Analytical and Clinical Studies Design Considerations for IVD Devices

2016 Pre-Submissions Workshop

Marina V. Kondratovich, Ph.D.
Associate Director for Clinical Studies
OIR, CDRH, FDA

April 20, 2016

Outline

I. Introduction

II. Precision

III. Clinical performance characteristics:
Risks, absolute risks, relative risks;
Likelihood ratios (LR).

IV. Potential Biases in Clinical Study

What Type are Device Outputs

How results of the device are reported to a physician?

Qualitative test: nominal type of outputs

Nominal

- Nominal refers to data such as names/categories. For example, five different genotypes. May have numbers assigned, not for arithmetic purpose.
- Easy to remember because nominal sounds like name.

Qualitative test: with 2 outputs (negative, positive)
with multiple outcomes (e.g. genotyping)

What Type are Device Outputs

How results of the device are reported to a physician?

Quantitative test: The amount or concentration of an analyte is measured and expressed as a numerical quantity value in measurements units.

Nominal

- Values that can be subtracted and can be divided:
  Total PSA: values 50, 100, 150 (units)

Linearity of the device should be evaluated.

What Type are Device Outputs

How results of the device are reported to a physician?

Semi-Quantitative test

Examples:

Ordinal

- Ordinal refers to quantities that have an ordering – order matters but not the difference between values. For example, urine dipstick with outputs: neg, trace, 1, 2, 3.
- Easy to remember because ordinal sounds like order.

Score (index) devices

Medical Laboratory Test

Metrological performance (measuring device)

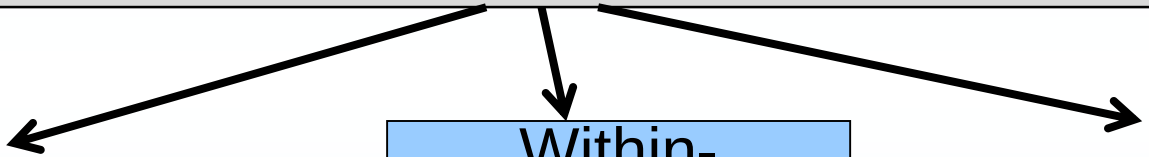
Clinical performance (related to the claim)

CLSI documents are major sources of terminology, study design, and statistical analysis

Precision Studies

Precision

closeness of agreement between … measured quantity values obtained by replicate measurements on the same .. objects under specified conditions.
NOTE: The 'specified conditions' can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement.



