

Performance Studies

Review of Information Provided to a
Regulatory Authority

Elliot P. Cowan, PhD

Partners in Diagnostics, LLC



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Purpose of this Talk

To understand:

- The purpose of analytical and clinical performance studies
- The different types of analytical and clinical performance studies and their specific roles
- Expectations for the conduct and reporting of analytical and clinical performance studies



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Introduction to Performance Studies



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Two Broad Categories of Studies



- Analytical performance studies
 - The ability of an IVD medical device to detect or measure a particular analyte.
- Clinical performance studies
 - The ability of a medical device to achieve its intended purpose as claimed by the manufacturer.

GHTF/SC/N4:2012 (Edition 2)

Two Broad Categories of Studies, Simplified



- Analytical Performance
 - What is this product capable of doing under ideal and controlled conditions?
- Clinical Performance
 - How does the product actually perform when used as intended?

Performance Studies Must be Well-Thought Out and Well-Organized



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



For All Performance Studies (1)

- Study description, study identifier, product identifier (for example, lot numbers), IFU version used, date of initiation and date of completion
- A summary of the study findings, including a conclusion that clarifies how the study objectives have been met



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



For All Performance Studies (2)

- **Study protocol and full report** containing the following information:
 - Study objectives, study design, the methodology used, and data collected
 - Site of study (e.g., manufacturer's R&D laboratory, hospital laboratory, health care clinic)
 - Operator of the assay (e.g., lab professional, non-lab, untrained)
 - Reference standard, if applicable
 - Specimen acceptance criteria, specimen characterization



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



For All Performance Studies (3)

- Study protocol and full report (cont.)
 - Specimen type and numbers of each type
 - Details of statistical methods, estimations and calculations applied
 - When performed by a party other than manufacturer, details of this party and relationship to manufacturer
 - Report only:
 - Actual test result summaries with their acceptance criteria and not just pass/fail statements
 - Clearly labeled data that are clearly linked to the study report
 - Study conclusion



General Overview of Conducting Performance Studies

1. **Identify** what studies need to be done
2. **Define** the study objective(s)
3. **Design** the study to meet the study objective(s)
4. **Write** the study protocol
5. **Conduct** the study
6. **Prepare** the study report
7. **Review** the report to ensure that the study protocol was followed and that the conclusions are appropriate



Analytical Performance Studies



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



What are analytical performance studies?



The ability of an IVD medical device to detect or measure a particular analyte.

GHTF/SC/N4:2012 (Edition 2)

or:



What this product is capable of doing under ideal and controlled conditions.

Analytical Performance Studies (1)

Specimen Type

Identify the different specimen types (matrices) that can be used with the IVD and storage and shipping conditions



Analytical Performance Studies (2)

Accuracy of Measurement

- Trueness (for quantitative tests)
 - Closeness of agreement between test result and the true value
- Precision (for all tests)
 - Closeness of agreement between independent test results
 - Repeatability: To establish the **variability of the test itself**
 - Reproducibility: To establish the **variability of the testing process**
- Calculate mean values of replicates and coefficients of variation



Analytical Performance Studies (3)

Analytical Sensitivity

Identify the inherent ability of the test to detect the analyte it was designed to detect compared to a reference assay (includes detection of variants)

Analytical Specificity

Identify the inherent ability of the test to detect only the analyte it was designed to detect, in the presence of other substances/agents in the specimen

Validation of assay cutoff

Determine the threshold above which the result is reported as positive and below which the result is reported as negative



Analytical Performance Studies (4)

Traceability

Identify how the values for calibrator and control values were determined and ensure that they are based on an appropriate standard

Measuring Range of the Assay

Define the range of analyte concentrations that can reliably be detected by the test

Validation of Assay Cutoff

Determine the threshold above which the result is reported as positive and below which the result is reported as negative

Reading time

- Lower limit: Time needed for chemical reaction to produce a test result
- Upper limit: Negative specimens may give false negative results (“over-development”)



Analytical Performance Studies (5)

