INTERNATIONAL STANDARD

ISO 21474-2

First edition
2022-05

In vitro diagnostic medical devices — Multiplex molecular testing for nucleic acids —

Part 2: Validation and verification

Dispositifs médicaux de diagnostic in vitro – Tests moléculaires multiplex pour les acides nucléiques —

Partie 2: Validation et vérification

Reference number
ISO 21474-2:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 212, *Clinical laboratory testing and in vitro diagnostic test systems.*

A list of all parts in the ISO 21474 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

The first generation of in vitro diagnostic (IVD) medical devices for nucleic acid-based molecular tests focused on detection or quantitation of a single nucleic acid sequence (eg, viral RNA, mRNA, and genomic DNA) within a clinical specimen. By comparison, a multiplex molecular test simultaneously measures multiple nucleic acid sequences of interests in a single reaction. The development and clinical use of multiplex IVD medical devices are rapidly expanding with the technological advances and new elucidation of clinical significance of the many biomarkers.

The competition among reactions in multiplex molecular tests can impose more stringent requirements for sample purity, input reagents and platforms to avoid nonspecific reactions and background signal. In comparison to single target analysis, multiplex molecular tests require an increased number of controls, more complex performance evaluation/data analysis algorithms and more complex reporting of results.

Laboratories can develop assays in-house ("home-brew, laboratory-developed, in-house") or use commercially available multiplex assays involving a variety of technologies and instrument platforms. With the increase in the availability and use of multiplex molecular tests, a guideline for the development, validation, verification, control, data analysis, and implementation of multiplex molecular tests is increasingly needed. For a multiplex molecular test to reliably achieve its intended use, there should be control of the process from the acquisition of the sample and preparation of the nucleic acid for testing to the evaluation of the data and the reporting of the results. Multiplex molecular testing provides significant challenges to the laboratory with regards to appropriate validation and verification, acquisition of appropriate control materials, data analysis, and reporting. The complexity of data analysis and reporting of results is increased relative to singleplex assays. Moreover, the availability of sufficient and appropriate control and reference materials (RMs) to properly validate and verify multiplex molecular tests is a major challenge. However, the use of partial or full sequencing techniques can be useful in qualifying control materials. This document describes the recommendations for various aspects of validation and verification of the measurement by multiplex molecular tests in order to ensure reproducible performance of such tests, in developing and implementing multiplex molecular nucleic acid tests for clinical use.

In vitro diagnostic medical devices — Multiplex molecular testing for nucleic acids —

Part 2: Validation and verification

1 Scope

This document gives the general requirements for validation and verification of multiplex molecular tests which simultaneously identify two or more nucleic acid target sequences of interest. This document is applicable to all multiplex methods used for examination using IVD medical devices and laboratory developed tests (LDTs). It provides information for both qualitative and quantitative detection of nucleic acid target sequences.

This document is intended as guidance for multiplex examinations that either detect and/or quantify human nucleic acid target sequences or microbial pathogen nucleic acid target sequences from human clinical specimens.

This document is applicable to any molecular in vitro diagnostic (IVD) examination performed by medical laboratories. It is also intended to be used by laboratory customers, IVD developers and manufacturers, biobanks, institutions, and commercial organizations performing biomedical research and regulatory authorities. This document is not applicable to metagenomics.

NOTE An examination procedure developed for a laboratory's own use is often referred to as a "laboratory developed test," "LDT," or "in-house test".

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 15189, *Medical laboratories — Requirements for quality and competence*

ISO 21474-1, *In vitro diagnostic medical devices — Multiplex molecular testing for nucleic acids — Part 1: Terminology and general requirements for nucleic acid quality evaluation*

3 Terms and definitions

For the purposes of this document, terms and definitions given in ISO 21474-1 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
analytical sensitivity
quotient of the change in a measurement indication and the corresponding change in a value of a quantity being measured

[SOURCE: ISO 18113-1:2009, A.3.3, modified —NOTES 1 to 4 were removed [1].]

3.2
diagnostic sensitivity
ability of an in vitro diagnostic (IVD) examination procedure to identify the presence of a target marker associated with a specific disease or condition

[SOURCE: ISO 18113-1:2009, A.3.15, modified —NOTES 1 to 4 were removed [1].]

