Similarly, an unknown or unexpected allele can occur at a given mutation site or polymorphism underlying the primers used in the PCR amplification. There are some genes in which this is known to occur with certain frequency (eg, allele drop-out in *CYP2D6*), but it is an expected limitation in most assays. Such unexpected sequence variation may also occur with greater frequency in infectious agents whose replicative polymerases lack proofreading functions, introducing substantial spontaneously arising genetic variation underlying amplification and sequencing primer sites.

2.5.6 Measuring Interval

The measuring interval is relevant to quantitative assays but is not within this guideline's scope.

2.5.7 Reference Interval

The reference interval for each target in qualitative nucleic acid assays is typically "detected" or "not detected." The specific reporting scheme depends on the clinical use of the assay results. A report may include a list of all measurand results or interpretation and a list of clinically significant results, as appropriate (see Subchapter 2.3.6).

2.5.8 Other Performance Characteristics

During assay validation, the laboratory needs to consider additional test characteristics and factors that could affect test system performance, including preexamination, examination, and postexamination considerations.

2.5.8.1 Sample Stability

Sample storage and shipping may lead to nucleic acid degradation. Therefore, whether the recommended sample storage, shipping times, temperatures, and/or repeated freeze-thaw cycles have affected sample stability should be determined. An analysis should be performed using sample aliquots at several time points during the recommended storage and transport at both ends of the recommended temperature range.

2.5.8.2 Patient Population

The laboratory should consider whether its patient population will affect test performance, such as through changes in sample quality or disease prevalence in the population. If the assay was validated using samples from one population, additional studies may be needed to apply it to a different population, and reporting may change (eg, residual risk for genetic disease based on negative test results in various ethnic groups).

2.5.8.3 Diagnostic Sensitivity

Diagnostic sensitivity is the ability of an assay to produce a positive test result in a patient with known disease. It may differ from analytical sensitivity depending on the measurand's association with the clinical illness. A full discussion regarding diagnostic sensitivity is beyond this guideline's scope, but it should be assessed during test validation. Literature and medical chart review are acceptable options for established tests when prospective clinical trials are not needed.

2.5.8.4 Diagnostic Specificity

Diagnostic specificity is the ability of an assay to produce a negative test result in a patient without known disease. It may differ from analytical specificity depending on the measurand's association with the clinical illness. A full discussion regarding diagnostic specificity is beyond this guideline's scope, but it should be assessed during test validation. Literature and medical chart review are acceptable options for established tests when prospective clinical trials are not performed.

2.5.9 Data Analysis for Validation

This subchapter suggests the criteria to be considered for validating the data analysis segment of an LDT. Validating data analysis should also include validating the software output results. The data output for target (measurand) detection can be a simple "yes" or "no" answer, or it can be multiple target detection in a sample with complex algorithms. The data types needed for validating a multiplex nucleic acid assay, as well as for analyzing data analysis and interpreting results, depend on the assay's intended purpose and the population in which the assay is used. For example, routine carrier testing for an autosomal recessive disorder is an indication that differs distinctly from diagnosis of the same disorder. The former may need mostly analytical evaluation, whereas the latter also needs data on clinical performance. This subchapter focuses primarily on analyzing data generated during the analytical validation for each assay characteristic described in Subchapter 2.5.

2.5.9.1 Accuracy

Challenges when measuring accuracy for a multiplex assay include:

- Availability of one method that covers all the measurands detected in the new method
- Selection of an appropriate comparison method that determines accuracy at each measurand level as well as the multiplex system

Accuracy for qualitative test results can be determined by comparison with a reference test method. For measurands with no available comparison method, accuracy can be determined by comparing the data to the clinical diagnosis results determined by definitive medical laboratory methods.

Positive percent agreement (PPA) is the assay's ability to obtain positive results for a measurand in concordance with positive results obtained by the reference method. For example, PPA is the proportion of heterozygous carriers for a specific mutation that test positive for that mutation. Negative percent agreement (NPA) for a measurand is the assay's ability to obtain negative results for a measurand in concordance with negative results obtained by the reference method. PPA and NPA are calculated for each individual measurand included in the assay and for clinically relevant combinations of measurands.

Thus, for one measurand, Table 5 shows an example of PPA and NPA for a single measurand.

