Surrogate Sample

DRAFT Framework for Public Comment

MEDICAL DEVICE INNOVATION CONSORTIUM

Align › Achieve › Accelerate

Clinical Diagnostics Surrogate Sample Working Group
4-5-2017
Introduction

The efficiency, predictability, and cost of *in vitro* diagnostic (IVD) development directly impacts patient access to IVDs. The collection of human specimens for testing is a central aspect of IVD validation. When specimens are difficult to obtain or the use of specimens introduces variables into the evaluation of analytical or clinical performance, surrogate samples may be used to improve the timeliness and efficiency of IVD validation.

The use of surrogate samples as an element of IVD analytical and clinical studies is not new. A review of historical data, knowledge, and experience of IVD innovators demonstrates surrogate samples play an important role in the validation of IVDs\(^1\). However, the lack of a common language and framework regarding the use of surrogate samples in IVD performance studies reduces the predictability, timeliness, and efficiency of IVD development, thus increasing cost and reducing patient access. The Surrogate Sample Framework sets forth a common language and methodologies for the use of surrogate samples in IVD analytical and clinical performance studies.

Reasons to use Surrogate Samples

IVD innovators rely upon surrogate samples for a variety of reasons. Surrogate samples provide an option when naturally occurring specimens of biological origin are rare or difficult to acquire. For example, for quantitative tests, acquiring a sufficient number of specimens across the measurement interval of a test, and especially at the high end, the low end, or medical decision points may prove challenging when there is a low prevalence of such specimens in the general population. Surrogate samples can help address these challenges.

Ethical considerations, such as invasive sampling methods or drawing large volumes of clinical specimens from sick patients, are another reason to use surrogate samples. Additionally, surrogate samples may simplify sample acquisition and characterization activities, as well as meet scientific demands of studies requiring large volumes of a sample. In summary, surrogate samples are used for a variety of reasons including, but not limited to:

- Inadequate stability of naturally occurring specimens
- Minimizing biological variability to meet the scientific rigor demanded by the study
- Analyte does not naturally occur at the level demanded by the study, for example, at the extremes of the measurement interval of a test or at the medical decision level (MDL) or test cut-off
- Insufficient volume of naturally occurring specimens
- Covering the measurement interval of a test
- Supplementing the number of positive samples to complete studies

**Terminology**

The authors of this paper elected to use the phrase “surrogate sample” to describe the samples that are the subject of this Framework. A review of relevant IVD and medical device vocabularies (e.g., JCGN, ISO, CLSI) revealed that a common definition of “surrogate sample” does not exist. Thus, the authors chose to define the term, adopting the following definition:

**Surrogate Sample**: material used as a substitute for body fluid or tissue taken for examination from a patient.

Examples include, but are not limited to:

- Materials supplemented (spiked) with an analyte of interest
- Pooled specimens of biological origin
- Material created to have functional characteristics of a body fluid or a tissue
- Material comprised of a combination of an analyte that emulates the analyte of interest and a material created to have the functional characteristics of a body fluid or tissue

Within this broader definition, surrogate samples are further characterized as:

- **Supplemented**: Individual specimens of biological origin spiked with an analyte of interest
- **Pooled**: Combined individual specimens of biological origin that may or may not be spiked with an analyte of interest
- **Simulated**: Treated materials of biological origin or artificial materials created to have functional characteristics of body fluid or tissue that may or may not be spiked with an analyte of interest

Terms such as contrived, altered, supplemented, or simulated are frequently used within the industry interchangeably with the term surrogate or as a subcategory of samples within the broader category of surrogate samples. As observed by the authors, the use of different terminology can impede the development of sound, scientific strategies for the use of surrogate samples. Therefore, we recommend adoption of the above definition of surrogate sample by the industry, including recognition and use in standards, guidance, and other relevant vocabularies.

For purposes of this Framework, we have opted to use the more commonly used term “analyte,” as defined in ISO 17511, rather than term “measurand.” Additionally, we have decided not to use the term “commutability” to describe a demonstration of comparable performance between a surrogate sample and specimen.
Commutability has a distinct definition in the industry and has been deemed by us to be more relevant to studies of method comparisons rather than for establishing validity of results from patient specimens by a specific test. Thus, we have not used the term “commutability” in the context of surrogate samples for purposes of this Framework.

We refer readers to the glossary for definition of other terms used in this Framework. Relevant IVD and medical device vocabularies (e.g., JCGN, ISO, CLSI) have been used to define terms.

Framework Organization

This Framework begins with general principles applicable to surrogate samples and discussion of a general surrogate sample hierarchy. It is further organized by performance study type. Each section expands upon the use of surrogate samples and related points to consider that are unique to the study. The reader is advised to consider the objectives of each study type, and recognize that what may be a suitable surrogate sample type for one performance study is not necessarily suitable for another study.