Repeatability

same lab,
same lot,
same operator,
same day,
same run

Within-Laboratory Precision

same lab,
different operators,
different days,
different runs,
different lots

Reproducibility

different labs,
different operators,
different days,
different runs

Precision Studies

Example of reproducibility study

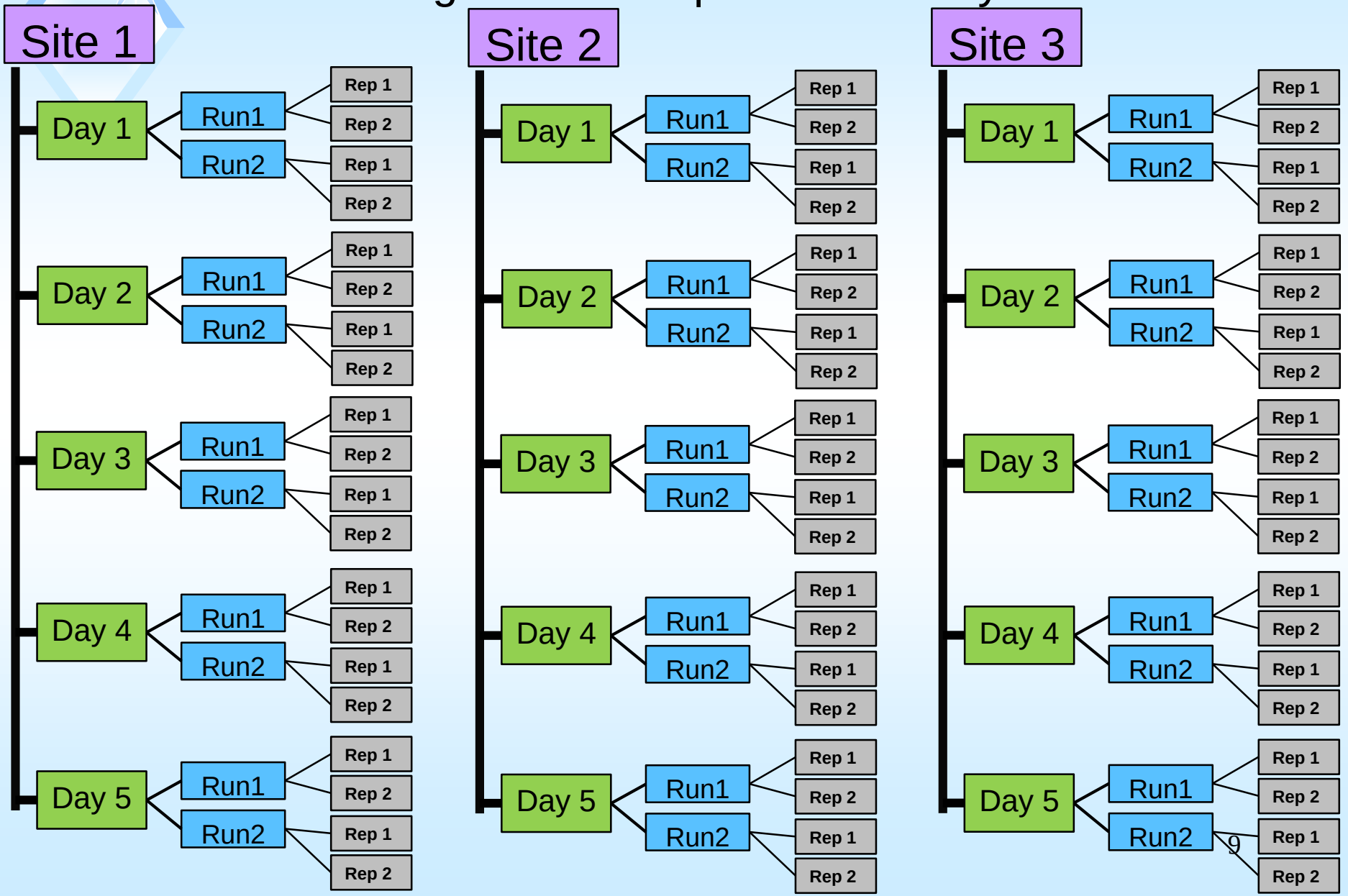
- 3 sites (1 internal + 2 external)
- 5 days per site
- 2 runs per day
- 2 replicate per run

❑Source of variability “operator-to-operator”

2 operators? Balanced design

Precision Studies

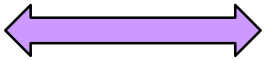
- Provide a diagram for the precision study



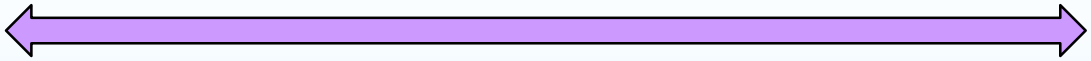
Precision Studies

For analysis of the data, use CLSI EP05-A3.

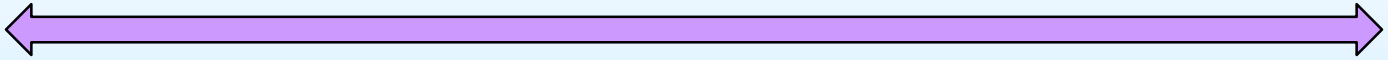
Reproducibility												
Mean	Repeatability (within-run)		Between-run		Between-day		Between-operator		Between-site		**Total**	
	SD	%CV	SD	%CV	SD	%CV	SD	%CV	SD	%CV	SD	**%CV**
….												



Repeatability



Within-Lab Precision



Reproducibility

Precision Studies

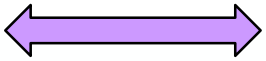
- **Source of variability "lot-to-lot"**

  Different study designs:

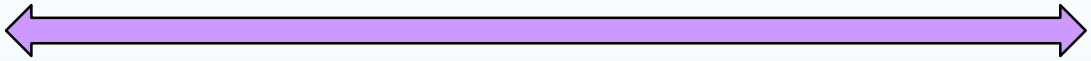
A) 3 sites
   - <u>each site has 3 lots</u>
   - 5 days
   - each day
     - 2 runs with Lot1,
     - 2 runs with Lot2,
     - 2 runs with Lot3
   - each run has 2 replicates

Precision Studies

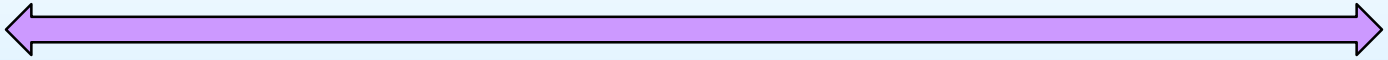
Reproducibility												
Mean	Repeatability (within-run)		Between-run		Between-day		Between-lot		Between-site		**Total**	
	SD	%CV	SD	%CV	SD	%CV	SD	%CV	SD	%CV	SD	**%CV**
….												