Table 5. 2 × 2 Contingency Table for Singleplex Assay vs Reference Method

Singleplex Assay	Reference Method	
	A Absent	A Present
A Negative	146	4
A Positive	4	46
Total	150	50

Statistics generated from information in Table 5 are:

- PPA = 92.0% (46/50), 95% CI = 81.2% to 96.8%
- NPA = 97.3% (146/150), 95% CI = 93.3% to 99.0%

In general, for a number (N) of clinically relevant combinations of multiplex assay outputs and the number of clinically relevant classes (M), the data table can be summarized as N × M as shown in Table 6. For N nonoverlapping clinically relevant outputs, which span all test outcomes, the data are shown in Table 6.

Table 6. N × M Contingency Table for Multiplex Assay vs Reference Method

Multiplex Assay	Reference Method			
	Outcome 1	**Outcome 2**	**…**	**Outcome M**
Outcome 1	a_{11}	a_{12}	…	a_{1M}
Outcome 2	a_{21}	a_{22}	…	a_{2M}
Outcome N	a_{N1}	a_{N2}	…	a_{NM}
Total			…	

Abbreviations: a_{xx}, number of samples with this result combination; M, number of clinically relevant outputs of the reference multiplex assay; N, number of clinically relevant combinations of outputs of the test multiplex assay.

Thus, for a number of multiplex test measurands (ie, features) (K), when each feature has only two outputs (ie, positive and negative), the table becomes $2^K \times 2^K$. If each feature has three outputs (eg, positive, negative, and invalid), the table becomes $3^K \times 2^K$. Some classes can be omitted because they are impossible or irrelevant from a biological or clinical point of view, and some classes can be combined if the clinical implications are the same.

As shown in Table 7, using an example with two measurands, A and B, four combinations are possible (A−/B−, A−/B+, A+/B−, and A+/B+) and may need separate validation against a reference test(s) for the pathogens. Because of cross-reactivity or other concerns, performance may also need to be evaluated analytically for each combination of the measurands' presence or absence. The study data can be presented fully through a 4 × 4 table cross-classifying the four possible assay results against those of the reference test. A 4 × 4 data table for the example is shown in Table 7.

Table 7. 4 × 4 Contingency Table for Multiplex Assay vs Reference Method

Multiplex Assay	Reference Method			
	A−/B−	**A−/B+**	**A+/B−**	**A+/B+**
A−/B−	95	2	1	0
A−/B+	1	48	1	2
A+/B−	4	0	38	1
A+/B+	0	0	0	7
Total	100	50	40	10

Depending on the assay structure, PPA and NPA are generally calculated for each individual measurand included in the assay. In the example of an assay to detect two measurands (measurands A and B), PPA and NPA for each pathogen are computed using two 2 × 2 tables, as shown in Tables 8 and 9.

Table 8. 2 × 2 Contingency Table for Individual Measurand A

Multiplex Assay	Reference Method	
	A Absent	**A Present**
A Negative	146	4
A Positive	4	46
Total	150	50

Statistics generated from information in Table 8 are:

- PPA (A) = 92.0% (46/50), 95% CI = 81.2% to 96.8%
- NPA (A) = 97.3% (146/150), 95% CI = 93.3% to 99.0%

Table 9. 2 × 2 Contingency Table for Individual Measurand B

Multiplex Test	Reference Method	
	B Absent	B Present
B Negative	138	3
B Positive	2	57
Total	140	60

Statistics generated from information in Table 9 are:

- PPA (B) = 95.0% (57/60), 95% CI = 86.3% to 98.3%
- NPA (B) = 98.5% (138/140), 95% CI = 94.9% to 99.6%

NOTE: The data in Table 7 can be used to construct two 2 × 2 tables in the form of Tables 8 and 9 (tables for evaluating the performance of the assay separately for measurands A and B), but the converse is not true.

In addition to PPA and NPA for each pathogen or measurand, the percent of correct classification by the multiplex test can be reported separately for each class:

- **For dual negatives (class A−/B−):** 95.0% (95/100)
- **For A only (class A+/B−):** 95.0% (38/40)
- **For B only (class A−/B+):** 96.0% (48/50)
- **For dual positives (class A+/B+):** 70% (7/10)

The larger the number of multiplex test measurands, the larger the number of possible outcomes. In the example above for measurands A and B, the table is 4 × 4 (2 × 2 different classes by the reference method by 2 × 2 different outcomes of the multiplex test).