4 General requirements

4.1 General

Multiplex molecular tests are IVD medical devices that detect and/or measure multiple nucleic acid sequences simultaneously, such as multiplex PCR, DNA microarray, and massive parallel sequencing-based methodologies.

In cases of multiplex molecular tests for nucleic acid, competition among individual reactions may impose more rigorous requirements for sample purity, sample input, reagents, and platforms to avoid nonspecific reactions and background signal. In comparison to singleplex analysis, multiplex molecular tests can have an increased number of controls, more complex performance evaluation/data analysis algorithms, and more complex reporting of results.

Validation activities are aimed to ensure determination of performance characteristics for intended use, which is performed in the test development process, by either the manufacturer or laboratory. The design verification and validation plan, created by manufacturers for an IVD and laboratories for an LDT, should aim to establish performance specifications such as accuracy, specificity, precision, LOD/assay range, and cross-reactivity/interfering substance. When planning the development of a test, the intended use shall be defined, including specimen types to be used. The examination manufacturer shall specify instructions for examination processses. These shall be followed. Whenever any processes or methods are changed during the verification process and after the implementation, a laboratory shall validate the assay with the new performance characteristics. The laboratory shall perform verification activities to confirm the design input specifications required for patient testing in advance. The laboratory then performs ongoing quality assurance (QA) of the multiplex molecular tests by document control of procedures, operator training, routine quality control (QC), proficiency testing, instrument calibration, and result correlation with clinical findings.

The validity and reliability of multiplex molecular test result can be impacted by all steps of a laboratory workflow, including primary sample collection, transport, storage, and processing in the preanalytical phase as well as by the analytical phase itself.

NOTE The CLSI guideline, MM17-A [2] provides recommendations for various aspects of verification and validation of multiplex testing.

4.2 Laboratory requirements

The requirement for quality and competence in medical laboratories shall be as described in ISO 15189.

For assays that are not fully integrated in a single disposable or on a single platform (from sample to result), sample treatment should be carried out in separate working areas/rooms as specified in ISO 22174:2005, Clause 6[3].

Accidental contamination of DNA and RNA can originate from aerosols containing template, such as amplicons generated from previously amplified NA (nucleic acid) targets. As a consequence, the organization of the work area in the laboratory and good practice shall be based on

— the systematic containment of the methodological steps involved in the production of the results, and

— a “forward flow” principle for sample handling.

The latter ensures that DNA and RNA in the test material remain physically segregated. Further details can be found in ISO 21571[4].

4.3 Reagents requirements

All reagents and materials used in the analysis should be identical, or equivalent, to those specified in the method. Otherwise, all reagents and materials should be of grade relevant for molecular biology, e.g. DNase and RNase free water.

Those reagents shall be stored and used as recommended by the supplier or according to the laboratory QA specifications.

The characteristics and quality of reagents (e.g. fluorescent dye(s), buffers) and the amount of an external measurement standard, or reference material (RM), that is added to the reaction mixture should be validated/verified.

All critical reagents should be tested for functionality before inclusion in formal performance evaluations. In multiplex molecular tests, functionality should be evaluated using synthetic controls, when not feasible to establish it for all the targets using genomic specimens.

Where the examination manufacturer provides instructions for storage and use of a commercial kit for measurement, these shall be followed. Where the examination manufacturer does not provide such instructions, they shall be specified and validated by the laboratory and followed.

4.4 Apparatus and equipment

The laboratory should use properly validated (i.e. installation qualified, operationally qualified and performance qualified) and maintained equipment according to the manufacturers' instructions and the requirements given in ISO 15189. In addition to standard laboratory equipment, specific apparatus for multiplex analysis is described in the corresponding individual standards, e.g. ISO 16578 for microarray[5].

Where available, calibration should be routinely performed and records maintained on equipment where performance can impact the data produced.

4.5 Reference and control materials

4.5.1 General

Whenever reference materials (RMs) are available, they shall be selected to be fit for their intended use. When RMs are not available, the alternative approaches shall be planned, developed and documented.