Stability: Claimed Shelf Life

Upper limit of the time interval during which the performance characteristics of the product stored under specified conditions can be assured

Stability: In-Use Stability

Upper limit of the time interval during which the performance characteristics of the product stored under specified conditions can be assured, *after the material has been altered from its original received state*

Stability: Shipping Stability

Upper limit of the time interval during which the performance characteristics of the product *shipped under specified conditions* can be assured



Analytical Performance Studies (6)

Robustness (Operational Characteristics)

Determine the degree to which the product can tolerate:

- Operator error/human factors
- Specimen integrity and handling
- Reagent integrity
- Hardware, software, and electronics integrity
- Stability of calibrators and controls
- Environmental factors



Clinical Performance Studies



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



What is the purpose of a clinical performance study for an IVD?

To demonstrate, **with objective evidence**, how the IVD will be expected to perform in routine clinical practice (i.e., “the real world”)



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF

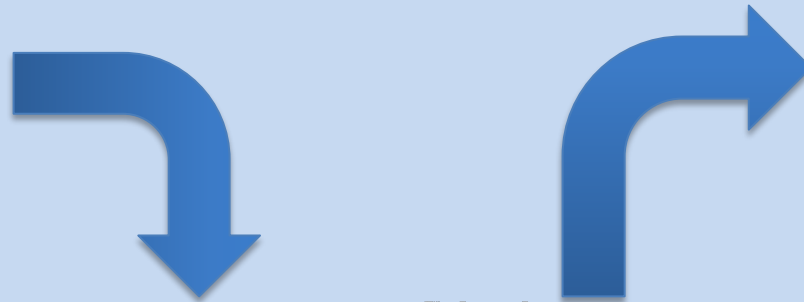


Breaking Down Clinical Performance Studies

INPUTS



- “Intendeds”
- Clinical performance study design
- Clinical performance study protocol

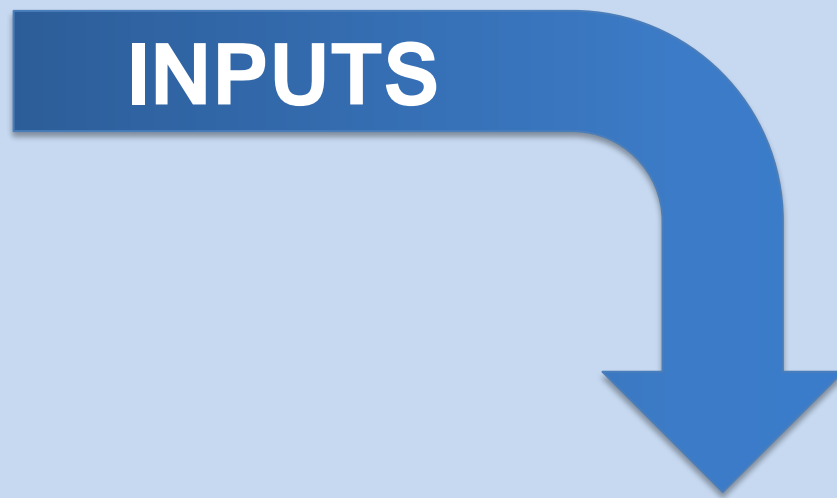


OUTPUTS



- Performance representative of the “real world”





INPUTS

INTENDED USE



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



The Important List of “Intendeds”

- Clinical performance studies should be conducted consistent with the intended use of the IVD

INTENDED:

Users

Use setting

Use population

Specimens

Test version



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Definition: Intended Use



The objective intent of the manufacturer regarding the use of a product, process or service as reflected in the specifications, instructions and information provided by the manufacturer.

(GHTF/SC/N4: 2011)



How the test is to be used: what condition, what sample, what patient

Includes indications for use: why an individual/patient would be tested

Indications for Use for IVDs

GHTF/SG5/N8:2012

- Diagnosis
- Aid to diagnosis
- Screening
- Monitoring
- Predisposition
- Prognosis
- Prediction of treatment
- Determination of physiological status



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Example of an Intended Use Statement

Intended use

X is a single-use immunochromatographic test for the detection of antibodies to Human Immunodeficiency Virus Type 1 (HIV-1) and Type 2 (HIV-2) in fingerstick whole blood, venous whole blood, serum or plasma specimens

as a point-of-care test to aid in the diagnosis of infection with HIV-1 and HIV-2. It is not intended to be used to screen donors of whole blood and plasma.