Some combinations can be of particular clinical interest. For example, when pathogens A or B are detected, the outcome of interest is the PPA for detecting coinfection (A+/B+). Correspondingly, NPA for detecting noncoinfection (A−/B−, A−/B+, or A+/B−) is based on pooling the three combinations that are complementary to coinfection. Using Table 7 (4 × 4), 10 specimens have coinfection A+/B+ according to the reference method, and the PPA to detect coinfection is 70% (7/10). Correspondingly, to detect noncoinfection, the NPA is 100% (190/190). When appropriate, overall PPA and NPA can be computed by considering two classes:

- A class in which any pathogen is present (eg, A−/B+, A+/B−, or A+/B+)
- A class for which neither pathogen is present (eg, A−/B−)

In other words, there are two outcomes of a multiplex assay: positive (A−/B+, A+/B−, or A+/B+) and negative (A−/B−). An appropriate example is an assay designed as an initial screen to detect the presence of any major respiratory viruses in a sample. The overall percent agreement (OPA) pools overall PPA and NPA together. Although a useful summary measure of accuracy, OPA masks the accuracy within measurand-positive and -negative populations, one of which could be poor. Therefore, OPA should not be reported alone but can accompany the PPA and NPA performance estimates. Uncertainty about estimates of PPA, NPA, and OPA is characterized by 95% CI, often computed using a score method. The associated 95% CI or other appropriate uncertainty estimates should accompany OPA, PPA, and NPA performance estimates.

2.5.9.2 Precision

Ideally, all assay variability sources should be evaluated in the precision study. For general information on precision studies, see CLSI documents EP05[84] and EP12.[34] For qualitative tests, variance components should be estimated using the appropriate statistical method for all the factors considered in the precision study, as well as overall. Particularly for qualitative tests that have underlying quantitative output, precision is often measured for each variation source, as well as the total variation, using variance analysis. For these tests, mean, SD, and % CV should be reported to summarize the results. These parameters should be estimated using samples that are nonreactive for all targets or measurands, samples that are low positive (near the LLoD), and samples that are positive for each target or measurand, ideally reflecting clinically relevant concentrations. In general, the test materials should be selected to simulate the characteristics of the appropriate patient samples.

In precision study analysis, factors held constant and varied during evaluation should be identified (eg, instrument calibration, reagent lots, and operators). Multiplex assay precision can be assessed and analyzed as overall assay precision and separately for each measurand call. For example, in a *CFTR* assay testing for 23 mutations, precision can be assessed separately for each mutation site and for the overall assay. In cases in which any one of the mutations is miscalled, the overall assay result is incorrect. Depending on the assay type, both measurand and overall precision can be reported. However, because calls resulting from a single specimen are not truly independent, calculating precision for the overall assay result from one specimen might be a more appropriate measure than using several single calls for every mutation site.

Additional precision study results analysis and reporting considerations for a multiplex *CFTR* assay stem from the low prevalence of some alleles. If low prevalence mutations are not represented in the precision study, a laboratory could unknowingly miss every allele present at a frequency of 0.1% and still report a high precision. For rare alleles, genotypes, or mutations, archived sample cell lines or nongenomic RMs (see Subchapters 2.2.2.2 and 2.2.2.5) containing the less prevalent mutations can be used in precision studies. During assay validation, the manufacturer or laboratory should perform precision studies and calculate precision for detecting every genotype reported by the assay, as well as demonstrate that the assay can distinguish between heterozygotes and homozygotes for mutations reported by the assay. Performing precision studies for each genotype may be excessively burdensome for some assays that detect a large number of mutations (eg, HPV subtyping and CF mutation analysis). An alternative is to perform precision studies with the most prevalent and most clinically significant genotypes.

Different assays might have distinctive scenarios for reporting indeterminate or “no call” results. For example, in an assay detecting 23 *CFTR* mutations in which one of the 23 loci tested returns an indeterminate result and the remaining 22 yield a correct call, the system could be set up to report the result for the whole assay as indeterminate or to report just the result for the particular mutation site as indeterminate. These two scenarios for reporting data need to be carefully considered when choosing how to correctly calculate and report assay precision.