General Principles for Design and Selection of Surrogate Samples

1. Identify the objective for using surrogate samples.
   a. Supplement specimens
   b. Replace specimens

2. Identify critical factors related to the performance study type, such as:
   a. Preserve biological variability
   b. Remove or reduce biological variability
   c. Meet scientific demands of study type requiring large volumes of sample
   d. Introduce a variable that can only be studied in a surrogate setting to understand how the test performs under the surrogate conditions

3. Consider factors based on whether the test is qualitative or quantitative

4. Consider the effect of extraction method, if applicable, in increasing or decreasing biological variability

5. When selecting the surrogate sample matrix consider which functional characteristics of the naturally occurring specimen are most relevant for the study type. Consider whether there are aspects of the surrogate sample that would interfere or otherwise influence test performance in a way that is

---

inconsistent with naturally occurring biological specimens. When the matrix is artificially prepared, describe the characteristics of the naturally occurring specimen that are represented or mimicked by the surrogate sample. Consider factors to minimize differences between specimens and surrogates, such as viscosity or endogenous cellular components.

6. Consider existing CLSI guidelines regarding the use of surrogate samples specific to the study type and FDA guidance regarding the use of surrogate samples specific to the analyte type.

7. Keeping the above principles in mind, consider a hierarchical approach to the selection of a surrogate to identify a surrogate similar to the specimens that will be used in clinical testing.

When constructing surrogate samples, it may be necessary to employ special methodologies. For example, when formalin-fixed paraffin embedded (FFPE) tissue is the specimen type, as might be used for cancer diagnostic tests, DNA or RNA containing the mutations, insertions, deletions, or sequences of interest may be added/spiked into the samples. This could include cell line DNA, plasmids, or transcripts cell lines that are formalin fixed and paraffin embedded. For swab specimens, addition of the analyte directly to the swab may not be feasible. In these cases, the analyte may need to be added to an appropriate liquid matrix in which various dilutions are then coated onto the swab, or surrogate swab samples may be created by dipping a swab in a solution or collection media spiked with the analyte of interest.

While the use of surrogate samples is well-established for certain technologies, analytes, or studies, there are other areas where the use has been less frequent. In these areas, a demonstration of functional characterization may prove useful in providing scientific evidence to support the use of a particular surrogate sample type.

Surrogate Sample Hierarchy

CLSI Guideline EP06-A “Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline” establishes a hierarchy to consider when selecting surrogate samples for linearity performance evaluations. We believe this hierarchy provides a reasonable path for consideration for all other study types as well. In concert with three categories of surrogate samples, identified above—supplemented, pooled, and simulated—and consideration of CLSI EP06-A, we have constructed a general surrogate sample hierarchy to serve as a starting point for all study types. Therefore, after considering the general principles and the performance study specific objectives, evaluate the hierarchy and select the best option for surrogate sample preparation and use. Refer to CLSI guideline EP06-A for specific elements to consider when preparing surrogate samples within the hierarchy.
Figure 1

[Diagram showing sample matrix and preferred usage hierarchy]

1. Under special circumstances, animal-sourced sample matrix could be acceptable.

2. The large arrow and staggering of the descriptive boxes represent general preferential directionality; however, actual preference of a sample type is situationally-dependent and may vary from that indicated by this figure.
**Performance Study Types**

In addition to the general principles, there are factors to consider in conjunction with specific performance study objectives. Additionally, application of the hierarchy varies with specific performance study objectives. This Framework outlines these factors and the application of the surrogate sample hierarchy by performance study type.

Based on our research\(^3\), the use of surrogate samples has been more prevalent in certain study types. Whereas in other study types, the use has been more situation dependent\(^4\). This Framework reinforces common usage embodied in standards and practice. Further, it is designed to formalize collective scientific knowledge through the use of uniform principles and terminology for those study types where surrogate sample use has been more prevalent. For those study types, in which surrogate sample use has been more situation dependent, this Framework sets forth practical applications aligned within the general principles and surrogate sample hierarchy.

This Framework does not provide comprehensive guidelines on study design, unless relevant to the use of surrogate samples. Readers are referred to scientific standards, guidance documents, and guidelines for factors to consider when designing performance studies.

The following performance studies are addressed in this Framework:

1. Linearity/Assay Analytical Measuring Interval
2. Analytical Specificity/Interference/Cross-Reactivity/Competitive Inhibition
3. Precision/Reproducibility
4. Detection Limit
5. Matrix Comparison
6. Method Comparison
7. Specimen Stability
8. Reagent Stability
9. Instrument Carry-Over Studies
10. Clinical Performance

---


Linearity / Assay Analytical Measuring Interval

Study Type
Linearity/assay analytical measuring interval is the range within a test providing results that are directly proportional to the concentration of analyte in the test sample.

Objective
The objective of this study is to determine the range across which the average value of test replicates is proportional to the true value of the concentration of the analyte. Increasing the number of replicates at each analyte concentration or increasing the number of analyte concentrations across the target range can be used for more definitive examinations of linearity.

Principles of the Use of Surrogate Samples
Surrogate samples are widely used in linearity/assay analytical measuring interval studies to produce samples in which there is a known proportional relationship in a representative matrix and that span the desired range of the test. The preferred type of surrogate sample is a high concentration specimen diluted with a negative patient sample to achieve the known proportional relationship of the analyte concentration. Sometimes this will not be possible and the surrogate specimen hierarchy should be followed, selecting the most preferred surrogate sample to help ensure the surrogate sample is representative.

How to Use
To provide sufficient volume of sample with an analyte concentration at different levels across the test’s measuring interval, surrogate samples can be prepared in a number of ways, including: dilution of a specimen with a high concentration of analyte with a matrix that has a very low amount of analyte; dilution of pure analyte with a matrix that has a very low amount of analyte; dilution of a specimen or pure analyte in a surrogate matrix. Preservation of patient-to-patient variability is a key consideration.