Repeatability



Within-Lab Precision



Reproducibility

Precision Studies

- **Source of variability "lot-to-lot"**

  Different study designs:

  B) Two precision studies

  Study 1

  Evaluation of lot-to-lot precision at 1 site (usually internal)

  Study 2

  Reproducibility

  3 sites but each site has the same lot

Precision Studies

Composition of precision panel:

Concentrations of samples

- qualitative test with two outputs (pos, neg)
- quantitative test

Nature of samples

- if is preferable to use patient samples
- you should work with FDA to define acceptable sample types
- include QC samples

Discuss important issues of composition of precision panel through pre-Sub process

For qualitative tests with available continuous signal, in addition to analysis of signal, present percent positives for each sample.

Clinical Studies

Guidance for Industry, Clinical Investigators, Institutional Review Boards and Food and Drug Administration Staff –

Design Considerations for Pivotal Clinical Investigations for Medical Devices (2013)

The web address

http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm373750.htm

Section 8, pages 38-46

Diagnostic Clinical Studies

Diagnostic Clinical Outcome Studies

Example:
one group of patients uses the Candidate test;
second group of patients uses the Old test;
Clinical outcome (for example, HBA1c)
comparison of clinical outcomes in both groups

Diagnostic Clinical Performance Studies

Example:
Qualitative test with two outputs
Sensitivity, specificity, risks (predictive values), likelihood ratios

Clinical Studies

Typical scheme

N subjects in the clinical study (N subjects from target population)

Every subject

Candidate Test:	Assessment of Target Condition (ATC) (Gold Standard):
Positive, Negative	D+ = Target condition present, D- =Target condition absent

Consider Test with Two Outputs (Pos, Neg)

Let us have 1,300 subjects who are representative subjects from intended use population (target population). Each subject has results of the Test (Pos, Neg) and ("Gold Standard") (D+, D-).

		Colposcopy		
		D+	**D-**	Total
T	**Pos**	66	694	760
	Neg	4	536	540
Total		70	1,230	1,300

Prevalence of 5.4% (70/1,300) reflects prevalence in the IU population.

Clinical Performance of the Test	
Sensitivity	94.3% (66/70)
Specificity	43.6% (536/1,230)

Risks (Absolute Risks)