2.5.9.3 Analytical Sensitivity (Lower Limit of Detection)

Data analysis for analytical sensitivity includes establishing the LLoD and the LoB (if applicable). The LLoD can be determined either statistically by calculating the point at which a signal can be distinguished from background (LLoD as function of LoB) or empirically by testing sample serial dilution with a known target concentration in the analytical range of the expected detection limit (see examples in Appendixes C and D). Measuring LLoD as a function of LoB ensures that the signal from a specimen with low levels of analyte can consistently be distinguished from the background signal that occurs in the measurand’s absence.[112] LoB and LLoD determinations for medical laboratory assays, verification of claimed limits, and the proper use and interpretation of the limits are discussed in CLSI document EP17.[89]

For qualitative tests, LLoD can be determined by probit analysis. Probit analysis transforms the concentration-response curve into a straight line that can be analyzed by regression using either least squares or maximum likelihood.[112] A common method for calculating LLoD is to plot the measurand's log concentration vs the percentage of replicates detected at that concentration. The LLoD is the concentration at which the best fitting line crosses 95% detection probability (see CLSI document EP17[89]).

2.5.9.4 Analytical Specificity

Data analysis for analytical specificity includes identifying cross-reacting substances, organisms, or genetic targets and the level at which the cross-reaction occurs.

2.5.9.5 Interfering Substances

Data analysis for interfering substances includes identifying interfering substances and the level at which interference occurs.

2.5.9.6 Measuring Interval

Data analysis for the measuring interval is generally applicable to quantitative assays and not qualitative assays. For qualitative assays, the measuring interval is the LLoD for each target as the lower end and the highest amount of input measurand that can be assayed as the upper end.

2.5.9.7 Reference Interval

Data analysis for the reference interval is generally applicable to quantitative assays and not qualitative assays. For qualitative multiplex molecular tests, the reference interval is typically "not detected" for disease targets and alleles or the normal genotype(s) in the population.

2.5.9.8 Diagnostic Sensitivity

A detailed discussion of data analysis for diagnostic sensitivity is beyond this guideline's scope. In general, if the definitive diagnosis is known, diagnostic sensitivity is calculated as the percentage of positive test results for patients with the known diagnosis (see CLSI documents MM01[1] and MM19[31]).

2.5.9.9 Diagnostic Specificity

A detailed discussion of data analysis for diagnostic specificity is beyond this guideline's scope. In general, if the definitive diagnosis is known, diagnostic specificity is calculated as the percentage of negative test results for patients without the known diagnosis (see CLSI documents MM01[1] and MM19[31]).

2.5.9.10 Other Performance Characteristics

Data analysis for other performance characteristics depends on the specific studies performed. In general, the parameters yielding acceptable or unacceptable test performance are determined. For example, the laboratory determines the maximum length of time that a sample can be stored at various temperatures or the maximum number of freeze-thaw cycles the sample can undergo while still yielding acceptable test performance.

2.6 Analytical Validation of a Modified *In Vitro* Diagnostic Assay

An IVD assay is considered modified when any change is made to a preexamination, examination, or postexamination feature of the assay.

Changes to an IVD assay may include:

- Test inputs (eg, specimen type, collection device)
- Analysis procedure (eg, reagents, incubation conditions)
- Interpretation (eg, analysis software), as well as the intended patient population and clinical use

Laboratories performing modified tests must demonstrate acceptable performance characteristics with the modification.

For assays that have been developed and previously validated in-house (LDTs), the requirements are less stringent. However, it is good practice to fully assess changes through a change control process before use in laboratory testing. If the laboratory director considers the change to have a potential material effect on the assay performance for its intended use, the assay should be revalidated for the change. Special consideration should be given to multiplex assays because of the increased complexity associated with analyzing more than one measurand and the increased potential for one or more of the measurands being affected by any change.

Examples of changes that can result in an IVD assay being considered modified are discussed in the following subchapters. In general, such changes should also be considered as needing a change control or revalidation for LDTs. An example protocol for validating a modified multiplex IVD assay is provided in Appendix D.

