

Performance Standards Reporting Requirements for Essential Assays in Clinical Trials

This document describes the information about in vitro assay and imaging test performance that will be expected in clinical trial *protocols* where such assays or tests are included as integral or integrated assays. The focus is on Phase 3 and large Phase 2 trials. Most, if not all, functional imaging methods can and should be thought of as in vivo assays, and will be subject to most of the same requirements as in vitro assays. The NCI recognizes that it will not be possible for every assay or imaging test to immediately meet all these standards, but they are provided to help guide development plans.

The rationale for imposing assay performance standards requirements is based on:

1. the need to ensure that tests used in trials can be rapidly translated to clinical practice for patient benefit;
2. the need to adhere to regulations that apply to tests to be used to guide clinical decisions; and
3. the principle that use of precious and non-renewable specimens collected in randomized trials should lead to clearly interpretable results that move the field forward.

The levels of evidence required for assays will reflect their role in the clinical trial, i.e., integral assays versus integrated assays. The appendices describe the information specifically required for integral assays. It is also important to note that *concepts* should provide sufficient information for evaluation of the chosen assay and its role in the trial, including specimen requirements for in vitro assays; the final protocol will have to provide all the identified information.

I. Integral assays refer to tests that must be performed for the trial to proceed.

A. In vitro assay(s) must be performed in laboratories with at least a CLIA Certificate of Compliance. The requirements and the information to support the use of the assay are as follows:

- Information to be submitted must include the categories of data that would be required for submission for FDA clearance (510k – substantial equivalence) or approval (premarket application)
 - The role the assay will play in the trial
 - Measurements of precision and reproducibility (within lab and between labs if more than one lab will be performing the assay)
 - Data to support proposed cut-point(s) if assay results are not reported as a continuous variable
 - Analytic specificity (cross-reacting substances, interfering substances, etc.) and analytic sensitivity (limits of detection)
 - Accuracy measurements

B. In vivo imaging assays (i.e., imaging tests) must be performed using standardized guidelines for image acquisition, analysis and interpretation. The requirements and the information to support the use of the imaging test are as follows:

- The role the imaging test will play in the trial.
- Measurements of precision and reproducibility (within lab and between labs if more than one lab will be performing the imaging test)
- Data to support proposed cut-point(s) if imaging results are not reported as a continuous variable
- Performance characteristics, including sensitivity, specificity and accuracy.

In addition to the information indicated above, the general background and description of the assay/imaging test should include:

- statement indicating whether the test is quantitative (includes degrees of positivity), semi-quantitative, or qualitative (positive vs. negative); what platform will be used; what is to be measured; controls; scoring procedures, including the values that will be used (e.g., pos vs. neg; 1+, 2+ 3+); interpretation; etc.
- specimen type(s) and preparation/handling/shipping SOPs, including definition for acceptability of a sample; for imaging tests, detailed SOPs regarding image acquisition, analysis and interpretation
- information about the statistical design used to establish the correlation with the clinical parameter of interest
- procedures to be used when results are not interpretable or are discrepant; this is especially important for assays or imaging tests used for eligibility or assignment to treatment arms

Further explanation of the information required for each of the categories in the first set of bullets above is presented in the appendix to this document. The requested information should be submitted as part of the correlative study section of the trial concept and protocol documents. ***N.B. The appendix refers to information required in protocols that include an integral biomarker assay or imaging study.***

II. Integrated assays include assays that will be performed on all samples or cases (for imaging studies) but are not required for the trial to proceed and will not inform treatment decisions or actions within the current trial. The requirements and the information to support the use of the assay are as follows:

- The assay should be well characterized, and as much information about the analytic performance as possible should be provided. Since the expectation is that integrated tests will often be used to inform the next set of trial hypotheses and will use precious specimens and/or significant patient participation and trial resources, the assay must have been adequately studied and shown to perform well using appropriate clinical specimens.
- Specific information required overlaps considerably with that for the integral assays:
 - Clear statistical design to assure that the marker/assay hypothesis aligns with the underlying trial and the sample size is sufficient to provide results to support potential use as an integral assay in a new trial