RMs shall be considered for the validation of multiplex molecular testing for nucleic acid. RMs can also be used for qualitative or quantitative methods, serving many different functions, such as calibration and assessment of an examination procedure. Especially for genomic analysis using massive parallel sequencing methods, genomic RMs can vary the sequence data depending on the chosen material (see Annexes A and B); therefore, RMs for such methods shall be considered for fitting to the intended use.

The laboratory should use QC materials that react to the examining system in a manner as close as possible to patient sample, as specified in ISO 15189. If these control materials are not available, it is not possible to test all the probes present in the panel for reagent acceptance or fidelity monitoring. The laboratory will adapt its arrangements notably by taking into account the clinical context.

Examination results shall be evaluated to confirm that the quality of the sample fits for the intended use. When the evaluation is performed, the use of RMs (see Annexes A and B) including nongenomic (see 4.5.3) and genomic (see 4.5.2) RMs shall be considered.

To validate the use of different matrices, overload assays (spiking) are necessary.

Where the manufacturer provides instructions for storage and use of RMs and control materials for measurement, these shall be followed. Where the examination manufacturer does not provide such instructions, they shall be specified and validated by the laboratory and followed.

4.5.2 Endogenous nucleic acid

Genomic DNA (gDNA) is the QC material that most closely resembles patient samples, and the most preferable, although not always the most efficient. The presence or absence of the mutations or variants in any material (e.g. immortalized human cell lines, purified gDNA, bacterial/microbial cultures, and bacterial gDNA) should be validated or verified by the laboratory for each new lot obtained before use.

Synthetic recombinant engineered or artificially constructed long DNA, e.g. recombinant artificial chromosomes, can be used as an alternative control material of gDNA (see 4.5.3). Several transcripts of housekeeping genes can be used as control materials.

NOTE Housekeeping genes include, e.g. beta-actin, GAPDH and cyclophilin.

4.5.3 Nongenomic reference materials (RMs)

Nongenomic QC materials are synthetic, recombinant, engineered, or artificially constructed, and are often composed of plasmids, due to their stability and ease of use. Nongenomic QC materials are of great value and particularly suitable for multiplex assays, since they can be engineered to contain all sequence variants in gene segments in one sample detected in an assay. Nongenomic controls should be chosen for QA purposes carefully to be sure that they are suitable for monitoring the desired system parameters. There are concerns and limitations regarding nongenomic controls. Some test systems and/or software are not designed to correctly identify multiple alleles contained in a multiplex control. Nongenomic controls have to be added to the patient material, and therefore cannot behave exactly as patient-derived material through the entire analysis process. This situation is of particular concern when there is a reasonable chance of mispriming or false-positive reporting, due to the presence of pseudogenes or other sequences normally present with high homology to the target of interest.

Nongenomic RMs can be used for not only DNA analysis but also RNA analysis.

NOTE DNA or RNA RMs can be evaluated by establishing metrological traceability from certified RMs such as NMIJ CRM 6205-a and NMIJ CRM 6204-b.

4.6 Calibration of the analysis

In the quantitative measurement, such as for qPCR-based assays, an appropriate number of calibration points and replicates covering the range of reliable signal should be applied. The calibration influences the measurement uncertainty (MU). More details on calibration of PCR are described in ISO 20395 [16].

An alternative to genomic DNA or mRNA calibration RMs can be considered for use, provided that it is demonstrated to perform in an equivalent way to the genomic DNA RM and the genomic DNA extracted from the sample, or the mRNA RM and the mRNA extracted from the sample.

EXAMPLE As an alternative to genomic DNA or mRNA calibration RMs, a dilution series of a plasmid or synthetic dsDNA containing the target sequence or a plasmid or synthetic dsDNA and ssRNA can be respectively used.

4.7 Input range

Input nucleic acid should be titrated to optimize the range of concentrations, in order to ensure the detection of each target sequence in a given assay, e.g. typical ranges for PCR, arrays, and sequencing.

5 Evaluation of performance characteristics

5.1 General

The validation protocol should start with an explicit statement of the intended use, which will determine the types of samples and the performance characteristics that need to be addressed.