2.6.1 Specimen Type

The specimen type has the potential to affect an assay. For example, blood is a commonly used specimen that typically yields good-quality, high-yield nucleic acid for downstream analysis. Recently, buccal cells (eg, cells collected on swabs or with saliva kits) have become relatively common specimens for DNA-based testing. Buccal cell specimens tend to have lower quality DNA and contain variable, but often high, amounts of microorganism DNA. Hence, the effective yield of the DNA of interest is lower. This is one of many possible reasons why a change in specimen type needs reassessment for an IVD assay and should be considered for an LDT.

Tissue samples can be more complex. For example, when an assay is validated for assessing FFPE lung tissue, the assay has applicability to other sample types as long as sufficient testable material of appropriate quality can be obtained and as long as testing of such tissues is appropriate for the assay's intended use. This assumes that other important preexamination factors are constant. It is important to consider whether the cell line used contains genomic rearrangements that may produce artificial binding sites for primers or probes, possibly generating false positives.[113]

If an IVD assay is regulatory approved or cleared only for lung tissue, testing an alternate tissue (eg, FFPE colon-derived tissue) would modify the device. If the assay was approved or cleared for all FFPE tissue, the device would not be considered modified.

Multiplex assays contain many measurands, and some may be affected more than others by a change in specimen type. For example, to detect a specific measurand, the DNA may need to be largely intact, but samples from certain specimen types may yield highly degraded DNA. Therefore, when the need for revalidation studies is assessed, specimen type and DNA quality and quantity should be considered. Test performance with the new specimen type should be assessed for all assay targets, or at least representative

targets. The potential for false negatives due to matrix inhibition or false positives due to endogenous substances should also be assessed.

2.6.2 Collection Device

Collection devices can also affect overall assay accuracy and performance. For example, for multiplex assays, the collection device may affect some measurands more than others or may contain interferents for one or more measurands. Considerations are similar to those described in Subchapter 2.6.1 for specimen type (ie, if the collection device is changed, this is a device modification and should be considered as an event that requires reassessment, or possibly revalidation, depending on the extent of the change, for an LDT).

Some assays are regulatory cleared or approved for use with a specific collection device. Typically, the device is independently cleared or approved as well as cleared or approved for use with the specific assay. Hence, it is important to understand that changing the collection device but not the assay, even if another cleared or approved collection device is substituted, is still considered to constitute a modified IVD assay.

2.6.3 Test Component

All changes to any test component, even if the reagent is theoretically the same but sourced from a different vendor, result in a modified IVD assay and should be considered for reassessment if it is an LDT. Typically, reagents for LDTs are separated by the laboratory director into those that are critical and noncritical. Given the multiplex nature of many assays, the critical reagent list can become extensive, and reagents that may be considered noncritical for a singleplex assay may be categorized as critical for multiplex assays. Examples of critical reagents include primers used to isolate or detect the measurands of interest or enzymes involved in DNA modification or copying. Those deemed noncritical are often not part of the assay chemistry (eg, a buffer used to run a gel to complete a QC check in between assay steps). The test performance with the new test component must be assessed for all of the assay targets, or at least representative targets for complex multiplex assays. The potential for false negatives due to interference or competition between measurands or for false positives due to endogenous substances should be assessed.

2.6.4 Analysis Software

Multiplex assays increasingly need analysis software or a pipeline of analytical tools that are used to process the data and to present them to the laboratory director in a meaningful and manageable way. These software and tools should be considered an important test component. Similar rules apply to those stated above: If an IVD assay is regulatory cleared or approved, any change to the analytical tools or software by the manufacturer should be verified. Any internal change to such tools leads to the assay becoming a "modified IVD assay" and hence subject to validation as an LDT.

2.6.5 Analyzing Data for Modified *In Vitro* Diagnostic Assays

As stated in Subchapter 2.6, an important factor to evaluate is the modification's effect on assay performance. Depending on the modifications, the analysis may be as elaborate as that of an LDT or may be more limited, similar to that of an unmodified IVD assay. Two types of analysis may need to be performed, including data analysis compared with the unmodified IVD assay and data analysis compared with the reference method or comparator test performed on that specimen type. For significant modifications (eg, different specimen type), the assay's performance, including analytical sensitivity and analytical specificity for each target in the multiplex assay, should be compared with the sensitivity and specificity of the unmodified multiplex IVD assay. In that case, the 2×2 contingency table for validation data analysis can be created by comparing the modified vs unmodified IVD assay for each measurand (see Table 10).