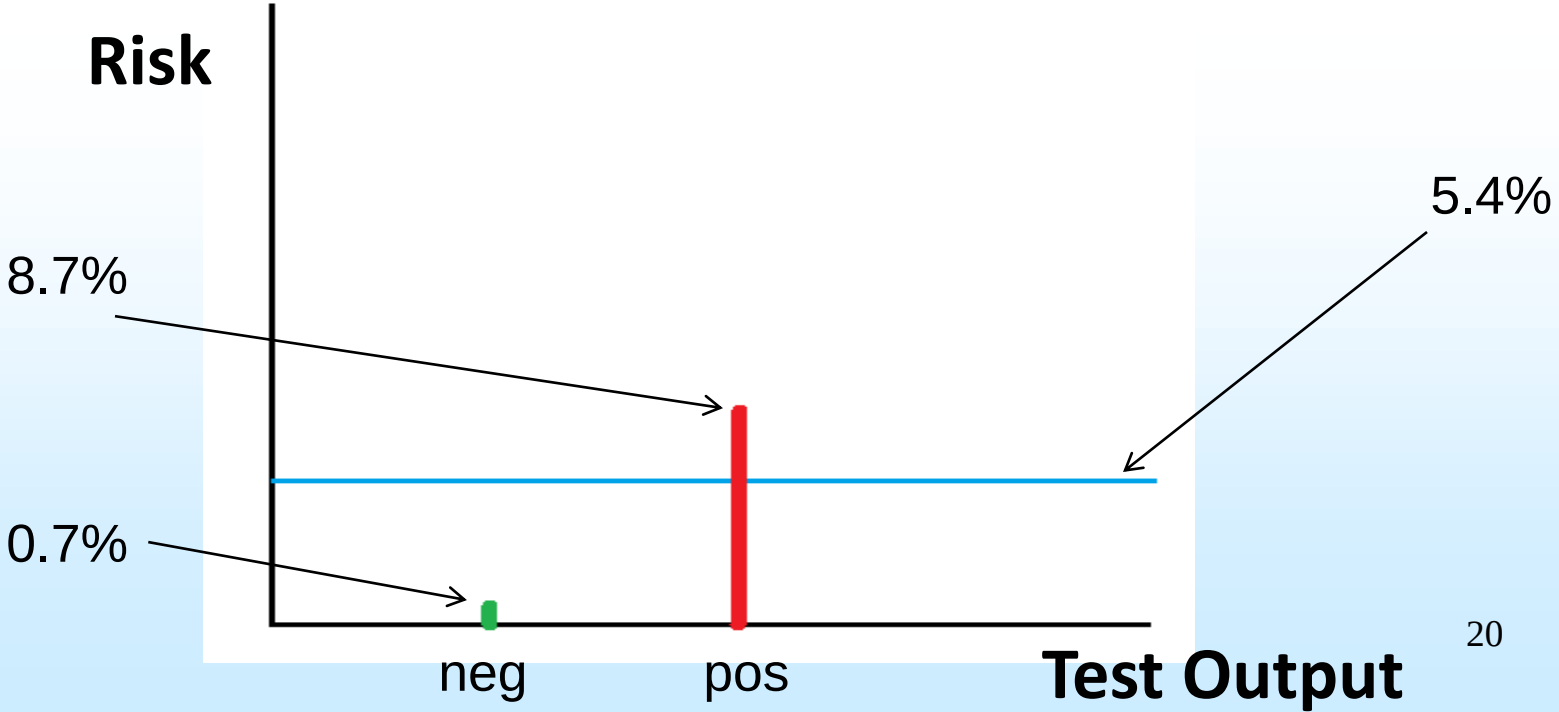
		D+	D-	Total
T	Pos	66	694	760
	Neg	4	536	540
Total		70	1,230	1,300

Clinical Performance of the Test	
R(Pos)=Risk of D+ for T Pos (PPV)*	8.7% (66/760)
R(Neg)=Risk of D+ for T Neg (1-NPV)*	0.7% (4/540)
π = Pre-test risk of D+ (baseline risk, prevalence)	5.4% (70/1,300)

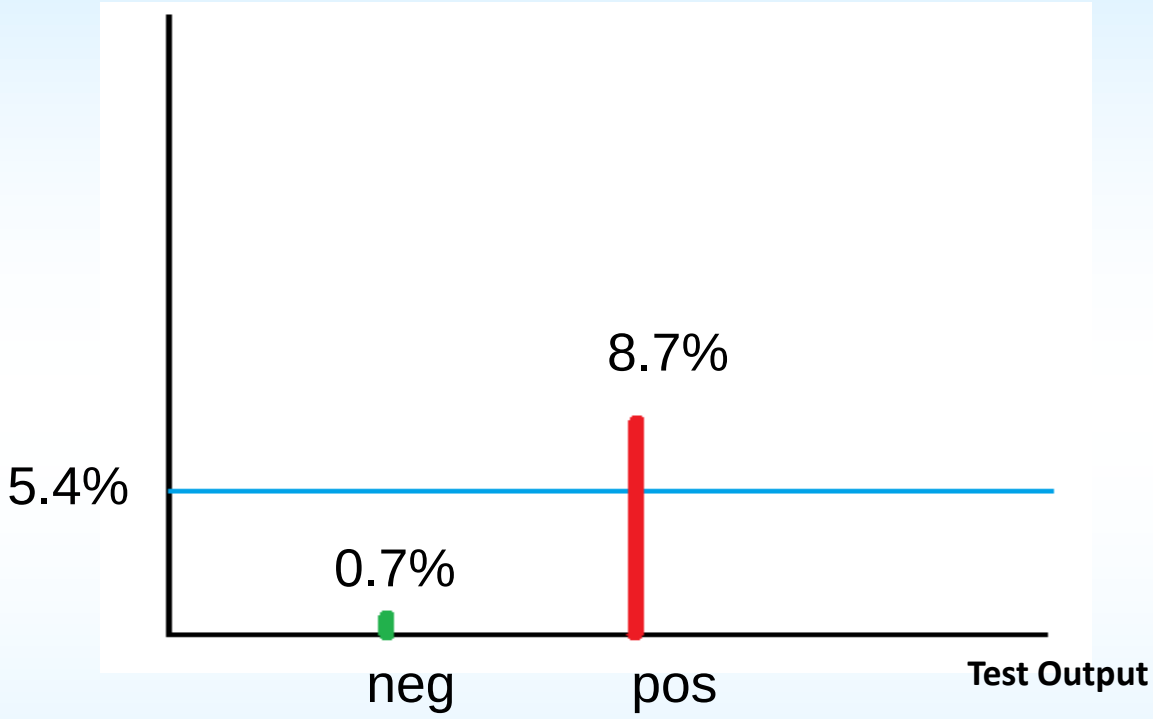
*Post-test risk for T +, post-test risk for T -.