The validation protocol should be designed carefully to ensure that all relevant parameters are addressed as efficiently as possible. Based on design and development planning, the laboratory shall perform validation of the intended use of the test in comparison to an existing predicate device, if available. When a predicate device is not available, the laboratory shall validate the intended use by appropriately designed studies. In accordance with the validation plan, the laboratory shall describe and document relevant performance characteristics of the test in relation to the disease or other clinical context assessed by the test. The laboratory shall identify those test characteristics that are critically associated with the clinical relevance and clinical utility of the test. These characteristics shall be documented as clinical performance characteristics. For more details regarding the test using nucleic acid amplification for microbial pathogens, see ISO 17822 [17].

Analytical performance evaluations should test the multiplex molecular test system in its final configuration, and not in separate singleplex experiments.

Considerations specific to the multiplex molecular tests should include, but not be limited to:

— a type and heterogeneity of the specimen;

— LOD that is clinically needed;

— genetic promiscuity and instability of target sequences.

EXAMPLE Examples of genetic promiscuity and instability of target sequences are sequence heterogeneity both within the disease and across the progression of it, specificity when considering closely related target of sequence, genetic origin of acquired or inherited when considering with disease inheritance, and genetic association when considering the disease diagnosis, prognosis or treatment.

NOTE Recommendations on practical guidance for laboratories regarding validation and verification of massive parallel sequencing for genetic diseases are provided in the literature [6][7][8].

In the verification process of a supplier method, major issues to be addressed should include as follows:

— the issue of testing all the probes present in the panel;

— the issue (or even the impossibility) of providing a number of samples that is statistically significant to conclude on the validity of the tests in particular “positive samples”;

— the issue of comparing the method to an earlier method;

— the complex problem of using internal QCs for all the targets of the panel.

The selection of the analytical performance evaluation tests that are performed should include the consideration of the design or operating principles of the IVD as well as the intended use or purpose. For example, limit of quantitation (LoQ) can not be applicable for an IVD that renders a qualitative or binary result.

5.2 Analytical specificity

5.2.1 Analytical reactivity

Since analytes such as sequence variants or pathogen-specific sequences often respond differently to reagent variability, all targeted analytes shall be tested, when possible, for accurate results. The major analytes detected with selected reagent shall be identified. The information of major analytes should be considered to include in the minimum specifications for the reagents. Those analytes most

sensitive to reagent parameters shall be identified and used to establish minimum specifications for the reagents. For the detection analysis of pathogens, weakly positive specimens, e.g. at the established limit of detection (LOD) concentration, or at a copy number empirically established at the lower end of the diagnostic range, should be included. It is also possible to use previously negative assayed samples loaded (spiked) with a defined amount of copies.

Analytical specificity may be established and demonstrated through a variety of tests. The following are the routine tests that establish analytical specificity and reactivity. Analytical specificity of an assay is its ability to detect and identify only the intended sequences or biomarkers. Challenging an assay to detect the intended target(s) across the reportable range is an approach to demonstrate specificity. An assay's ability to detect and identify contrived or RMs containing the intended targets at varying concentrations (e.g. at or near the LOD, two to three times the LOD and higher) is an approach in combination with other tests (e.g. cross-reactivity) to demonstrate specificity.

In multiplex molecular tests based on PCR, relative concentrations of primers and probe should be balanced and optimized in order to ensure the detection of each target sequence, by maximizing amplification efficiency and minimizing secondary structures, e.g. stem-loop, pseudoknot, so that product detection is optimized.

In PCR, in order to balance the competition of the reactions, efficiency can be maximized by evaluating each individual parameter in relation to the accumulation of products. Probe interactions can be measured by singleplex and multiplex reactions with the same probe sets; signal intensities and crossover thresholds can be compared to identify detrimental interactive effects resulting from multiplexed primers and probes.

5.2.2 Limit of blank

The term ‘limit of blank’ is the highest measurement result that is likely to be observed for a blank sample. A signal can be observed in the absence of any analyte. This is generally caused by primer dimerization. Primer dimerization can occur more frequently as the number of primer pairs increases in a reaction. To minimize primer-dimer formation, primer sequences should be screened to avoid homology at 3’ end among all the primers in the reaction; it is also useful to contain a G or C at the 3’ end of the primers to clamp the primer and prevent "breathing" of ends increasing priming efficiency. DNA "breathing" occurs when ends do not stay annealed but fray or split apart. Modified polymerases are available which prevent nonspecific amplification until activated and released into the reaction after an initial incubation at a high temperature.