When to Use
Because the linear range experiment requires enough samples to prepare dilutions across the linear range, the use of surrogate samples is generally accepted. In all cases the surrogate samples should be representative of a specimen.

Additional Considerations
When using surrogate samples for linearity/assay analytical measuring interval preservation of biological variability should be considered. For this reason, first consider use of a large volume from one patient diluted prior to the use of pooled samples.

Reference
Analytical Specificity (Interference, cross-reactivity, competitive inhibition)

Study Type
Analytical Specificity (Interference, cross-reactivity, competitive inhibition) studies are conducted to evaluate the ability of a measurement procedure to determine solely the quantity it purports to measure.

Objective
The objective of this study is to demonstrate that test performance is not affected by potentially interfering materials present in samples.

Principles of the Use of Surrogate Samples
Surrogate samples should be used when clinical specimens are not available or if it is not possible to obtain specimens with known amounts of potentially interfering material.

How to Use
For analytical specificity, there are two types of samples to which potentially interfering materials could be added: naturally occurring human specimens and surrogate samples. Once the potentially interfering materials have been added to either sample type the result is a surrogate sample that can be used to assess analytical specificity with the test. In some cases, the surrogate sample could be composed of a surrogate analyte, a surrogate sample matrix, and a potentially interfering substance.

It should be noted that not all potentially interfering substances are added to a sample. Potentially interfering substances may be introduced through the preparation of a sample. For example, the fixation process for formalin-fixed paraffin embedded (FFPE) tissue may introduce materials that interfere with test performance or otherwise diminish the integrity of the sample. In these cases, the potential interference can be evaluated through the comparison of specimens or surrogate samples processed using different procedural conditions.

When to Use
In order of recommended preference, surrogate samples used for analytical specificity should be as similar to a clinical specimen as possible. As a consequence, surrogate samples that are categorized as supplemented with interferent should be given the highest priority, followed by the pooled category, and finally the simulated category.

In some cases, the potentially interfering material may not have an established method with which the material concentration can be measured. For these surrogate samples it is acceptable to determine the concentration of the potentially interfering materials based on the amount added.

In other cases, the addition of the potentially interfering material may not properly reflect physiological conditions. For example, adding a drug or diluted drug to a sample may not yield the biologically active metabolite, adding triglycerides may not reflect chylomicron or micelle formation that would happen in a naturally occurring lipemic specimen, or the binding constants for the potentially interfering material may influence test results as with the use of caffeine or theophylline when measuring free versus protein bound bilirubin. For these surrogate samples the sponsor may need to refer to literature for ‘best practices’ based on
historic or otherwise reasonable expectations of how the potentially interfering material has been used in surrogate samples.

Finally, some potentially interfering conditions do not have a corresponding surrogate counterpart. For example, necrotic tissue in FFPE tissue cannot be emulated using a surrogate sample. In these cases, there is no suitable surrogate sample and clinical specimens containing the potentially interfering conditions should be used.

Reference

Precision/Reproducibility

Study Type
Precision/reproducibility studies are conducted to establish the within-laboratory precision and site-to-site variability (reproducibility) of a test. In addition, evaluation of different components of variance as within-run (repeatability), between-run, between-day, and between-lot are also conducted.

Objective
Precision/reproducibility studies are used to establish the level of agreement between repeated measurements on the same or similar sample under different conditions. These different conditions can include factors such as analyst, instrument, time, location, or conditions of use.

Principles of the Use of Surrogate Samples
Surrogate samples are widely used in precision/reproducibility studies to obtain sufficient sample with consistent levels of analyte to perform the required replicate measurements. The preferred type of surrogate sample for precision/reproducibility studies are pooled patient samples or spiked individual patient samples. If the desired amount of analyte is not available in the pooled patient samples, the concentration of analyte could be increased with spiking or decreased by dilution with a negative sample pool.

How to Use
To provide sufficient volume of sample with a known analyte concentration at different levels across the test’s measuring interval, surrogate samples can be prepared in a number of ways, including: spiking individual patient specimens, pooling of samples of similar concentrations across the measuring range, spiking a pooled sample with a high concentration analyte, or dilution of a specimen with a high concentration of analyte with a matrix that has a very low amount of analyte; dilution of pure analyte with a matrix that has a very low amount of analyte; dilution of a specimen or pure analyte in a surrogate matrix.

When to Use
Due to the volume of sample required to perform precision/reproducibility testing the use of surrogate samples is generally accepted. In all cases the surrogate samples should be representative of a specimen.

Additional Considerations
The benefit of individual spiked patient samples to better represent the variability between samples should be considered when selecting the type of surrogate samples to use. Spiking some amount of analyte into patient individual specimens would retain the matrix variability.

Reference
- CLSI EP05: 2014 (Evaluation of Precision of Quantitative Measurement Procedures) recognizes that typically for precision studies pooled patient specimens are used and that the spiking or dilution of samples is necessary to achieve analyte levels at the appropriate levels.
Detection Limit (LoB, LoD, and LoQ)

Study Type
Test developers must determine the ability of a test to detect an analyte at the limits of test performance and accurately determine when an analyte is not present.

