Clinical Studies Design Considerations for Diagnostics

OIVD Pre-Submissions Workshop

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

April 19, 2012

Outline

I. Introduction

II. Clinical performance characteristics:
    Risks, absolute risks, relative risks;
    Likelihood ratios (LR) and odds ratios (OR).

III. Advantages of LR and OR:
    Comparison of tests;
    Tests with multiple outcomes.

I. Introduction

Key Elements

- Intended Use (IU)
  What is device supposed to do?

- Indications for Use (IFU)
  When should it be used?

- Both analytical and clinical data are supporting evidence for Intended Use and Indications For Use

Intended Use Statement (how/by whom device is used)

- What is the device measuring, identifying or detecting? (analyte, organism, ..)
- Specimen types, matrix (whole blood, serum,..)
- Conditions for use (hospital lab, home use,..)
- What type of data output? (quantitative, qualitative, semi-quantitative)

Indication for Use Statement (for what/on whom device is used)

- *Target condition*

- a particular disease, a disease stage, health status, or any other identifiable condition of event within a patient

- *Target population* (intended use population)

- those subjects for whom the test is intended to be used

- *Medical Testing Contexts*

- as, for screening, diagnosis, monitoring, prognosis, etc.

Examples of Medical Testing Contexts for cancer IVDs

- **Diagnosis** (target condition is present or not during the time of testing);
- **Screening** (maybe in a general population (asymptomatic subjects at average risk) or a subpopulation (subjects at high risk);
- **Risk assessment** (assessment of predisposition to disease in future);
- **Monitoring** (is therapy working for a patient?);

* This is not a comprehensive list

Examples of Medical Testing Contexts for cancer IVDs

- **Prognostic Biomarker**
  - The biomarker indicates disease aggressiveness in patients
  - Compare outcomes for biomarker positive patients vs biomarker negative patients
- **Treatment Predictive Biomarker**
  - The biomarker distinguishes patients who will benefit from those who will not benefit by treatment with a particular drug
  - Compare drug effect (i.e., treatment A vs treatment B) for biomarker positive vs biomarker negative patients

Intended Use/Indication For Use

Example

The HPV HR test is an *in vitro* diagnostic test for the qualitative detection of DNA from 14 high-risk Human Papilloma Virus (HPV) types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) in cervical specimens. *To screen* patients with atypical squamous cells of undetermined significance (ASCUS) cervical cytology results *to determine the need for referral to colposcopy.*

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

Every subject

Candidate Test:

Positive,
Negative

Clinical Reference Standard (Gold Standard):

D+ = Target condition present,
D- =Target condition absent

		Clinical Reference Standard		Total
		Disease Present	Disease Absent	
Candidate Test	Pos	66	694	760
	Neg	4	536	540
Total		70	1,230	1,300

Clinical performance refers to the degree of agreement between the results of the Candidate test and the results from the Clinical Reference Standard ("Gold" Standard).

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

Clinical Reference Standard

Clinical Reference Standard (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.

Clinical Reference Standard

Basic principles:

1) Candidate test results CANNOT be used in the Clinical Reference Standard

2) Clinical Reference Standard can classify each subject from the target population as “Target condition present” (Disease present) or “Target condition absent” (Disease absent).

Banked (retrospective) samples

A good reason for pre-IDE

May be allowed

- How representative are banked 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

Provide unbiased estimates of performance

II. Clinical Performance Characteristics

Clinical Performance Characteristics

- Clinical sensitivity and clinical specificity
- Positive and negative predictive values along with prevalence
- Absolute risks and relative risks

Consider Test with Two Outcomes (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 a Clinical Reference Standard ("Gold Standard") (D+, D-).

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

Clinical Performance of the Test	
R_1=Risk of D+ for T+ (PPV)*	8.7% (66/760)
R_0=Risk of D+ for T- (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 -.

Risks (Absolute Risks)

Example 2
Prognosis
Survival analysis

- Pr (metastatic disease within 5 years given the device outcome is “Positive”;
- Pr (metastatic disease within 5 years given the device outcome is “Negative”)

These are PPV and 1-NPV.
Performance is impacted by the prevalence of “metastatic disease within 5 years” in the clinical study.