A carrier shall be included in a blank, since blanks that contain no nucleic acid are not good at estimating low level contamination due to absence of carrier effect.

NOTE Well designed probe based nucleic acid amplification test (NAAT) methods will have no signal in absence of template and thus no limit of blank.

5.2.3 Cross-reactivity

A panel of closely related organisms/alleles shall be assessed in order to determine whether a multiplex assay cross-reacts with analytes other than the ones it is designed to measure. Primer/probes shall be designed considering species specificity. In a viral detection assay, primer/probes shall be designed not to accidentally cross-react with human and bacterial genome DNA. In the designing process, in silico analyses including nucleic acid database search should be performed to confirm the species specificity, avoiding the possibility of accidental cross reaction with human and bacteria.

For multiplex detection of pathogens, the effects of clinically relevant coinfections should be assessed to determine that they do not interfere with test results for any of the pathogens probed by the assay. The effect of a sample mimicking clinical specimens with high background levels of human genomic DNA shall be also assessed.

5.2.4 Exclusivity

For multiplex detection of target variants within the alleles, primer/probes shall be designed considering variant specificity.

5.2.5 Interfering substances and carryover

Interfering substances refer to inhibitory compounds from sample, or extraction process, and typically cause false negatives. The possible source of test interference shall be specifically identified, and the impact of each shall be systematically evaluated during verification and validation. The risk of carryover contamination shall be evaluated at each step of the assay. A no template control should be included in every run, where relevant.

The selection of interfering substances, endogenous and exogenous, is dependent on the sample type and sample processing or pre-examination steps prior to testing.

Carryover is a recognized problem with massive parallel sequencing for human genome that are designed to detect variants with low allele burden. During examination design, procedures shall be in place to avoid carryover from one sample to another. Bioinformatics approaches can be used to detect human-human sample contamination to monitor carryover.

NOTE In massive parallel sequencing for human genome, there is a considerable interference from repetitive sequence and pseudogenes. For highly fragmented DNA, short reads can be misaligned if derived from pseudogenes and yield false positive results. Pseudogenes and highly repetitive sequences can reduce on-target reads by depleting capture probes, leading to underestimation of detection of targets.

5.3 Range of reliable signal, reportable range and reference range

Determination of the range of reliable signal such as LOD and linear range in a multiplex molecular test shall be ensured in the verification and validation study. The method will only be applicable in that range. The signal of the analyte shall be valid within the range of reliable signal.

Cross-over signals shall be empirically established. When criteria are used to discriminate among the range of analytes probed by the multiplex molecular test system, they should be adequately tested in cross-reactivity studies and related analytical evaluations, such as LOD and precision assessments.

The reportable range is the span of all examination results that are considered valid. The reference range is the range of normal values. The reportable range and reference range shall be determined, depending on the intended use.

NOTE In multiplex molecular examinations for human genome based on massive parallel sequencing, the reference range includes variants that are considered benign or nonpathogenic, in addition to the reference Human Genome which is a compilation of multiple genomes from healthy individuals.

5.4 Limit of detection of multiplex molecular test platform (LODP)

In the case of multiplex molecular tests, LOD should be evaluated for all targets within the assay where possible. Where possible, the LOD of each major target should be established. In selection of appropriate targets for the LOD testing, prevalence or difficulty of detection should be considered a criterion for selection. The target should at least include the most prevalent and the most problematic to detect. For the remaining targets, the lowest concentration of target analyte that can be consistently detected in ≥95 % of sample measurements should be provided.

Use of LODP should be considered instead of LOD, when the target sequences consist of a mixture of sequences to be detected and not to be detected, regardless of platform used.

In multiplex molecular tests, it is not always possible to determine LOD of each target being probed. If LOD of individual targets are required, an external measurement standard (or RMs) could be used for experimental determination of the limit of representative targets on a given platform. The quantity of each analyte to be detected should be higher than LODP.