Objective
Detection capability studies are used, depending on the type of test, to establish the Limit of Blank (LoB), Limit of Detection (LoD), and/or Limit of Quantitation (LoQ). The studies are usually the first analytical studies conducted in order to determine the appropriate spiking level of an analyte for other analytical studies (e.g., interfering substances, reproducibility) for qualitative tests.

Principles of the Use of Surrogate Samples
The matrix selected should closely represent a true, individual clinical specimen. It is important that the matrix selected for conducting detection capability studies does not contain extraneous properties or elements that would contribute to the variability of results that would not typically be encountered when testing clinical specimen or would otherwise confound test results obtained. Such exogenous substances may introduce bias that could impact the ability of the study to establish the “true” detection limit of a test or test system. If the matrix must be artificially prepared, the characteristics of the surrogate specimen should be clearly defined so that the surrogate matrix resembles an appropriate patient specimen.

How to Use
For evaluation of LoB, for most material, it is preferred to collect individual patient specimen(s) known to be negative for the intended analyte(s). The analyte intended for measurement is then added to the matrix either directly from a high concentration stock or using a clinical specimen known to contain the intended analyte(s), if quantitatively measurable in the latter case. For qualitative tests, from this spiked sample, serial dilutions can be made to create the varying levels required to determine the detection limit; dilutions should preferably be made with the unspiked, negative sample matrix. Should the required volume requirements not be met with a single-source sample, pooling a minimal number of clinical specimens is preferred to preserve biological variability. If a large amount of a known negative specimen is hard to obtain (e.g., cerebral spinal fluid), then a simulated artificial matrix that closely resembles the naturally occurring specimen can be used.

When using negative patient specimens, it is important to reduce bias associated with the variability introduced by such specimens. Specimens obtained from different patients in a clinical setting, collected to create a representative clinical sample matrix pools, may contain different levels of endogenous and/or exogenous components based on their genetics and/or disease state as well as different substances used to manage their disease. Although presence of the endogenous elements may be beneficial for establishing analytical sensitivity in the intended use population, in excess these substances may create a matrix effect, which is a phenomenon that may generate bias in the results obtained from the study. Patient specimens used to create negative pools should be free of obvious, potential interferents that generate bias. It is acceptable to dilute positive specimens or spike the analyte into negative specimens in order to provide low-level samples at desired levels of analyte. When spiking analyte(s) into the sample matrix, high concentrations of naturally occurring or artificial material is preferred, to limit any initial dilution effects.

In several situations, (e.g., qualitative ID molecular/NAAT assays), the ability to create reproducible negative sample matrix pools is critical. Therefore, it is preferred to use the same pooled/supplemental sample matrix that is spiked at concentrations representing multiples of the LOD determined for the test (e.g., at
concentrations 2 to 3 times the determined LOD). This approach will help reproduce variability in results associated with different specimen pools.

**When to Use**

When establishing analytical detection limits, surrogate specimens are needed to reliably determine these limits since in many cases sufficient volumes of clinical specimens containing known concentration(s) of the analyte(s) of interest cannot practically be found in a clinical population.

**Additional Considerations**

Preparation of surrogate specimens for use in detection limit studies may be challenging as the method of preparation may introduce bias. For example, the spiking of very low levels of analytes may be challenging as the ability to reliably add small amounts of the analyte may exceed the capability of the system used to add these low levels. Therefore, it may be necessary to create serial dilutions of the matrix, whereby the analyte is added at a high level to a matrix that is then diluted multiple times to achieve the appropriate level.

Some specimen types may create additional challenges where special methodologies may be required to assess the detection limit. For example, when formalin-fixed paraffin embedded (FFPE) tissue is the specimen type used for cancer diagnostic tests, it may be necessary to provide insertions or deletions of one or more nucleotides. For swab specimens, the analyte may not be able to be added directly to the swab. In these cases, the analyte may need to be added to an appropriate liquid matrix. Various dilutions are then coated onto the swab.

Additionally, as large specimen pools are used for such detection limit studies, the stability of the specimen matrix should be established so as not to introduce result variability through degradation of the matrix.

**References**

Matrix Comparison Study

Study Type
Matrix comparison is a study for demonstration of similarity of specimen types.

Objective
The purpose of the matrix comparison study for a Candidate type of samples (e.g., plasma) and a Primary type of samples (e.g., serum) for a quantitative test is an evaluation of a systematic difference between these two types of samples at medical decision levels (MDLs). For a qualitative test, the purpose of the matrix comparison study is an estimation of positive and negative percent agreements of the test results with the Candidate samples vs test results with the Primary samples for a population. In the matrix comparison study, usually every patient in the study provides Primary and Candidate specimens which are measured by the test and results are compared.

When Surrogate Samples can be Considered for Use

Covering the measuring interval of the test
For a quantitative test, test values with Primary samples should cover the measuring interval of the test. Sometimes, it is difficult to find patients who have Primary samples with high values or low values. Pairs of surrogate samples can be used in the matrix comparison study.

Patients with Primary samples values close to MDL
In the study, it should be few Primary samples (3-5 samples) with test values close to the MDL. If it is difficult to find patients with Primary samples values close to the MDL, pairs of surrogate samples can be used.