Absolute Risks

Clinical Performance of the Test	
R (Pos) =Risk of D+ for T Pos	8.7% (66/760)
R (Neg)=Risk of D+ for T Neg	0.7% (4/540)
π = Pre-test risk of D+	5.4% (70/1,300)

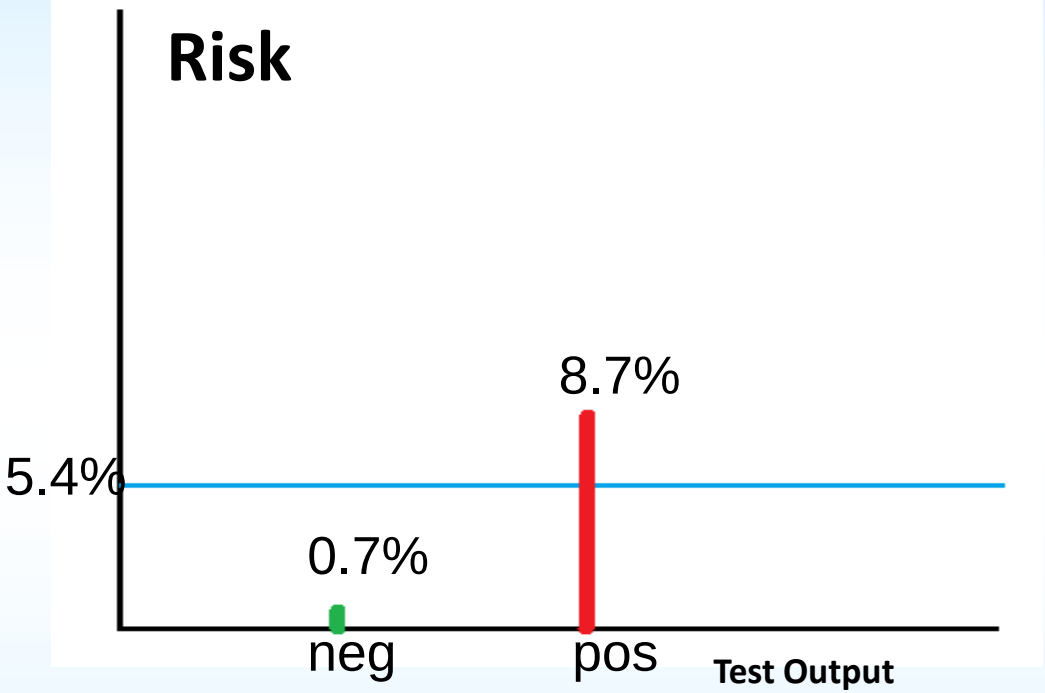


Statistically informative test output



Output of the test is statistically informative if risk of Disease for this test output is different from pre-test risk (prevalence).

Clinically informative test output



- Before test management:
  All women go to colposcopy;
- **Positive** test output is statistically informative but **not clinically informative**;
- It is acceptable not to sent to colpo if risk <1%
- **Negative** test result is statistically informative and **clinically informative**

Result of the test is clinically informative if

(A) Result of the test is statistically informative;

AND

(B) Risk for this test result is so much different from the pre-test risk that clinical management can be made based on this test result;

AND

(C)This clinical management is different from the clinical management without the test.