Table 10. 2 × 2 Contingency Table for Individual Measurand A

Modified Multiplex Assay	Unmodified Multiplex Assay	
	A Absent	A Present
A Negative	146	4
A Positive	4	46
Total	150	50

In a second analysis, sensitivity and specificity for each target or group of targets (see Subchapters 2.5.3 and 2.5.4) in the modified IVD multiplex assay should be evaluated to a reference method similar to the analysis performed for an LDT (see Table 11).

Table 11. 2 × 2 Contingency Table for Individual Measurand A

Modified Multiplex Assay	Reference Method	
	A Absent	A Present
A Negative	146	4
A Positive	4	46
Total	150	50

In addition to accuracy, the modification's effect on the assay's LLoD with the new specimen type should be analyzed by comparing the LLoD with the IVD specimen type to the LLoD with the new specimen types. For multiplex assays, the new specimen type LLoD needs to be determined for each target and can be compared with that reported in the unmodified IVD test's manufacturer product insert. Alternatively, parallel dilution studies for each target in both the IVD specimen types and the specimen type under study may be performed to evaluate the change, if any, in the assay LLoD. The laboratory director should establish acceptable criteria for confidence level and CI that may be set using the statistical assessment described in Subchapter 2.4.2.

2.7 Verification of an *In Vitro* Diagnostic Assay

For an unmodified IVD assay, the laboratory must confirm with a high level of confidence that the system performs as claimed when used by the personnel who routinely perform patient testing. The IVD manufacturer has already validated the multiplex assay, and the laboratory must only verify that the established performance specifications (eg, accuracy, precision, measuring interval, reference intervals, for the intended population[s]) can be achieved when performed under its laboratory conditions by the laboratory's personnel. Specific verification activities may be outlined by regulations or standards governing laboratories.[29,37,105] If a medical laboratory using an unmodified IVD assay cannot verify the expected analytical performance specifications, the assay needs to be validated. The laboratory director needs to determine whether the new analytical performance characteristics are acceptable for the intended clinical use(s) before the laboratory implements the test. Medical laboratories also must meet accreditation requirements, which may incorporate verification elements.[107,114] If a laboratory chooses to modify an IVD assay for off-label use, it must perform a validation for the modification before patient testing. For a more detailed discussion, see CLSI document MM03.[2] An example protocol for verifying a multiplex IVD assay is provided in Appendix E.

2.7.1 Analytical Verification

The goal of verification studies is to confirm that the laboratory can perform the test correctly and that the performance characteristics meet those specified by the manufacturer. For multiplex tests, verification of the test accuracy and precision should be performed for each target on the panel. However, depending on the total number of targets on the panel, verifying performance specifications for each target may not be

feasible or cost effective, especially for rare targets. Different approaches may be considered as described in Subchapter 2.4.2. For IVD assays that have already undergone extensive testing, the laboratory does not need to reestablish assay performance characteristics for each measurand but rather establish that the assay performs as expected for each measurand on the panel in the intended patient population.

2.7.2 Analyzing Verification Data

For unmodified IVD multiplex assays, data obtained during verification processes need to be compared with the manufacturer's claims as stated in the assay package insert for:

- Accuracy
- Precision
- Sensitivity
- Specificity
- Measuring interval
- Reference interval

Performance characteristics for each target on the panel need to be evaluated. Data analysis methods as described for singleplex qualitative assays in CLSI document EP12[34] may be used. Depending on the comparator method (reference vs nonreference), either sensitivity and specificity (if comparator is a reference standard) or positive and negative agreement (if comparator is a nonreference standard) should be determined.