Not enough patients with positive test results for Primary samples
For a qualitative test, an estimation of the positive percent agreement (PPA) requires a certain number of patients with positive test results for Primary samples. If it difficult to find patients with positive test results with Primary samples, surrogate pairs of samples (Primary and Candidate samples) can be added for estimation of PPA.

Samples close to the cutoff
In matrix comparison studies, it is required to have some percent of sample pairs with values close to the test’s cutoff with Primary samples among all positive samples. If it is difficult to find patient samples close to the cutoff, pairs of surrogate samples can be included in the matrix comparison study.

Principles of the Use of Surrogate Samples
For pairs of surrogate samples with high test values with Primary samples in the matrix comparison study of quantitative test, add the minimal amount of analyte as possible to the Primary and Candidate samples.

For pairs of surrogate samples with low test value with Primary samples in the matrix comparison study of quantitative test, add the minimal amount of the sample with zero concentration (as diluent) as possible.

The use of the surrogate samples in the matrix comparison study for a qualitative test should preserve the percent of samples close to the cutoff, which is required in the matrix comparison study. When pairs of surrogate samples are used in the matrix comparison study, first, it should be determined C5 – C95 interval (or LoD) of the qualitative test for defining of “close to the cutoff.” If a spiking or dilution technique will be used to prepare surrogate samples, the relationship between amount of spiking material (or amount of true negative...
patient sample) and how close to the cutoff these Primary samples can be should be understood. If the qualitative test does not have available signals, then the use of surrogate samples can be problematic.

Reference:

- No current CLSI document (CLSI EP35, regarding matrix comparison, is in development, projected publication: end of 2017)
Method Comparison

Study Type
Evaluation of bias or systematic difference between two quantitative or qualitative tests.

Objective
The purpose of the method comparison study for a candidate quantitative test and the comparator test is an evaluation of bias or systematic difference. For a qualitative candidate test, the purpose of the method comparison study is an estimation of positive and negative percent agreements of the candidate test results with a comparator test results for a population. The comparator can be a qualitative test or a quantitative test with an appropriate cutoff.

When Surrogate Samples can be Considered for Use

Not enough volume
In the method comparison studies, each sample needs to be measured by the candidate test and by the comparator test. Sometimes, volumes of neat patient samples do not allow this; therefore, pooling of patients samples are used.

Covering the measuring interval
For a quantitative candidate test, samples in the method comparison study should cover the measuring interval of the test. Sometimes, it is difficult to have samples with high values or low values; surrogate samples can be used in the method comparison study to meet these needs.

Samples close to medical decision levels (MDLs)
For a quantitative candidate test, it should be few samples (3-5 samples) close to the MDL in the method comparison study. If it is difficult to find patient samples close to the MDL, surrogate samples can be used.

Not enough samples positive by the comparator
For a qualitative candidate test, an estimation of the positive percent agreement (PPA) with some level of uncertainty (confidence intervals) requires some number of samples positive by the comparator test. If it is difficult to find patient samples positive by the comparator test, surrogate positive samples can be added for estimation of PPA. An inclusion of surrogate samples can be limited to only some special type of studies and most of the time, archived samples (not surrogate samples) should be considered.

Samples close to the cutoff
In some method comparison studies (e.g., migration studies), it is required to have a certain percentage of samples close to the cutoff of the candidate test among all positive samples. If it is difficult to find patient samples close to the cutoff, surrogate samples can be included in the method comparison study.

Principles in Preparation of Surrogate Samples
Preserve individual biological variability as much as possible (if possible, do not pool samples). For pooling, use each individual sample only one time to preserve variability.

For very high concentration surrogate samples in the method comparison study of a quantitative Candidate test, use patient specimens with the highest concentration of naturally occurring analyte to minimize the use of spiking material available (e.g. high concentration stock solutions). For low value surrogate samples in the
method comparison study of the quantitative Candidate test, use patient specimens with low concentrations
of naturally occurring analyte to minimize the use of diluent.

For obtaining unbiased estimates of positive and negative percent agreements in the method comparison
study of the qualitative Candidate test, the percent of samples close to the cutoff of the Candidate test should
reflect the intended use population or requirements applied for the patient samples (e.g., migration study).
Therefore, the use of the surrogate samples in the method comparison study should preserve the percent of
samples close to the cutoff. First, when surrogate samples are used in the matrix comparison study, it should
be determined C5 – C95 interval (or LoD) of the Candidate test for defining of “close to the cutoff”. Second, the
percent of samples close to the cutoff of the Candidate test should be evaluated in the intended use
population. For this, consider all patient samples in the comparison study, which were randomly selected from
the intended use population, and estimate the percent of the patient samples close to the cutoff of the
Candidate test. The surrogate samples should have the same percent of the samples close to the cutoff. To
determine whether a patient sample result was close to the cutoff, one can use values of the signal from the
Candidate test (if available). If the signal of the Candidate test is not available (e.g., influenza test with visual
reading of the line), consider a signal of the Comparator test (if available) or quantitative values of the
quantitative Comparator. Third, if a spiking or dilution technique will be used to prepare the surrogate
samples, one should understand the relationship between the amount of spiking material (or amount of true
negative patient sample in dilution) and how close to the cutoff these samples can be. If the Candidate test
and Comparator test do not have available signals then the use of surrogate samples can be problematic, for
instance, a method comparison study for influenza test with visual reading of the line (Candidate) vs culture
(Comparator).