Candidate Test

- Finalize assay steps before the pivotal clinical study
- Define interpretations of all outputs, including equivocal

Example:

S/Co ≤ 1.0, Negative;
S/Co > 1.0, Positive

Example:

S/Co ≤ 0.9, Negative;
0.9< S/Co ≤1.1, Equivocal;
S/Co > 1.1, Positive

- Invalid result (control failed) ≠ Equivocal
- All results should be reported

Assessment of Target Condition

ATC (Gold Standard)-
best available method for establishing the presence or absence of the target condition (for example, colposcopy/biopsy for cervical cancer)

- Target condition is not necessary a disease (for example, it can be a success of some treatment).

- Target condition can be present at the same time when test T is applied; it can be present in future.

Confusion may sometimes arise when distinguishing between:

- **Reference Method**

  related to analytical performance (best method for measuring of analyte (quantitative) or for detection of analyte (qualitative))

- **Assessment of Target Condition**

  related to clinical performance
  (no recognized term, other terms as “gold standard”, “reference standard”, “clinical reference standard”, “diagnostic accuracy criteria”)

- Most of the time, reference method and an assessment of the target condition are different (e.g., HPV test for cervical cancer, total PSA for the prostate cancer).
- Sometimes, reference method and ATC are identical (e.g., flu test).

Archived (retrospective) samples

A good reason for pre-Sub

May be allowed

- How representative are archived samples (inclusion/exclusion criteria)
- Clinical context on specimens
- Only leftovers from big tumors (sample volumes)? Re-testing of samples close to the cutoff (sample volume)?
- Storage does not impact analyte of interest

Basic principle: Archived sample should provide unbiased estimates of test performance.

Test with Multiple Outcomes

Example #1: Multiplex test detecting two biomarkers A and B

These biomarkers are related to disease D

Four outcomes of the test:

(A+, B+)

(A+, B-)

(A-, B+)

(A-, B-)

Example #2: Test detects one biomarker (one SNP).

This biomarker is related to disease D.

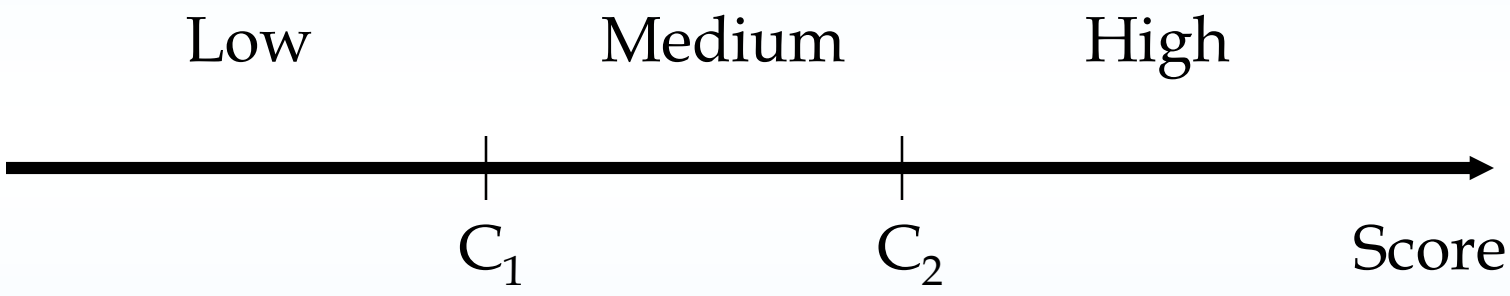
The biomarker has 3 possible results

(aa, aA, AA).

Example #3:

10 biomarkers combined in a score.

2 cutoffs are established that the score is reported as (High, Medium, Low)



How to describe performance of these tests?

Example : HPV Genotyping - 3 outcomes
(HPV16/18);
(Other High HPV types),
(HPV neg)

Test Results	Colposcopy/Biopsy		Total
	CIN2+	Not-CIN2+	
HPV 16/18	46	314	360
Other HPV types	20	380	400
HPV neg	4	536	540
Total	70	1230	1300