The experimental LODP is related to the analytical portion, the quality/quantity of the analyte, and the absolute LODP of the detection method. These values should be confirmed via an interlaboratory comparison using appropriate reference and control samples, and the lowest level of the external measurement standard (or RM) obtained experimentally should have a false negative rate of less than or equal to 5 %.

In the performance evaluation process including LOD determination, any cross-reactivity among the primer/probes used in the system should be assessed. The acceptance criteria of the performance evaluation should include a description that comprehensively considers the whole analysis system which consists of the individual analyte detections.

NOTE 1 NMIJ CRM 6204-a is available for an example of the external measurement standard, which can be universally used for RNA expression.

NOTE 2 When appropriate reference or control samples are not available, alternatives such as international conventional calibrators or working calibrators prepared in-house by laboratories and manufacturers are used as RMs. The laboratory assures the commutability of any RM used for validation purposes.

5.5 Measurement precision and uncertainty

In precision studies, sources of variability that shall be considered are instrument, laboratory, operator, sample concentration, sample source, reagent lot, run, day, and the time of day if relevant.

Precision studies should include challenging specimens such as near LOD (e.g. 2 to 3 times the LOD) or near the clinical cut-off.

In repeatability studies, representation of rare alleles shall be included to ensure their presence is able to be detected with precision. A reasonable mixture of variants should be included in each run of repeatability testing.

The reproducibility panel shall include each analyte.

For qualitative tests with underlying quantitative outputs, precision can be measured with coefficients of variation (CV) for each source of variation, and analysis of variance components for the total variation. This should be established using specimens that are reactive for analytes measured in the multiplex molecular tests, as well as at least some specimens that are nonreactive for all target/analytes.

NOTE 1 Use of CV is not appropriate for NAAT methods reporting arbitrary fluorescent units such as Cq/Ct. CV is also not mathematically appropriate for nucleic acid measurements that follow a log normal distribution.

MU can be defined as an estimated range of values within which the true value of the measurement resides. For qualitative tests with underlying quantitative outputs, MU can be estimated by information on precision and bias provided by experiments performed in method verification.

In multiplex molecular tests, it is not always possible to determine the MU of each target being probed. If MU of individual targets are required, an external measurement standard (or RMs) could be used for experimental determination of the MU of representative targets on a given platform. The combined MU of examination should be larger than the MU of each target.

NOTE 2 A practical guidance for the estimation of MU is provided in ISO/TS 20914[9].

5.6 Accuracy and method comparison studies

Accuracy has various elements such as analytical/diagnostic sensitivity and specificity, positive predictive value (PPV) and negative predictive value (NPV), percent agreements, false-positive and false-negative rates. The performance of the assay should be compared to an established reference method or comparator, other validated molecular IVD medical device or to RMs or both. They shall be established through relevant statistical analyses of a dataset generated from an appropriately designed

method comparison study. Appropriate data analysis techniques and reference methods shall be chosen to establish accuracy, both at the level of individual analyte and the overall system.

NOTE 1 When using a non-reference standard to be compared for evaluation of a new test, unbiased estimates of sensitivity and specificity cannot be directly calculated. The same numerical calculations can apply where the estimates are called positive percent agreement (PPA) and negative percent agreement rather than sensitivity and specificity.

Analytical/diagnostic sensitivity/PPV for some target/alleles on the multiplex panel are often equivocally established. In such instances, the test should be developed by procuring as many positive samples as possible for a method comparison, and obtaining complementary data using an alternative specimen source. Publications that have investigated the diagnostic sensitivity for these rare targets with the same assay can provide additional data.

In case of massive parallel sequencing, for each reported variant class (i.e. SNVs, indels, CNAs, SVs), the performance as PPA and PPV shall be established and documented, if relevant.

NOTE 2 Some reports describe a framework for an appropriate minimum number of samples for multiplex molecular testing including massive parallel sequencing test validation/verification, when analysed with 95 % confidence and 95 % reliability, which are not described in this document[6][7].