Additional Considerations
First, always perform separate statistical analyses for patient specimens and for surrogate samples. If
performances are similar, as described by estimates of biases (systematic differences) at MDLs for quantitative
test or PPA and NPA for the qualitative candidate test, then perform a statistical analysis for all samples
combined.

Note that for a quantitative candidate test, estimates of the slopes, intercepts, and biases (or systematic
differences) at MDLs using patient specimens and using the surrogate samples should be compared. If the
surrogate samples show less dispersion around the regression line compared to the dispersion around the
regression line for the patient samples, then confidence intervals about slope, intercept, and biases at MDL
with combined data can be misstated.

References
- CLSI. Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved
  2013.
Specimen Stability

Study Type
Specimen stability is the ability of a specimen to maintain its measured characteristic(s) within acceptable limits over an established period of time.

Objective
Assess the magnitude of analyte degradation through its measured characteristic(s) over time when stored under defined conditions.

Principles of the Use of Surrogate Samples
Acceptable changes of bias or variability that may arise from sample stability are identified by assessing analyte stability in a known surrogate matrix. Surrogate samples may be used to facilitate specimen stability studies by providing sufficient sample volume to conduct tests or to obtain targeted analyte levels, such as levels near the medical decision level (MDL). In these situations, selection from the surrogate hierarchy begins at the level designated “pooled unspiked” followed next by the level designated “pooled spiked.” When the use of surrogate samples is solely to address sample volume, thoroughly consider the scientific rationale as to why the higher ranking sample type is not feasible before moving to the next level of the hierarchy. When the use of surrogate samples is to address scientific challenges with the functional characteristics of naturally occurring specimens (e.g., whole blood, cerebral spinal fluid), begin at the surrogate sample hierarchy level designated “supplemented” and move down through each level of the hierarchy after thoroughly considering the scientific rationale as to why the higher ranking sample type is not feasible.

How to Use
Although the use of surrogate samples to establish specimen stability is not well-documented in standards or literature, assessment of analyte stability may occur with the use of a well-characterized matrix of known stability and with reagents of known stability. Prepare surrogate samples with very high and very low concentrations of the analyte, as well as at levels at the MDL. When preparing surrogate samples follow well-designed pre-analytical steps to mimic processing steps of native specimens.

When to Use
Consult the surrogate sample hierarchy at Figure 1, so that your selection begins with a surrogate that is most representative of a specimen. Stability of the selected matrix should be known and understood to avoid the introduction of bias.

Additional Considerations
Consider how well the selected matrix resembles a specimen in terms of viscosity or other key elements of the naturally occurring specimen. Well-designed pre-analytical steps to mimic processing steps of native specimens is an important consideration, so that spiked analyte is processed as the naturally occurring analyte in native biological specimens.

Reference
Reagent Stability

Study Type
Stability is the ability of an in vitro diagnostic (IVD) reagent to maintain its performance characteristics within pre-established acceptance limits over time. Manufacturers establish shelf-life and in-use stability for quantitative and qualitative IVD reagents.

Objective
IVD reagent stability is evaluated on the basis of ensuring that key performance metrics continue to meet predefined acceptance criteria within the expiry date. Eliminating or reducing variation of other study parameters permits the evaluation of the IVD reagents. The use of surrogate samples in stability studies plays a key role in minimizing sample variability because it enables the creation of sufficient volume for the total number of tests.

Principles of the Use of Surrogate Samples
Surrogate samples facilitate reagent stability studies because they ensure sufficient volume of sample for testing over the study’s duration. Stability studies assess performance at the extremes of the reportable range, which are often the regions where significant reagent changes are likely to occur, thus affecting stability, as well as at the medical decision level (MDL). Surrogate samples allow for sufficient sample volume targeted at the appropriate levels for testing. When an appropriate surrogate sample type is selected, stability studies may be conducted with one hundred percent surrogate samples. As used in this context, an appropriate surrogate sample is one of known stability duration within the unspiked or spiked with naturally occurring analyte pooled section of the surrogate sample hierarchy at Figure 1. Because IVD reagents may show little or no significant analyte drift when tested with non-native analyte material, design stability studies to use unspiked pooled or spiked with naturally occurring analyte pooled surrogate samples when available.

How to Use
Patient sample pools, controls, and other suitable materials of known analyte stability are recognized as appropriate surrogate sample types and are tested alongside associated calibrators. Surrogate sample types may be neat, diluted, or supplemented to represent the MDL, low level of measuring interval, and high level of the measuring interval for quantitative tests. For qualitative tests, surrogate sample types may be neat, diluted, or supplemented to represent the MDL. Determine surrogate sample stability duration as a prerequisite to use in reagent stability studies.

When to Use
Although the use of surrogate samples is generally acceptable in establishing reagent stability, consult the surrogate sample hierarchy at Figure 1, so that your selection begins with a surrogate that is most representative of a specimen (e.g., patient sample pool). Calibrators and controls may be used as part of the study. However, they should not be used at the exclusion of surrogate sample types ranking higher on the hierarchy without clear scientific rationale as to why the higher ranking sample type is not feasible. For surrogate samples selected from the spiked artificial pooled section or simulated section of the surrogate sample hierarchy, it is advisable to include within the study native human specimen, unspiked pooled surrogate samples, or spiked with naturally occurring analyte pooled surrogate samples comparable to one surrogate sample level or the MDL.

