over a pre-determined period of time by multiple measurements of an aliquot derived from a homogeneous sample. For qualitative assays random error can be established by testing the PC and NC material in triplicate over a period of 10 days on two different thermocyclers. For qPCR assays estimates can be determined from the quantification standards and two positive controls at the lower and upper ends of the assay's reportable range.

Accuracy refers to the closeness in agreement between a single measurement and the true value of the analyte under investigation. Trueness is determined by analyzing the average value obtained from a series of measurements with the new assay and the true value of the analyte (if an international standard is available) or a reference method (if a standard is not available). There are two approaches to evaluating trueness (Burd, 2010):

- (1) A comparison of methods study, where split samples are tested in parallel with an appropriate gold standard method.
- (2) A recovery study, where proficiency samples, or other verified sample types are compared with the assay under development.

Although either of these methods is acceptable, a split sample, comparative study using clinical samples is preferred. However, in some cases, e.g., when developing an assay for a novel pathogen, a suitable gold standard assay may not be available and a recovery study will have to suffice.

The Gold Standard Comparative Study

The comparative study is the cornerstone of the validation exercise. Estimates of sensitivity and specificity are derived from comparisons between the LDT and an established gold standard, which, ideally, is the same type of test and has an assumed sensitivity and specificity of 100%. By definition the gold standard assay is an error-free diagnostic method, which, rarely (if ever) exists. Consequently, an earlier PCR assay, or a culture-based method is often used.

Samples for the comparative study should be representative of the disease and population being investigated and also suitably distributed across both age range and gender. Where only limited samples are available, it may be necessary to use archived positive and negative specimens, or cultured material spiked into negative clinical samples. Other sources may be commercial standards, quality control materials or proficiency panels. Sample numbers can be determined as described previously. However, the sample numbers are more often selected for more pragmatic reasons, such as cost and feasibility and therefore, less than the statistically optimal number will be tested. This is often the case for a new or re-emerging infectious disease.

The difficulty comes in interpreting the results of a study where the comparative test is not a "perfect" gold standard, (i.e., the alloyed standard). This is particularly so if the comparative method is cell culture, which is thought to be 100% specific, but provides less than optimal sensitivity. The strength of the PCR lies in its sensitivity (theoretically a single DNA copy). Thus a newly developed PCR assay may produce many more positive results than the reference method, leading to a misleading, lower estimate of specificity. The researcher is then left with the problem of deciding whether these are true or false positives. A thorough and detailed assessment of the sample derivation, clinical history and sequencing can help to determine the true status of any sample, together with discrepant analysis. In discrepant analysis the discordant results are resolved by a third test, such as another PCR directed to a different gene-target and the results from this resolver test are used to definitively assign the final results. It is important to include a number of randomly selected samples with concordant results for discrepant analysis so as to reduce bias.

Further statistical analysis can be carried out to improve confidence in the results from the new LDT, particularly when the true disease status of the samples is difficult to evaluate. The agreement between the new test and comparator test can be expressed as the

New Test	Gold Standard		Totals
	Positive	Negative	
Positive	TP	FP	TP + FP
Negative	FN	TN	FN + TN
Totals	TP + FN	FP + TN	n

Fig. 2 Fig. 2 2×2 Contingency Table.

kappa value. The kappa statistic is a generic term for a number of similar measures of agreement applied to categorical data and is a measure of the proportion of agreement between results beyond chance. It is used in assessing the degree to which two or more raters, (i.e., tests) examining the same data, (i.e., specimens) agree in assigning the data to categories (positive or negative results). A kappa value of 1.00 indicates perfect agreement; a value of 0.00 indicates no agreement above that expected by chance, and a kappa value of -1.00 indicates complete disagreement. Generally, kappa statistics above 0.80 are considered almost perfect. The kappa value is calculated from the results in the standard 2×2 contingency table, used to record the results from the comparative test (see Fig. 2).

$$OP = \frac{TP + TN}{n}$$

$$EP = \left[\frac{TP + FP}{n} \right] \times \left[\frac{TP + FN}{n} \right] + \left[\frac{FN + TN}{n} \right] \times \left[\frac{FP + TN}{n} \right]$$

$$K = \frac{OP - EP}{1 - EP}$$

where:

K = Kappa value.

OP = observed proportion of agreement.

EP = expected proportion of agreement.

TP = true positive.

FP = false positive.

TN = true negative.

FN = false negative.

n = total number of samples.

Diagnostic sensitivity and specificity (DSe and DSp) are also calculated from the results in the contingency table.

$$DSe = \frac{TP}{TP + FN}$$

$$DSp = \frac{TN}{FP + TN}$$

Two further important probabilities can be calculated from the data derived from the trueness study, the positive and negative predictive values (PPV and NPV). The PPV is a measure of a positive result from the test to truly predict the presence of disease/infection, while the NPV is the probability that a negative result accurately indicate a non-diseased/uninfected status.