Absolute Risks, Relative Risks

Clinical Performance of the Test	
R_1=Risk of D+ for T+ (PPV)	8.7% (66/760)
R_0=Risk of D+ for T- (1-NPV)	0.7% (4/540)
π = Pre-test risk of D+	5.4% (70/1,300)

- $\mathbf{R_1/\pi = 1.6}$: For a subject with T+, the risk increases by 1.6 times with regard to pre-test risk (=8.7%/5.4%);
- $\mathbf{R_0/\pi = 0.14}$: For a subject with T-, the risk increases by 0.14 times (decreased by 7.3 (1/0.14) times) with regard to pre-test risk (=0.7%/5.4%);
- $\mathbf{R_1/R_0 = 11.7}$: For a subject with T+, the risk increases by 11.7 times with regard to the subjects with T- (=8.7%/0.7%)

Absolute risks and relative risk depend on the sensitivity, specificity and also on the pre-test risk.

Se Sp

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

R_1

R_0

		D+	D-	Total
T	+	132	654	786
	-	8	506	514
Total		140	1,160	1,300

Se = 94.3% (66/70)
Sp = 43.6% (536/1,230)
Pre-test risk =5.4% (70/1,300)
R_1=8.7% (66/760)
R_0=0.7% (4/540)
R_1/π=**1.61**; R_0/π=**0.14**; R_1/R_0= **11.7**

Se = 94.3% (132/140)
Sp = 43.6% (506/1,160)
Pre-test risk =10.8% (140/1,300)
R_1=16.8% (132/786)
R_0=1.6% (8/514)
R_1/π=**1.56**; R_0/π=**0.14**; R_1/R_0= **10.8**

II. Clinical Performance Characteristics

Likelihood ratios are an alternative way to describe the performance of a test.

What is "likelihood ratio";

***Advantages of likelihood ratios*.**

Likelihood Ratio

Definition

Test has K possible results: $Result_1$, $Result_2$, …, $Result_K$.

Likelihood ratio for $Result_i$ is

$$\text{Likelihood Ratio}(\text{Result}_i) = \frac{\Pr(\text{Result}_i \mid \text{Disease})}{\Pr(\text{Result}_i \mid \text{No Disease})}$$

- LR is the probability of obtaining this test result in those with disease divided by the probability of obtaining this result in those without the disease;
- Each possible test result has a likelihood ratio.

Likelihood Ratios

Advantages

- Risks depend on corresponding LR and prevalence;
- Two qualitative tests should be compared using LR;
- Applicable for the tests with more than two test outcomes;
- Likelihood ratios are useful when there is verification bias*.

* For details, see Kondratovich, Marina (2008) Comparing Two Medical Tests When Results of Reference Standard Are Unavailable for Those Negative via Both Tests, *Journal of Biopharmaceutical Statistics,* 18: 1; 145-166

Likelihood Ratios

Consider a qualitative test T with 2 outcomes (Positive, Negative):

- LR (T+) = PLR and
- LR (T-) = NLR

Likelihood Ratios, Odds Ratios

Likelihood Ratios (LR) are another way to describe the performance of a test.

"Odds" are the ratio of the probability of one outcome to the probability of its opposite outcome.

Example:

Single fair coin with outcomes {Head, Tail}:

odds =1 because Pr (Head)=0.5 and

Pr (Tail)=1-0.5=0.5 =>

odds=1 (0.5/0.5=1).

Likelihood Ratios, Odds Ratios (Continued)

Subject from the IU population with pre-test risk π, two outcomes (D+, D-); Pr(D+) = π.

$$\text{Pre - test odds} = \frac{\pi}{1 - \pi}$$

After the test is performed (with knowledge of the test results):

$$\text{Post - test odds(T+)} = \frac{\Pr(D+ \mid T+)}{1 - \Pr(\text{D+} \mid \text{T+})} = \frac{R_1}{1 - R_1}$$

$$\text{Post - test odds(T-)} = \frac{\Pr(D+ \mid T-)}{1 - \Pr(\text{D+} \mid \text{T-})} = \frac{R_0}{1 - R_0}$$

Is there a relationship between post-test odds and pre-test odds?

Likelihood Ratios, Odds Ratios (Cont.)

$$\frac{R_1}{1-R_1} = LR(T+) \times \frac{\pi}{1-\pi}$$

$$LR(T+) = \frac{Se}{1-Sp}$$

$$\frac{R_0}{1-R_0} = LR(T-) \times \frac{\pi}{1-\pi}$$

$$LR(T-) = \frac{1-Se}{Sp}$$

Post-test odds = Likelihood Ratio x Pre-test odds

Likelihood Ratio of Positive Result

- How more often a positive test result occurs in persons with target condition compared to those without the target condition

$$LR(T+) = PLR = \frac{\Pr(T+ \mid D+)}{\Pr(T+ \mid D-)} = \frac{Se}{1-Sp}$$

$$\text{PLR} = \frac{\Pr(T+ \mid D+)}{\Pr(T+ \mid D-)} = \frac{Se}{1-Sp} = \frac{66/70}{694/1230} = \frac{0.943}{0.564} = 1.67$$