---

Additional Considerations

Describe the surrogate sample in the stability testing plan. Consider differential recognition of the analyte changing over time when using surrogate samples. When able, include native human specimen comparable to one surrogate sample level or the MDL. Consider performance of surrogate samples in other testing to support the use in stability studies when native human specimens are unavailable.

References

Instrument Carry-Over Studies

Study Type
For automated liquid handling systems used in clinical diagnostics, developers must demonstrate that carry-over and cross-contamination will not occur during use of the system with patient specimens. These types of phenomena may contribute to a false negative, false positive, incorrect concentration level, indeterminate, or no result.

It should be noted that these types of studies might not be necessary if system design prevents the possibility of carry-over and cross-contamination.

Objective
The results of carry-over and cross-contamination studies should demonstrate that there is no detrimental impact of the liquid handling system when used with patient specimens on the results of the tests run on the automated analyzer.

Principles of the Use of Surrogate Samples
The matrix selected should closely represent a true clinical specimen. If matrices other than patient specimens are used, the sample may not properly mimic the intended specimen type. For example, the viscosity of the surrogate sample may not accurately reflect the opportunity for specimen/analyte carry-over effects.

Surrogate specimens spiked with a high amount of analyte may be used to simulate high positive (or high concentration) samples. Individual patient samples known to be negative for the target analyte are also used. The samples are to be used in series alternating with negative (or low concentration) samples in patterns dependent upon the operational function of the device.

How to Use
The surrogate samples should be derived from the specimen types intended to be used with the automated system. Therefore, the preferred specimen types are pooled patient specimens that are unspiked and spiked with the target analyte. The surrogate specimens should be spiked with the analyte to an appropriate level in order to measure the impact of sample/analyte carry-over. Based on manufacturer instructions, it may be necessary to process the specimen (e.g., in a sample buffer) prior to analysis in the carry-over study.

Simulated sample matrices should be avoided when evaluating the impact of patient sample carry-over and cross-contamination.

When to Use
When determining the impact of potential carry-over and cross-contamination effects using patient samples, surrogate specimens are needed due to the lack of specimens with known, high concentrations of the target analyte, in the clinical population.

Additional Considerations
Note that some specimen types may have inherent variability that may be difficult to mimic with a surrogate specimen (e.g., excessively mucoidal specimens, sputum, hemolyzed blood samples). It may be necessary to obtain negative specimens with these characteristics and spike in the appropriate levels of the target analyte(s).
References

Clinical Performance

Study Type
Clinical performance relates to the clinical sensitivity and sensitivity of the test.

Objective
Clinical sensitivity is the ability of an IVD test to identify the presence of a target marker associated with a particular disease or condition, which is based on criteria that are independent of the test under consideration. Clinical specificity is the ability of an IVD test to recognize the absence of a target marker associated with a particular disease or condition, which is based on criteria that are independent of the test under consideration. Clinical cut-off is the value that is understood to differentiate two conditions, usually health from disease.

Principles of the Use of Surrogate Samples
Since clinical performance is contingent upon the intended use of the test in defined populations, the use of surrogate samples is typically reserved for use as a means to supplement clinical specimens, rather than replacement. In cases of rare disease, surrogate samples may be considered for replacement purposes. However, in such cases, carefully consider functional comparability. Preservation of biological variability is a key consideration. Thus, it is advisable to begin at the top of the Surrogate Sample Hierarchy when considering the use of surrogate samples for clinical performance studies.

How to Use
Use well-characterized surrogate samples. When preparing surrogate samples follow well-designed pre-analytical steps to mimic processing steps of native specimens.

When to Use
Quantitative Tests
Use when the analyte of interest is absent from specimens of healthy subjects. In those cases, the cutoff is often set as the limit of blank (the lowest concentration of analyte that can be consistently detected) or the limit of quantification (the lowest concentration of analyte that can be quantitatively determined with stated accuracy). Supplementing the analyte of interest into individual specimens (preferred) or pools is commonly used to establish these parameters. The indication for use would prescribe which is appropriate.

Similarly, surrogates can be used to better define the Confidence Interval (CI) for the cut-off values of the reference interval.

Further, with increased use of nucleic acid diagnostic techniques, when “copy” numbers are used to establish a cutoff, supplementing samples with known oligonucleotides are often employed.

Qualitative Tests
Qualitative tests are frequently used for the diagnosis of infectious diseases by identifying an analyte that is not present in healthy subjects. For these tests, the cutoff is defined as the limit of blank. Supplementing the analyte of interest into individual specimens (preferred) or pools is commonly used to establish these parameters.