- Description of whether the test is quantitative (includes degrees of positivity), semi-quantitative, or qualitative (positive vs. negative); what platform will be used; what is to be measured; controls; scoring procedures, including the values that will be used (e.g., pos vs. neg; 1+, 2+ 3+); interpretation; etc.
- Definition of the specimen type(s) and preparation/handling/shipping SOPs, including definition for acceptability of a sample; for imaging tests, detailed SOPs regarding image acquisition, analysis and interpretation
- Explanation of the statistical design used in previous studies to establish the correlation with the clinical parameter of interest during development of the assay; available data on clinical performance and/or clinical utility
- For in vitro assays, information about the laboratory(ies) performing the assay – is it CLIA certified? Will the assay be performed with controls that would meet CLIA requirements? If not, what controls will be used to ensure reproducibility and facilitate transferability?

APPENDIX

Explanation of information required

- Role the assay will play in the trial ii
- Measurements of precision and reproducibility iii
- Data to support proposed cut-point(s) iv
- Analytic specificity and analytic sensitivity v
- Accuracy measurements vii

The role the assay will play in the trial

Definition of item and rationale for requesting this information:

The role the assay will serve in the clinical trial should be clearly defined. Examples of trial-specific roles would include eligibility determination, assignment to therapy, or risk stratification for randomization to ensure a balance of marker-defined subgroups within each treatment arm.

The intended use of the assay/test in clinical practice may differ from its role in the trial. The intended use for which the assay is being evaluated in the trial should be described. Examples of intended clinical uses would include prognostic indicator, predictive variable for benefit from a particular treatment or class of therapeutic agents, or indicator of favorable response or toxic reaction to a specific drug.

Use of an assay in a trial-specific role requires that the assay has already been shown to provide information that will help to conduct the trial in a more efficient, safer, or focused way. If the clinical trial will evaluate the assay for clinical use, the specifics of the clinical decision the assay is intended to guide must be clearly defined, including patient population to which it is applicable and benchmark performance characteristics that are desired. This is particularly important if there is intent to use the data collected on the assay's clinical performance for submission to the FDA for pre-market approval or clearance of the assay.

Information to be provided

This section of the document should provide the background of the marker/assay and present the data that support the use of the assay for the defined role. For imaging studies, the performance of the imaging method should have been well characterized. Issues such as the test-retest variance or reproducibility as well as the accuracy of the imaging test should be established. The history of clinical studies using the assay should be presented. This is NOT the place to present the assay performance characteristics since these are to be presented separately.

Issues to consider

This section should be very carefully focused on the supporting data to demonstrate that the assay is fit for the trial-specific role that it will play or to show that background data are sufficiently strong to support further evaluation of the assay for a clinical use. The data should be described in sufficient detail to demonstrate relevance to the context of the trial (e.g., patient characteristics, clinical stages, treatments, etc.). If such data do not exist, then an explanation should be provided for why this assay was chosen.

This section should be no longer than two pages (not including bibliography).

Measurements of Precision and Reproducibility

Definition of item and rationale for requesting this information

Precision and reproducibility address closeness of agreement between independent test results obtained under stipulated conditions. The precision of an assay procedure refers to repeatability of measurements under essentially unchanged assay conditions, often referred to as "within-series precision" or "within-run precision." For imaging tests, "within-patient test-retest" reproducibility data would be relevant. Intermediate precision refers to measurements taken when there is variation in one or more factors, such as time, calibration, operator, and equipment - usually within a laboratory. Reproducibility generally refers to inter-laboratory precision and relates to changes in conditions such as different operators and measuring systems (including different calibrations and reagent batches). Independent test results refer to results obtained in a manner that is not influenced by previous results obtained on the same or similar test samples. Information about precision and reproducibility is critical to the ability to discriminate noise from biological meaning. Information on expected variation in assay procedures that might impact measurement results is critical.

Information to be provided

Information to be provided should include the protocol followed, the conditions of the study, what factors were varied, and summary metrics including calculations of standard deviation (SD), coefficient of variation (CV) and descriptions of relationships between variation measures and means. Precision studies will optimally be performed in ranges of assay values corresponding to important clinical decision points (e.g., near a cut-point that separates different clinical states). Imaging tests, whenever possible, should have quantitative or semi-quantitative analyses.

Most of the information to be included in this section can be in the form of tables accompanied by short descriptions of the inter- and intra-laboratory tests performed.