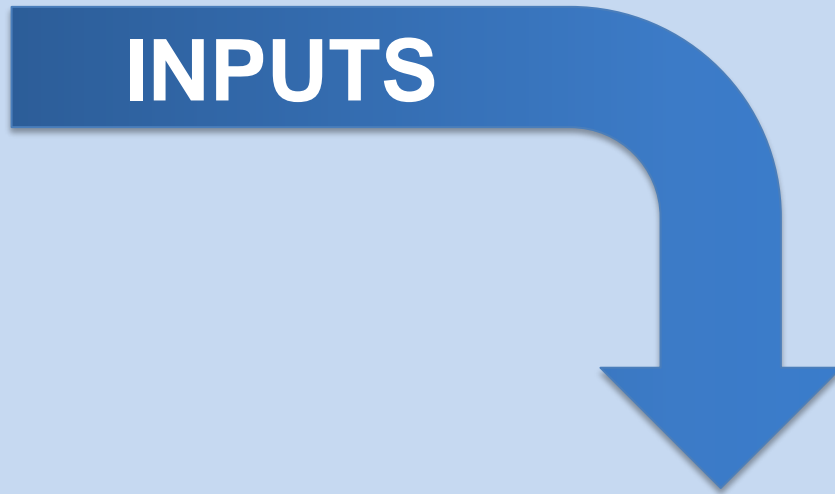
Indications for use



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



INPUTS



STUDY CONSIDERATIONS



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Elements of a Clinical Performance Study

- Clinical performance study design
- Clinical performance study protocol
- Informed consent/study participant protection
- Study monitor
- Study investigators



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Establishing Truth

- Purpose of the clinical performance study is to establish the performance characteristics of a new IVD
- There must be a way to determine the true status of a specimen



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Definition: Reference Standard (CLSI)

- The best available method for establishing the presence or absence of the condition or characteristic of interest
 - Can be a single test or method, or a combination of methods and techniques, including clinical follow-up
- The reference standard will evolve with the advancement of analytical systems



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Assessing the Selection of the Reference Standard

- Does the reference standard have a known and appropriate level of clinical performance with the same specimen types?
- Is the reference standard well-characterized with known state-of-the art performance characteristics that is manufactured under a stringent quality management system?