---
Additional Considerations

Surrogates prepared by supplementing analyte or pooling specimens, defined here, are best prepared with matrices most similar to those of the clinical specimens used by the test. If the analyte is added to samples that are devoid of material of interest, several patient sources are preferable for determining limits of detection. In this way, any bias caused by patient-specific causes can be evaluated and compared to pre-determined acceptance criteria. If studies of aggregate properties are being conducted, such as in an agreement study or a correlation, pools are often more convenient for preparation, especially if a large sample volume is needed. In addition, when analytes are extracted before analysis, such as is often needed for infectious disease or nucleic acid tests, surrogates can be prepared using the extraction medium as the solvent.
Glossary

Analyte: Component represented in the name of a measurable quantity. In this framework, we have used analyte preferentially over the term measurand (ISO 17511)

Analytical Specificity: Ability of a measurement procedure to measure solely the measurand (ISO 17511)

Biospecimen: A biological sample fluid or tissue obtained from an organism (CLSI I/LA28)

Cut-off value (laboratory medicine): Quantity value used as a limit to identify samples that indicate the presence or the absence of a specific disease, condition, or measurand (ISO/TS 17822-1:2014 (en))

Detection Limit (Limit of Detection, LoD): The lowest concentration of analyte that can be consistently detected. (Adopted from CLSI EP17-A2:2012)

i. Limit of Blank (LoB): The highest measurement result that is likely to be observed, with a stated probability (typically, in ≥95% of samples tested under routine clinical laboratory conditions and in a defined type of sample), for a blank sample (CLSI EP17-A2:2012)

ii. Limit of Quantitation (LoQ): The lowest amount of measurand in a material that can be quantitatively determined with a stated accuracy under stated experimental conditions (CLSI EP17-A2:2012)

Interference (analytical): Systematic error of measurement caused by an influence quantity, which does not by itself produce a signal in the measuring system, but which causes an enhancement or depression of the value indicated (CLSI MM17-A:2008)

In-use stability: Duration of time over which the performance of an IVD reagent within its expiration date remains within specified limits after opening the container system supplied by the manufacturer, and put into use under standard operation conditions (e.g., storage on the instrument). (CLSI EP25-A:2009)

Linearity: The ability (within a given range) to provide results that are directly proportional to the concentration (amount) of the analyte in the test sample (CLSI EP06-A:2003)

Matrix (of a material system): Components of a material system, except the analyte (ISO 15193)

Matrix comparison: Equivalence of specimen types

Matrix effect: Influence of a property of the sample, other than the analyte, on the measurement, and thereby on the value of the measurable quantity (ISO 15194)

Measurand: Quantity intended to be measured (ISO 15193)

Method comparison: Evaluation of bias or systematic difference between two quantitative or qualitative tests

Precision (of measurement): Closeness of agreement between independent results of measurements obtained under stipulated conditions

i. Repeatability: Precision under essentially unchanged conditions, often termed “within-run precision”
ii. Reproducibility: Precision under changes in conditions (e.g., time, different laboratories, operators, and measuring systems [including different calibrations and reagent batches]), (ISO 17511)

Reference Range or Reference Interval: The range of test values expected for a designated population of individuals (42 CFR 293.2)

Shelf life: Period of time until the expiration (expiry) date, during which an IVD reagent in the packaging configuration provided to the user maintains its stability under storage conditions specified by the manufacturer (CLSI EP25-A:2009)

Sensitivity (of a measuring system): Quotient of the change in an indication of a measuring system and the corresponding change in a value of a quantity being measured (JCGM 200:2012)

Specimen: Discrete portion of a body fluid or tissue taken for examination, study, or analysis of one or more quantities or properties assumed to apply for the whole (ISO 15189)

Surrogate Material: Material used as a substitute for biologic material without having all of the functional characteristics of such materials (CLSI H26)

Surrogate Sample: Material used as a substitute for body fluid or tissue taken for examination from a patient. Examples include, but are not limited to:

- Materials supplemented (spiked) with an analyte of interest
- Pooled specimens of biological origin
- Material created to have functional characteristics of a body fluid or a tissue
- Material comprised of a combination of an analyte that emulates the analyte of interest and a material created to have the functional characteristics of a body fluid or tissue.

Sources:

42 CFR 293.2 Laboratory requirements. General provisions. Definitions.


ISO 17511: In vitro diagnostic medical devices—Measurement of quantities in biological samples—Metrological traceability of values assigned to calibrators and control materials

ISO 15194: In vitro diagnostic medical devices—Measurement of quantities in samples of biological origin—Requirements for certified reference materials and the content of supporting documentation

ISO 15193: In vitro diagnostic medical devices—Measurement of quantities in samples of biological origin—Requirements for content and presentation of reference measurement procedures

ISO 15189: Medical laboratories—Requirements for quality and competence

ISO 11843: Capability of detection


ISO/Guide 30:2015 (en) Reference materials—Selected items and definitions

JCGM 200:2012 International vocabulary of metrology—Basic and general concepts and associated terms.

International Bureau of Weights and Measures.
MDIC has assembled a work group comprised of member organizations and other subject matter experts to guide work on this project.

**Industry:**

- Khatereh Calleja, JD, AdvaMed
- Tremel Faison, BARDA
- Ronald Freeze, PhD, Abbott
- Ian Giles, MD, Sysmex
- Patrick O’Donnell, Roche
- Brad Spring, BD
- Mark Del Vecchio, BD
- April Veoukas, JD, Abbott (Working Group Chair)

**FDA:**

- Marina Kondratovich, PhD, CDRH | OIR
- Zivana Tezak, PhD, CDRH | OIR
- Yun-Fu Hu, PhD, CDRH | OIR

**Expert Advisors:**

- Susan Alpert, MD, PhD
- Constantine Gatsonis, PhD
- Fred Lasky, PhD

**Program Director:**

- Carolyn Hiller, MDIC

© Medical Device Innovation Consortium 2017