Sensitivity and specificity can be calculated using a 2×2 table for each target, as described in Subchapters 2.5.9.1 and 2.6.5. For rare or uncommon targets for which calculations were not obtained during the verification, a review of peer-reviewed literature may be used to verify performance as stated in the manufacturer's product insert. A benefit of multiplex assays with several targets is that positive results for one target may be evaluated as negative results for other targets, increasing the total number of test results used for calculating specificity or negative agreement. If cross-reactivity is a limitation stated by the manufacturer, it should be considered during data analysis. A multiplex assay under evaluation might produce indeterminate or equivocal results for some or all of the targets on the panel. In this case, data should be analyzed depending on how the results might be used for clinical care. If an equivocal result will be interpreted as positive, that result should be analyzed with positive result data. If the opposite is true, the result should be analyzed as a negative result. Analysis of invalid results—which can occur for several reasons, including internal control, reagent, or instrument failure—should not be used in the data analysis. However, the invalid or failure rate for each target and for the test as a whole should be determined, because this information plays an important role in understanding the assay performance.

Assay precision (interassay and intra-assay, if applicable) is calculated from repeat testing as described in Subchapter 2.5.9.2. Assay precision may be expressed as % CV and compared with that reported by the manufacturer. Similarly, measuring interval (if applicable) and reference interval (normal vs abnormal) should be confirmed.

Acceptance criteria for each performance specification should be established when the verification plan is prepared. For example, it may be decided that agreement between the test being verified and the reference test should be 90% or greater for each target and for all combined target tests. However, it may not be possible to demonstrate acceptable performance for all measurands, and assessment of poor performance for one or a few measurands needs to be interpreted within the clinical significance and expected effect of performance in the patient population. Multiplex molecular assays may have limitations in certain situations and for certain measurands, and these limitations should be clearly described (see Subchapter 2.8), with alternative methods available (eg, sending to a referral laboratory) when needed. Because a high number of positive samples may be needed to reach statistical significance, the complexity and the challenge for a

multiplex assay is substantial. The agreement strength between methods (kappa value) and 95% CI should be calculated.

2.7.3 Clinical Verification

Clinical verification is performed to establish an assay's clinical validity and utility. Diagnostic sensitivity and diagnostic specificity are closely related to the PPV and negative predictive value, respectively.[90] For IVD molecular tests, clinical verification is not required. However, this practice helps the laboratory to better understand the assay's effect on patient care.[115] In addition to analytical performance, IVD test manufacturers may attach clinical utility claims to their assays. Specific guidelines on how the test's diagnostic sensitivity and diagnostic specificity should be established have not been published, and clinical verification should be performed at the laboratory director's discretion. This is especially relevant when a test is applied to a population other than the one intended, such as testing for genetic variants in different tumor types. To assist in this process, two questions may be asked:

- What is the assay's intended purpose?
- Does the assay fulfill the intended purpose?

An example is clinical verification of a multiplex syndromic infectious disease assay. The intended purpose would be to aid in the syndrome's diagnosis (eg, gastroenteritis or bacteremia) in conjunction with other clinical data. To evaluate the assay's diagnostic sensitivity, patient charts might be reviewed to determine if symptoms consistent with the syndrome (eg, diarrhea or fever) are present in patients with positive results. Similarly, diagnostic specificity might be evaluated for patients with symptoms consistent with the syndrome (ie, diarrhea or fever) who were determined to be negative for the pathogen(s) of interest by reference testing.[115] Answering these questions may be challenging, especially for complex multiplex panels. When clinical verification of the assay is performed, important variables to consider include:

- Local prevalence of the targets detected by the multiplex assay
- Host factors (eg, in the clinical context, immunocompromised vs immunocompetent patients, or adult vs child)
- Outbreak situation for infectious diseases

Additionally, the probability and clinical significance of recovering or detecting multiple targets or mutations also need to be considered and a baseline or acceptable cutoff established (eg, one patient could be infected with multiple pathogens on the panel). Multiplex assay clinical verification might be particularly challenging, especially for rare or uncommon targets, and peer-reviewed publications for clinical verification of the methods may be used.[90]

2.8 Validation and Verification Summaries

Validation and verification study summaries should briefly present the assay design, studies performed, and relevant results. Tables and graphs can be an efficient way to present large amounts of primary data obtained during these studies. Any deviations from the validation or verification plan should be noted, and performance specifications for the assay should be clearly defined.

Assay limitations can severely affect the analysis of any assay. Limitations can be inherent to the assay (eg, methodology, assay components, primer location and sequence, amplification efficiency or reaction kinetics, LLoD, sample type limitations, sample processing effects, nucleic acid purification method) or can exist because of the sample. Some of these issues are discussed in CLSI document MM03.[2]