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



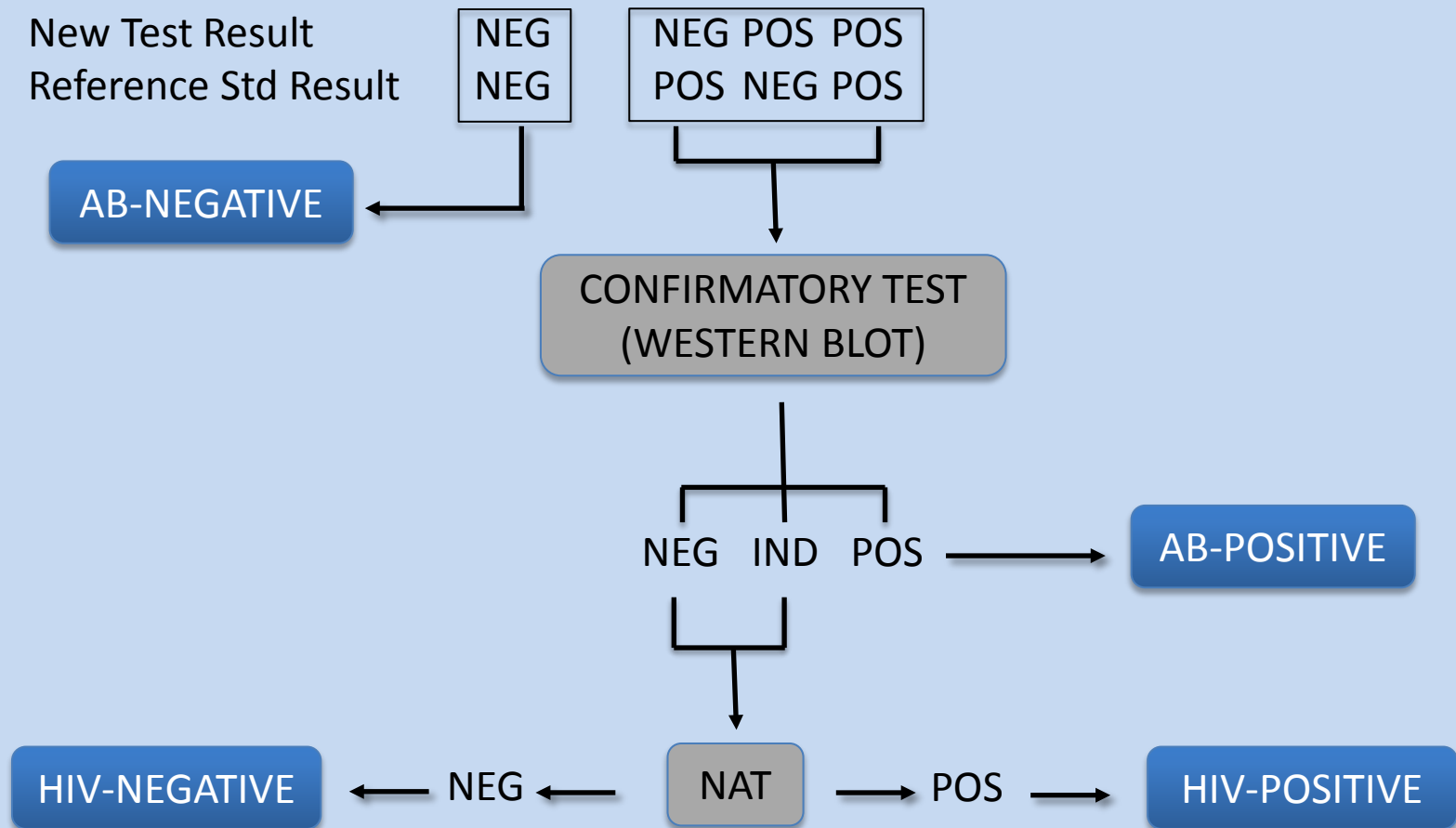
Use of the Reference Standard: The Testing Algorithm

- What is a testing algorithm?
 - A predetermined order of testing to establish the true status (positive or negative, correct quantitation, etc.) of a specimen
- Characteristics of testing algorithms
 - All specimens should be tested using both the test under investigation and the reference test
 - In the case of a binary result, negative results on both tests may be considered negative, any positive result should be tested further
 - Testing methods should go as far as possible to establish truth
- **For assessment: Is the testing algorithm adequate to establish truth?**



Example of a Testing Algorithm

Rapid HIV antibody test as an aid in the diagnosis of HIV infection



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Reference Standard Influences Performance Characteristics

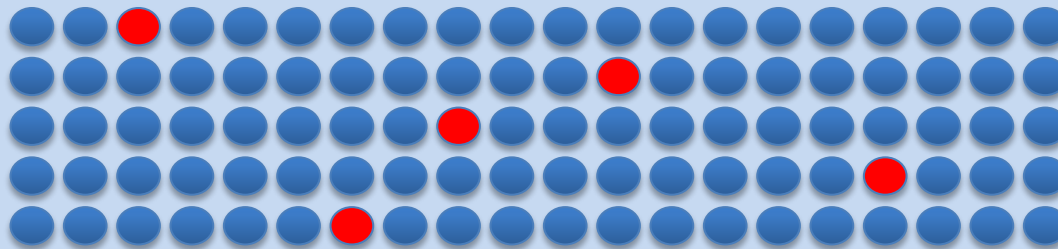
- In assessment, ensure that the performance of the reference standard is as high as possible
- The “Slippery Slope”: Clinical performance is only be as good as the reference standard
 - Negatives on both the test under investigation and the reference standard (concordant negatives) are not tested further
 - Done for practical reasons – most results negative in prospective clinical performance studies