Issues to consider

There is well defined guidance for the study of precision in quantitative tests (CLSI Evaluation of Precision Performance of Quantitative Measurement Methods EP 5-A2). Precision studies in qualitative tests have been less well defined but can also be characterized using repeat testing and percent agreement. There are no analogous guidances for imaging tests at present, but increasingly there are publications giving benchmark reproducibility data for various imaging modalities. Investigators proposing to use an imaging test as an assay in a clinical therapy trial, as described in this document, should provide data showing how their implementation of the imaging test compares with published benchmark reproducibility data.

Data to support proposed cut-point(s) if assay results are not reported as a continuous variable

Definition of item and rationale for requesting this information

Cut-points are thresholds that are applied to continuous or semi-quantitative assay measurements for purposes of reducing the assay or imaging test result to a positive/negative determination or perhaps to a few categories (e.g., low, medium, high). Any cut-point(s) must be clearly pre-specified because the statistical strength of the association between the categorized marker and a clinical endpoint, and the clinical interpretation of the assay result, may vary depending on the particular cut-point(s) used.

Information to be provided

The cut-points to be applied to assay measurements, the rationale and the background data for the selection as it relates to the intended clinical use must be provided. In the case of a continuous marker that will be used to predict a binary outcome (e.g., treatment response or toxicity), cut-point rationale might be based on ROC analysis aimed at achieving a desired level of sensitivity or specificity. For time-to-event endpoints cut-points might be selected to achieve a specified separation of survival curves. The background information should include the sample sizes of any previous studies, a comparison of the characteristics of the previously studied patients and specimens to those that will be examined in the proposed study, and a brief explanation of how the cut-points were selected in those studies.

This section should be no more than one page.

Issues to consider

Frequently, cut-points are applied to assay or imaging test measurements for convenience of analysis without careful thought as to why a particular cut-point is appropriate or whether it is appropriate to apply a cut-point at all. For example, there might not be a strong rationale for applying a certain cut-point if the relationship between the assay measurement and clinical endpoint represents a biological continuum. Particularly problematic is the practice of cut-point optimization, i.e., choosing a cut-point for a continuous or semi-quantitative measurement to maximize the degree of statistical significance (e.g., minimize the p-value) of the difference between the clinical outcomes in the two resulting marker-defined groups. Not only does this method overestimate the true magnitude of difference in outcome between the two marker-defined groups, but it disregards the relative costs of misclassifying patients in either direction between the two groups. In general, choosing cut-points based on observed data can lead to biased results, and operating characteristics of the cut-point (e.g., sensitivity, specificity, predictive values) should be demonstrated on data sets independent of the ones used to derive them.

Analytic sensitivity, limits of detection and quantitation, and analytic specificity

Definition of item and rationale for requesting this information

Analytic sensitivity is the ability of a test to detect an analyte or entity when it is present. When the output of a test is binary, sensitivity traditionally refers to the proportion of positive test results obtained on cases that are truly positive for the entity or analyte of interest. For tests with quantitative output, the sensitivity refers to the change in the test output relative to the change in the actual amount of analyte, and this relation may depend on the absolute amount of analyte present.

The limit of detection is defined as the smallest amount of analyte that an analytical method can detect with a specified probability. A related term is limit of quantitation, the smallest amount of an analyte in a sample that can be quantitatively determined with acceptable precision, and trueness as measured by bias.

Analytic specificity is the ability of a test or procedure to correctly indicate absence of an analyte or entity when it is truly absent or to accurately quantify an entity or analyte in the presence of interfering or cross-reacting substances. Almost all assays demonstrate potential for false positive results due to interfering substances. (Sensitivity may also be affected by interfering substances.) When the output of a test is binary, specificity traditionally refers to the proportion of negative test results obtained on cases that truly do not possess the entity or analyte of interest.

Information to be provided

Information to be provided about the design of the sensitivity and specificity studies that were performed should include characteristics of the samples and positive and negative controls, the rationale for interfering substances studied, analyte or entity (e.g., tumor cells harboring a particular mutation) spike-in amounts and matrices used in any dilution experiments. Summary results such as sensitivity and specificity rates over the range of test samples considered, and calibration or dilution curves should be presented, as appropriate. For imaging tests, information should be provided about the populations studied and design of clinical trials used to determine the sensitivity and specificity of the imaging method.