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{FN + TN}$$

The Study Without a Gold Standard

In many cases, particularly for novel pathogens, a comparative method will not be available. Under these circumstances the gold standard may be a diagnosis determined by accepted clinical methods. It may also be necessary to incorporate results from more than one method to provide the comparative results. Whatever the approach chosen, estimating the performance characteristics of a new assay without a true comparative test is a challenging task. Typically, agreement is measured as the overall percentage or fraction of samples that have the same result (i.e., both either positive or negative). The difficulty with these simple agreement measures is that they do not take agreement by chance into account. A number of models have been proposed, based on latent class analysis, log-linear modeling and other techniques. Latent class analysis involves multiple imperfect tests that are used to construct a gold standard. Such models are based on the concept that the observed results of different, imperfect, tests for the same disease are influenced by a common but unobserved (latent) variable i.e., the true disease status.

Inter-Laboratory Testing

The ultimate evidence of an assay's fitness for purpose is its successful integration into other laboratories. The LDT reagents and samples used for the comparative study should be sent to a minimum of three laboratories willing to participate in testing. This will also provide additional data on the assay's reproducibility and robustness.

Maintaining the Validated Status and Re-Validation

The validated assay must be monitored consistently for repeatability through the performance of the run controls to evaluate any potential changes in the assay's precision and accuracy. All reagents should be assigned batch numbers and new batches tested with the controls and a suitable number of positive and negative samples from a previous run to ensure consistent efficiencies between batches. If more than one thermal cycler is used, each machine should be referenced and each test carried out on it identified. The results of the PCR controls and quantification standards must be plotted daily on Shewhart charts and assessed by Westgard rules to ensure that reaction efficiencies are maintained.

Participation in regular IQA and EQA schemes is fundamental to maintaining an assay's validated status. The purpose of EQA testing is not solely to monitor the performance of the assay, but assesses all aspects of the test procedure, including the extraction method and reporting accuracy. If an external panel is not available, alternatives can include blind sample testing, sample exchange with other laboratories, or clinical note reviews.

Technical modifications continually arise as manufacturers develop and improve their reagents for extracting nucleic acids, the PCR and instrumentation. Assessment of new probe chemistry or the transfer of an assay to full or semi-automated instrumentation, normally only requires a methods comparison study. In this way any changes to the diagnostic sensitivity and specificity can be assessed quickly and accurately and the updated assay introduced with the minimum delay.

Over time it is likely that an assay will require some degree of revalidation, either for purely technical reasons, or because of changes in the nature of the analyte detected. Regular monitoring of amplification efficiency together with clinical information will give an early warning of any changes in the sequence of the amplicon detected in an assay. Point mutations are common in the genomes of many infectious microorganisms, particularly RNA viruses. New viral lineages may also be introduced into the population being tested due to travel from abroad. Under such circumstances, significant modifications to the primers and probe may be required, necessitating reverification and revalidation of the assay. Amplicon sequencing should be part of the validation plan and an integral part of routine trouble-shooting algorithms, since it is capable of resolving errors due to non-specific binding of primers and probes.

Acknowledgment

This article, including tables and figure, is based on author's article "The Validation of Real-time PCR Assays for Infectious Diseases", published in *Real-Time PCR: Advanced Technologies and Applications (2013)*, edited by: Nick A. Saunders and Martin A. Lee, Health Protection Agency, Colindale, UK and Porton Consulting Research Ltd, Salisbury, UK (respectively), Caister Academic Press.

References

- Burd, E.M., 2010. Validation of laboratory-developed molecular assays for infectious diseases. *Clinical Microbiology Reviews* 32, 550–576.
Bustin, S.A., Benes, V., Garson, J.A., *et al.*, 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 55 (4), 611–622.

Further Reading

- Corman, V.M., Landt, O., Kaiser, M., *et al.*, 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25 (3), 23–30.
Hadgu, A., Qu, Y., 1998. A biomedical application of latent class models with random effects. *Journal of the Royal Statistical Society Series C* 47 (4), 603–616.
Hadgu, A., Dendukuri, N., Hilden, J., 2005. Evaluation of nucleic acid amplification tests in the absence of a perfect gold-standard test – A review of the statistical and epidemiologic issues. *Epidemiology* 16, 604–612.
Rådström, P., Knutsson, R., Wolfs, P., Lövenklev, M., Löfström, C., 2004. Pre-PCR processing: Strategies to generate PCR-compatible samples. *Molecular Biotechnology* 26, 133–146.
Raymaekers, M., Smets, R., Maes, B., Cartuyvels, R., 2009. Checklist for optimization and validation of real-time PCR assays. *Journal of Clinical Laboratory Analysis* 23, 145–151.
Rutjes, A.W.S., Reitsma, J.B., Coomarasamy, A., Khan, K.S., Bossuyt, P.M.M., 2007. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment* 11, 1–69.
Saunders, N.A., Martin, A.L. (Eds.), 2013. *Real-Time PCR: Advanced Technologies and Applications*. UK: Caister Academic Press.
Sloan, L.M., 2007. Real-time PCR in clinical microbiology: Verification, validation, and contamination control. *Clinical Microbiology Newsletter* 29, 87–95.
Westgard, J.O., Barry, P.L., Hunt, M.R., 1981. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clinical Chemistry* 3, 493–501.
Wilson, I.G., 1997. Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology* 63, 3741–3751.

Relevant Websites

- <https://www.ebi.ac.uk/Tools/msa/clustalo/>
Clustal Omega.
<http://www.ncbi.nlm.nih.gov>
National Center for Biotechnology Information.
<http://www.oligo.net/>
oligo.net.