Example of the “Slippery Slope”

100 true POS among a larger number tested

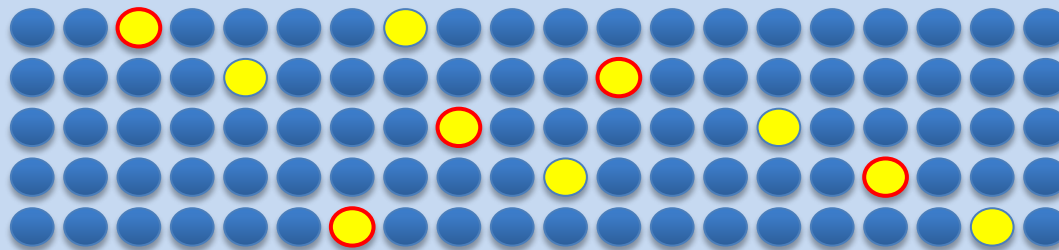
Reference Standard Test: 95% Sensitivity



● = undetected
By ref test;
no further
testing



Test X
(misses same specimens as Ref Std)



● = undetected
by both tests;
no further
testing



5 *detectable* false negative
Apparent Sensitivity: 95%

Actual Sensitivity:
 $95\% \times 95\% = 90\%$

What if there is no reference standard?

Options:

- Identify research test method that has been validated to an acceptable level
- Identify clinical evidence of infection (if the IVD is intended to detect infection)



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Statistical Design Issues (1)

- Statistical significance (power)
 - Appropriate sample size to estimate performance (e.g., sensitivity and specificity) with appropriate confidence intervals
- Appropriate subject and specimen inclusion and exclusion criteria



Statistical Issues (2)

- Criteria for resolution of discrepant results (testing algorithm to establish truth)
- Criteria for data exclusion
- Methods of analysis
- Clinically relevant performance measures
- Minimization of bias



Importance of Sample Size in Clinical Performance Studies

- Want to know how much confidence there is that the data from the clinical performance studies will be what the user can expect
- The generally accepted way to express data from studies is to identify the actual values (referred to as point estimates) along with 95% confidence intervals



95% Confidence Intervals

- Definition
 - There is a 95% likelihood that the values obtained by users will fall within the statistically determined range
- Basic concept:
 - The larger the number of samples tested, the smaller the confidence intervals around the point estimate (i.e., the more confidence there is in the point estimate, and less uncertainty)



Influence of Indications for Use on Sample Size (1)

- Screening indication
 - Most people tested will be negative for the analyte
 - Need to screen a large number of individuals to detect a positive
 - Supplement with known positives or individuals from a population with a higher prevalence of the analyte
 - Blinded to the user to avoid bias
 - More confidence in the specificity (expected false positive rate)



Influence of Indications for Use on Sample Size (2)

- Diagnostic indication
 - Testing of individuals with signs and symptoms
 - More limited numbers realistically possible, and larger confidence intervals may be acceptable (per risk analysis)
 - Supplement with known positives



Bias

- **Definition:**

A flaw in the study design or the method of collecting or interpreting information that can lead to incorrect conclusions about what the study or clinical trial showed.



Types of Bias in IVD Evaluation Studies

- Spectrum composition
- Workup bias
- Review bias (blinding)

Spectrum Composition Bias

- What it is:
 - Bias that results from the clinical performance studies not representing the intended use or intended user population
- How to assess studies for spectrum composition bias
 - Should contain information on 3 of the following 4 criteria:
 - Age distribution
 - Sex distribution
 - Summary of presenting clinical symptoms and/or disease stage
 - Eligibility criteria for study subjects and users
 - Pertinent subgroups
 - Sensitivity and specificity may represent average values for a population