Annex A
(informative)

Certified reference materials (CRMs)

A.1 Uses of nucleic acid certified reference materials (CRMs)

A.1.1 General

A more detailed description of the use of certified reference materials (CRMs) and reference materials (RMs) are available in ISO Guide 33[10].

A.1.2 Examination verification and measurement uncertainty (MU)

Estimation of bias (the difference between the measured value and the true value) is one of the most difficult elements of examination verification, but appropriate RMs can provide valuable information, within the limits of the uncertainty of the RMs certified value(s) and the uncertainty of the examination being validated. Although traceable certified values are highly desirable, the estimation of bias differences between two or more examinations can be established by use of less rigorously certified RMs. Clearly the RMs need to be within the scope of the examination in terms of, e.g. matrix type, analyte concentration. Ideally, it is recommended that a number of RMs covering the full range of the examination be tested. Where minor modifications to a well-established examination are being evaluated then less rigorous bias studies can be employed.

Replicate measurement of the RM, covering the full range of variables permitted by the examination being validated can be used to estimate the uncertainty associated with any bias, which will normally be corrected for.

The uncertainty associated with an RM is controlled to be no greater than one third of that of the sample measurement.

A.1.3 Calibration

Normally, a pure substance RM is used for calibration of the measurement stage of an examination. Other components of the examination, such as sample digestion, separation and derivatisation are, of course, not covered and loss of analyte, contamination and interferences and their associated uncertainties are addressed as part of the verification of the examination. The uncertainty associated with RM purity contributes to the total uncertainty of the measurement. For example, an RM certified as 99,9 % pure, with an expanded uncertainty U (k=2) of 0,1 % will contribute an uncertainty component of 0,1 % to the overall MU budget. In the case of trace substance analysis, this level of uncertainty will rarely be important but for analysis work, it can be expected to be significant.

Some other examinations based on spectroscopic analysis, such as X-ray fluorescence (XRF) analysis, use matrix RMs for calibration of the complete analytical process. In addition to a close matrix match, the analyte form will be the same in the samples and RMs, and the analytical concentrations of the RMs will be designed to span that of the samples.

ISO Guide 33 [10]and Reference [13] provide additional useful information.

A.1.4 Quality control (QC) and quality assurance (QA)

For in-house QC, the requirement of certified property value(s) can be relaxed, but adequate homogeneity and stability are essential. Similar requirements apply to samples used to establish how well or badly measurements made in different laboratories agree. In the case of proficiency testing, homogeneity is

essential and sample stability within the timescale of the exercise is assessed and controlled. Although desirable, the cost of certifying the property values of proficiency testing samples often prohibits this being done and consensus mean values are often used instead. As a consequence, there often remains some doubt concerning the reliability of assigned values used in proficiency testing schemes. This is because, although the consensus mean of a set of data has a value, 'the majority' is not necessarily correct and as a consequence the values carry some undisclosed element of uncertainty. The interpretation of proficiency testing data thus needs to be carried out with caution. Proficiency testing organization that meet the requirements of ISO 17043 [12] take uncertainty into account when operating the results.

A.2 Assessment of the suitability of reference materials (RMs)

As previously indicated, the key quality parameter is the uncertainty associated with the certified value and the reliability of the uncertainty estimate. Uncertainty budgets are calculated by using the following: 'Certification' data will be stated together with the expanded uncertainty, U, using a coverage factor k=2 which gives a level of confidence of approximately 95 %.

However, the full uncertainty data is often not available, and so other quality criteria should be considered. Also, the non-expert might not be in a position to fully evaluate the 'certification' data and a quality check list, or a third-party quality approval system, is desirable. Such systems are under development but will take some time to become fully established.

A.3 Certificates and supporting reports

Ideally, a certificate in accordance with ISO Guide 31 [13] and a report covering the characterisation, certification and statistical analysis procedures, in accordance with ISO Guide 35 [14], should be available. However, many RMs, particularly older materials and materials not specifically produced as RMs, cannot be fully in conformance with ISO Guide 31 and ISO Guide 35. Alternative, equivalent information in whatever form it is available, that provides credible evidence of conformance can be considered acceptable. Examples include the following: technical reports, trade specifications, papers in journals or reports of scientific meetings and correspondence with suppliers.