Issues to consider

The sensitivity and specificity will be influenced by the imprecision of the measurements due to technical factors, and therefore it is important that the same assay technical protocols and same level of replication as will occur in the clinical setting be used in the analytical performance studies. The identification of which interfering substances to evaluate will depend on what is known about the analyte being measured and the matrix in which it is being measured, what is known from the literature about interfering substances, and what can be gleaned from false positive or false negative results in studies in the intended use population or in cross sectional studies of patients with conditions likely to produce cross reactivity. These substances may be common metabolic analytes that alter the underlying test principles (hemolysis, bilirubin, lipids);

cross-reacting analytes (antigens or antibodies) that are mis-identified as analyte by the detection system; materials that interfere with test mechanics such as human anti-mouse antibody (HAMA) or heterophile interference; pharmacologic or physiologic factors that cause measurement errors; or closely linked protein or nucleic acid targets which are inadvertently picked up as positive signals by the measurement system being used. Interference testing should be performed in ranges of analyte measurements corresponding to important clinical decision cut points using levels of interfering substances that would be expected to occur in the intended use population. Determination of limit of detection and limit of quantitation are particularly critical for laboratory determinations in which small amounts of an analyte are of importance in diagnosis of disease states associated with that analyte. Although intended for standard chemistry analytes, many of the principles apply to other forms of testing.

1) FDA International Conference on Harmonization; Guideline on Validation of Analytical Procedures: Definitions and Terminology; Availability. Docket No. 94D-0016, March 1995.

2) CLSI Protocols for Determination of Limits of Detection and Limits of Quantitation; Approved Guideline (EP17- A).

3) CLSI voluntary standard: Interference Testing in Clinical Chemistry (EP7-A2). Although intended for standard chemistry analytes, many of the principles apply to other forms of testing.

Accuracy Measurements

Definition of item and rationale for requesting this information

Accuracy is defined as the closeness of agreement between the test results obtained using the new biomarker test and results obtained using a reference standard method widely accepted as producing “truth” for the analyte. For example, a reference method considered standard for detection of DNA mutations is sequencing. The observed level of agreement will depend on both the bias and precision of the new test. Bias is the amount by which an average of many repeated measurements made using the new test systematically over- or underestimates the reference standard method result. Precision is discussed separately. For many new biomarker tests, there will not be a universally accepted reference standard method.

Information to be provided

The reference standard method, if any exists, should be clearly stated. Accuracy measures such as overall percent agreement, and sensitivity and specificity relative to the reference standard results should be reported for tests that yield binary results. For continuous marker values, accuracy measures such as average bias, mean absolute deviation or mean squared deviation should be reported over the relevant range of true (reference standard method) values.

Issues to consider

For situations in which there is no universally accepted reference standard method, it may still be helpful to compare the new biomarker test to a non-reference standard test for which results are expected to show some correlation with the new test results. For example, results of a new immunohistochemical test using a novel combination of antibodies to assess protein expression might be expected to show some correlation with results obtained using an older test with a single antibody, although some differences would be expected as well. Presentations of comparisons to a non-reference standard method should be accompanied by a discussion of reasons why some differences would be expected. For many imaging tests, accuracy measurements can usually only be made with respect to an inanimate phantom (test object). Data regarding accuracy and precision measurements of a relevant phantom should be provided. In some cases, previous studies may have been done using histologic markers to demonstrate “accuracy” of an imaging method (for example, comparing DCE-MRI results to microvessel density in biopsied tissue as a “truth” marker for angiogenesis).

When there are no reference or non-reference standard methods for measuring the biomarker, sometimes an assessment of true clinical state or condition will serve as the reference. For example, if a new test is developed to predict a toxic reaction to a drug, the results of that test could be compared to the outcome or clinical manifestation of an adverse event. If this approach is taken, then the new biomarker test results must not be used to make the clinical assessment. A biomarker test used in this way would be an example of an integrated assay because the test is performed on all patients but no action is taken on the results directly. If the assay result was shown to predict the adverse

outcome, then in a subsequent trial the assay might be used for clinical decision making (and the biomarker test would be considered an integral assay). For this example in which clinical assessment is used as the reference, evaluation of analytical accuracy is bypassed and clinical accuracy is evaluated directly.