Workup Bias

- What it is:
 - When patients with positive or negative test results preferentially receive verification of diagnosis by the reference procedure
- How to assess studies for workup bias
 - Ensure that all samples are tested using the reference test



Review Bias (Blinding)

- What it is:
 - Bias introduced by operators who are aware of a prior test result or the clinical status of the individual from whom the test specimen was obtained
- How to assess studies for review bias
 - Were investigational test and reference test interpreted separately by persons unaware of the results of the other?
 - Ensure that only blinded specimens were used for testing



Ethical Considerations (1)

- Rights, safety and well-being of subjects participating in a clinical performance study should be protected
 - Study should generate new data
 - Benefits to health must outweigh risks to study participants, and risks must be minimized
 - Confidentiality must be respected



Ethical Considerations (2)

- Informed consent
 - For specimens or personal data collected specifically for the study or for specimens/data that can be traced to an individual
- Ethics Committee
 - Required by some jurisdictions to review, approve, and monitor studies to protect human subject rights and welfare
- Communicating test results outside of the study
 - Mechanism to report results to physicians or public health authorities cleared by Ethics Committee?



Study Considerations: Specimen Collection and Handling

- Specimen source
 - Specimens taken from patients with the ***intention to use them in a particular clinical performance study***
 - Tested immediately (fresh) or stored (e.g., refrigerated or frozen) for later use
 - ***Leftover specimens*** collected for routine diagnostic testing that would otherwise be discarded, or specimens collected for research purposes
 - ***Archived specimens*** that were collected in the past and were stored in repositories
- Issues to consider
 - Specimen integrity (especially for leftover and archived)
 - Measures taken to avoid review bias



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Study Considerations:

Clinical Performance Study Site Location

- To simulate real world use, best to conduct clinical performance studies at sites other than the manufacturer (for multiple sites, manufacturer may be one)
 - Intended user, intended use environment
- If testing is only done at the manufacturer, then justification is needed
- Procedures for specimen collection and/or testing at the site must minimize bias



INPUTS

STUDY PROTOCOL



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Definition: Clinical Performance Study Protocol



Document that states the rationale, objectives, design and proposed analysis, methodology, monitoring, conduct and record-keeping of the clinical performance study.

GHTF/SG5/N8:2012



How the study is intended to be conducted.

Reviewing the Study Protocol (1)

- Completeness
 - Is there a study protocol?
 - Are all of the necessary elements present in the study protocol?
- Rationale and objectives
 - Are the rationale and objectives appropriate to conduct a study that will establish the performance of the IVD?
- Is study monitored and are clinical investigators appropriate?



Reviewing the Study Protocol (2)

- Did the study design reflect the diversity of real-world use of the product?
 - Intended users
 - Level of training?
 - Intended use setting
 - Laboratory or point-of-care?
 - Under conditions in which test expected to be used?
 - Intended use population
 - Reflects diversity of individuals who would be tested?
 - Intended specimens
 - Specimens tested are those for which a testing claim is sought?



Reviewing the Study Protocol (3)

- Is sampling based on sound statistical principles and methodology?
- Were the statistical methods used for data analysis (confidence interval calculations, significance calculations, etc.) appropriate?
- How did the study determine truth (what the correct test result should be)?
 - Appropriate reference standard?
 - Testing algorithm to establish clinical status of patient
- Were controls in place to minimize bias?
 - Blinding users to previous test results or clinical status of individual being tested



Reviewing the Study Protocol (4)

- Was informed consent needed?
- If needed, was the informed consent adequate to inform the study participant of:
 - Nature and objectives of the study
 - The individual's rights by participating in the study
 - What will be done with the specimen
 - Risks to the individual as a result of participating
 - Contact information for questions or concerns