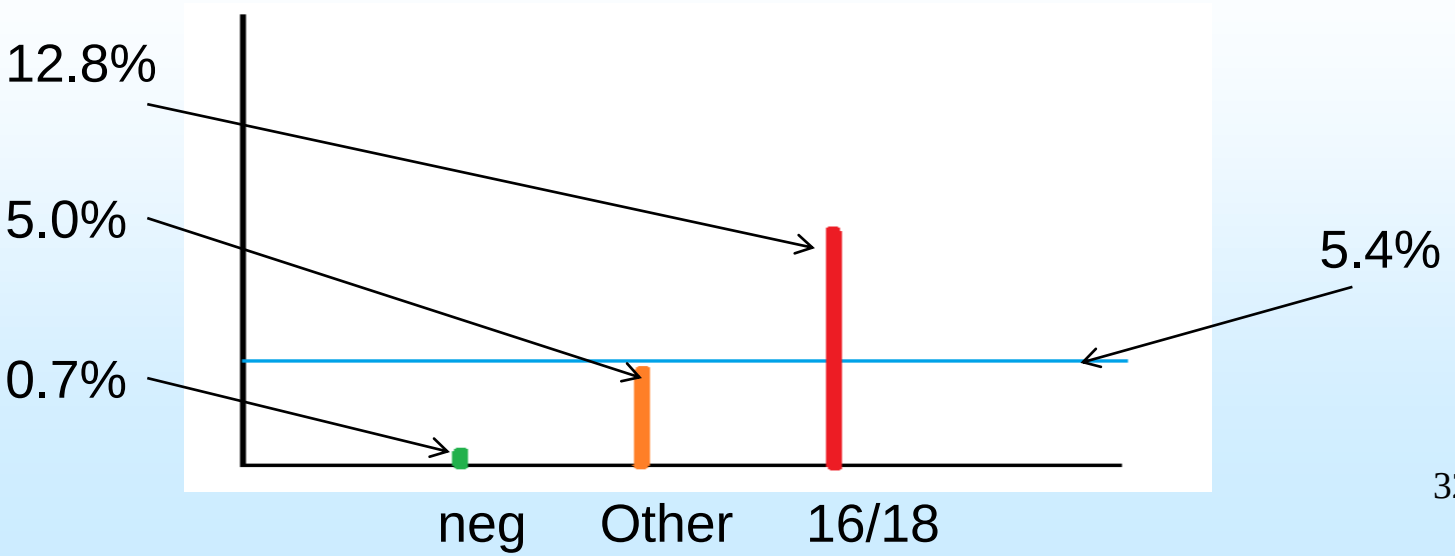
How to describe performance of this test?

Test with 3 outcomes: there are 3 risks R_X=Pr(D+|T=X)

Test Results	**Colposcopy/Biopsy**		Total	**Risk of CIN2+**
	CIN2+	**Not-CIN2+**		
HPV 16/18	46	314	360	**12.8%** **(46/360)**
Other HPV types	20	380	400	**5.0%** **(20/400)**
No HPV	4	536	540	**0.7%** **(4/540)**
Total	70	1230	1300	**5.4%** **(70/1300)**

Performance of the test is described by:
1) three Risks; 2) three frequencies (percent) of results; 3) pre-test probability.

Test Results	LR	Risk of Disease	Percent of results
HPV 16/18	2.6	12.8%	27.7%
Other HPV types	0.93	5.0%	30.8%
No HPV	0.13	0.7%	41.5%
Pre-test probability of CIN2+ is 5.4%			



III. Potential Biases

We considered an ideal scenario when N randomly selected subjects are from the intended use population and each subject has result of the test and verification of disease (D+, D-).

Potential Biases

1) **Selection bias** (when the study population does not represent the IU population)
2) **Spectrum bias**

1) Selection Bias

Examples of inappropriate study design

- Alzheimer's disease

In the study, the subjects with severe AD and healthy subjects were included => Selection bias – overestimation of performance.

- If the healthy subjects are not part of intended use population, do not include them in the clinical study (overestimation of specificity).

- Healthy subjects are used for determination of reference intervals.

2) Spectrum Bias

Example

Test ABC

Intended Use population		
Stage I	**50%**	**Sen=50%**
Stage II	**50%**	**Sen=90%**
Overall	**100%**	**70%** **0.5*50% + 0.5*90%**

Archived Specimens		
Stage I	**20%**	**Sen=50%**
Stage II	**80%**	**Sen=90%**
Overall	**100%**	**82%** **0.2*50% + 0.8*90%**

Sensitivity is biased (overestimated)

Guidance for Industry and Food and Drug Administration Staff –

Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and De Novo Classifications (2012)

The web address

http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm267829.htm

Example 3, pages 17-19

Thank you!

Marina.Kondratovich@fda.hhs.gov