Note on Test Version

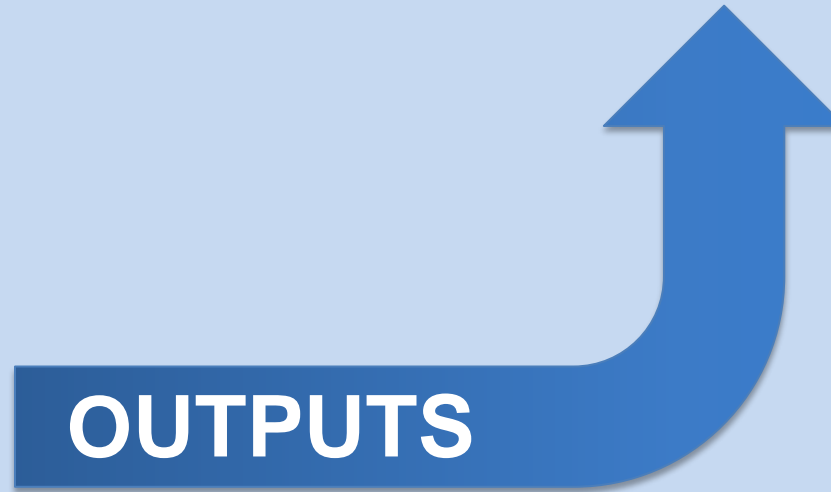
- The version of the test used in the clinical performance studies should be one that is intended to be marketed.
- No changes should be made during the clinical performance studies.
- All sites must use the same version of the test.
- If any changes are made after conducting the clinical performance studies, those should be validated to show that they do not impact safety and performance.



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



PARAMETERS



OUTPUTS



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



What are the expected outputs of a clinical performance study?

- Results from ***clinical evaluation*** of the IVD according to the study protocol to
 - Generate ***clinical data***
 - Which serve as ***clinical evidence***
 - To support the ***clinical performance*** of the IVD



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Outputs of Clinical Performance Studies

- Clinical performance may be expressed in several ways
 - Sensitivity
 - Specificity
 - Positive predictive value (PPV)
 - Negative predictive value (NPV)
 - Invalid test rate
 - Accuracy



Performance Measures

(in the context of an HIV test)

Sensitivity

- % who are infected and have a positive test result
- Ability of the test to **correctly identify an infected individual**

Specificity

- % who are not infected and have a negative test result
- Ability of the test to **correctly identify an uninfected individual**

PPV

- % with a positive test result and who are infected
- Measures the **reliability of a positive test result**

NPV

- % with a negative test result and who are not infected
- Measures the **reliability of a negative test result**



Sensitivity and Specificity vs. Positive and Negative Predictive Values

- Sensitivity and specificity are inherent characteristics of a test
- Positive and negative predictive values vary by prevalence in the population



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Example: HIV

For a test that is 99% sensitive and 99% specific, and a sample of 10,000 individuals:

- If the prevalence of the analyte is 10%:

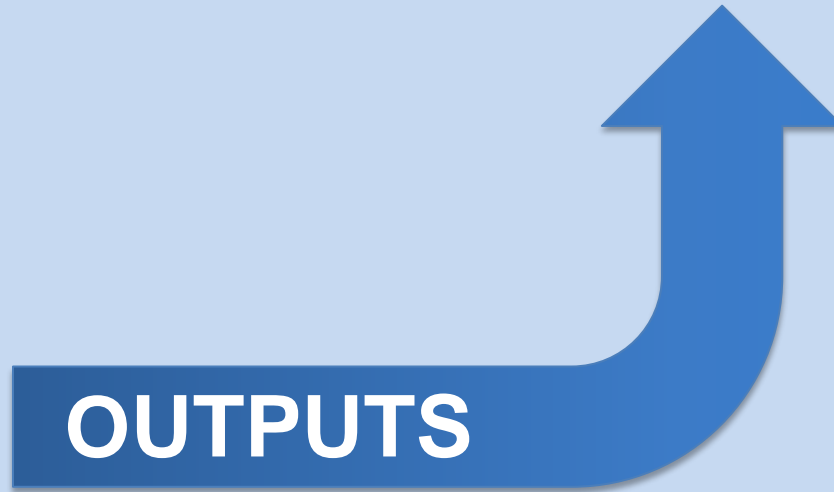
$$PPV = 990 / (990 + 100) = 91\%$$

- If the prevalence of the analyte is 1%:

$$PPV = 99 / (99 + 100) = 50\%$$



QUALITATIVE TEST RESULTS



OUTPUTS



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Binary Test Results

- Results are Yes/No, Reactive/Non-reactive, Positive/Negative
 - Enzyme immunoassays: Signal to cutoff ratios
 - Rapid Diagnostics: Presence or absence of a line or spot
 - Early Infant Diagnosis test: Above or below a cutoff value
 - Sometimes referred to as semi-quantitative



Performance Measures Associated with Binary Test Results

- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value
- Invalid test rate



Cutoff Values: Additional Comment

- The manufacturer may modify the cutoff after the clinical performance study and re-analyze the data to optimize the sensitivity and specificity
- Before regulatory approval!



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Reviewing Binary Data (1)

- Are the numbers correct?
 - Numbers in summary tables and text
 - Calculations
 - Confidence intervals and results of other statistical methods
- Are all data taken into account?
 - Ensure that there is a justification for any data point not included in calculations

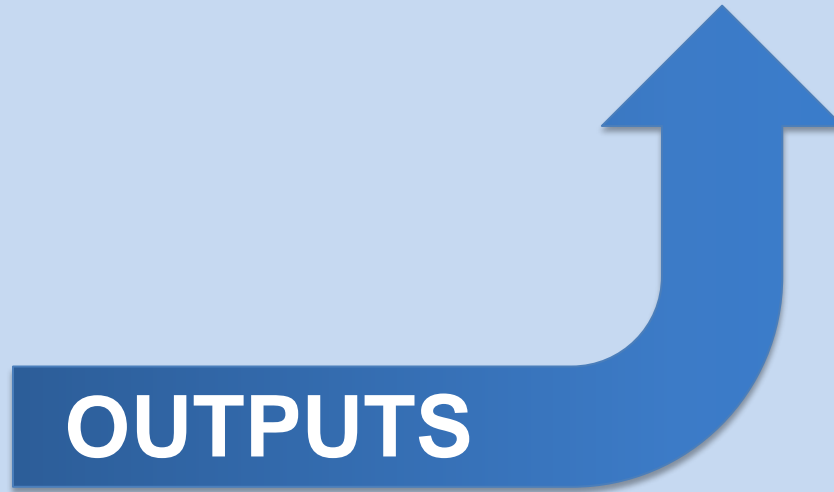


Reviewing Binary Data (2)

- Is there an excessive number of invalid test results?
 - Source of invalids (test, operator)?
- Are the performance numbers (including confidence intervals, results of statistical analysis, etc.) acceptable?
 - Do the benefits outweigh the risks?



QUANTITATIVE TEST RESULTS



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF



Quantitative Test Results

- Numerical readout that is interpreted
- Performance measures:
 - Accuracy
 - How much difference is there in the quantitative result produced by the test method compared to the reference test?
 - Where there is a difference, is it consistent?
 - Methods:
 - Linear regression
 - Difference plots
 - Invalid test rate



A Note on Comparing Quantitative Nucleic Acid Tests

- Readout of nucleic acid tests is copies of RNA/mL
- “Copy number” may vary significantly from test to test using the same sample
- To compare results in this case, establish the number of copies per unit of RNA when such a standard exists
 - Testing of standards (that define units) and convert copies to units to normalize the test results



Reviewing Quantitative Data (1)

- Are the numbers correct?
 - Numbers in summary tables and text
 - Calculations
 - Confidence intervals and results of other statistical methods
- Are all data taken into account?
 - Ensure that there is a justification for any data point not included in calculations



Reviewing Quantitative Data (2)

- Are there an excessive number of invalid test results?
 - Source of invalids? (e.g., test, operator?)
- Are the performance numbers (including confidence intervals, results of statistical analysis, etc.) acceptable?
- Do the benefits outweigh the risks?
 - Consider the implications of incorrect and invalid test results



Thank You



U.S. FOOD AND DRUG ADMINISTRATION
U.S. PRESIDENT'S EMERGENCY PLAN FOR AIDS RELIEF