		Colposcopy/Biopsy		Total
		CIN2+	Not-CIN2+	
HPV	Pos	66 94.3%	694 56.4%	760
	Neg	4	536	540
Total		70	1,230	1,300

Likelihood Ratio of Negative Result

- Likelihood ratio of a negative test result

$$LR(T-) = NLR = \frac{\Pr(T- \mid D+)}{\Pr(T- \mid D-)} = \frac{1-Se}{Sp}$$

- How less likely a <u>negative</u> test result occurs in persons <u>with</u> the target condition compared to those <u>without</u> the target condition

$$NLR = \frac{\Pr(T-\mid D+)}{\Pr(T-\mid D-)} = \frac{1-Se}{Sp} = \frac{4/70}{536/1230} = \frac{0.057}{0.436} = 0.13$$

		Colposcopy/Biopsy		Total
		CIN2+	Not-CIN2+	
HPV	Pos	66	694	760
	Neg	4 5.7%	536 43.6%	540
Total		70	1,230	1,300

- ***Likelihood Ratios and Predictive Values***

Predictive values (risks) depend of the corresponding LR and prevalence

$$\frac{R_1}{1-R_1} = LR(T+) \times \frac{\pi}{1-\pi} \qquad LR(T+) = \frac{Se}{1-Sp}$$

$$\frac{R_0}{1-R_0} = LR(T-) \times \frac{\pi}{1-\pi} \qquad LR(T-) = \frac{1-Se}{Sp}$$

Odds Ratio (OR) = LR(T+)/LR(T-)

If π is close to 0, then R_1 is close to LR(T+) x π and
R_0 is close to LR(T-) x π

If π is close to 0, then R_1/R_0 is close to OR=LR(T+)/LR(T-)

Likelihood Ratios, Odds Ratios (Cont.)

Post-test odds = Likelihood Ratio x Pre-test odds

$$LR(T+) = \frac{Se}{1-Sp} \qquad LR(T-) = \frac{1-Se}{Sp}$$

Likelihood Ratios do not depend on the pre-test risk.

$$\text{Odds Ratio (OR)} = \frac{LR(T+)}{LR(T-)}$$

Odds Ratio does not depend on the pre-test risk.

Summary:

$$PLR = \frac{Se}{1-Sp}; NLR = \frac{1-Se}{Sp}$$

- PPV (risk of disease for pos. result) depends on PLR and prevalence.
  The larger PLR, the higher risk of disease for pos. result (higher PPV).

- NPV depends on NLR and prevalence.
  The smaller NLR, the lower risk of disease for neg. result (higher NPV).

- The larger OR (odds ratio), the larger RR (relative risk).

If prevalence (baseline risk) is <u>low</u>, then

- $R_1/\pi \approx$ PLR
- $R_0/\pi \approx$ NLR;
- Relative Risk (R_1/R_0) ≈ Odds ratio (OR=PLR/NLR);

Pre-test Probability of Disease	Test Result	Likelihood Ratio	Post-test Probability of Disease
5.4%	Positive	PLR=1.67	8.7%
	Negative	NLR=0.13	0.7%

- R_1=8.7%; π=5.4%; R_1/π=1.61 ------- PLR=1.67
- R_0=0.7%; π=5.4%; R_0/π=0.14 ------- NLR=0.13;
- R_1/R_0= 11.7 (8.7%/0.7%) ------- OR = PLR/NLR = 12.8

LR is a way of quantifying how much a given test result changes the pre-test probability of disease in a patient.

LR	Interpretation
> 10	*Large* increase in the likelihood of disease
5 - 10	*Moderate* increase in the likelihood of disease
2 - 5	*Small* increase in the likelihood of disease
1 - 2	*Minimal* increase in the likelihood of disease
1	*No change* in the likelihood of disease
0.5 - 1.0	*Minimal* decrease in the likelihood of disease
0.2 - 0.5	*Small* decrease in the likelihood of disease
0.1 - 0.2	*Moderate* decrease in the likelihood of disease
< 0.1	*Large* conclusive decrease in the likelihood of disease

What is clinically acceptable, depend on disease and the pre-test risk of disease (prevalence).

Likelihood Ratios and Comparison of Two Qualitative Tests

Example: T_1 – pre-surgical assessment (pos, neg)
T_2 – qualitative test (pos, neg)

Can T_2 be used instead of T_1?

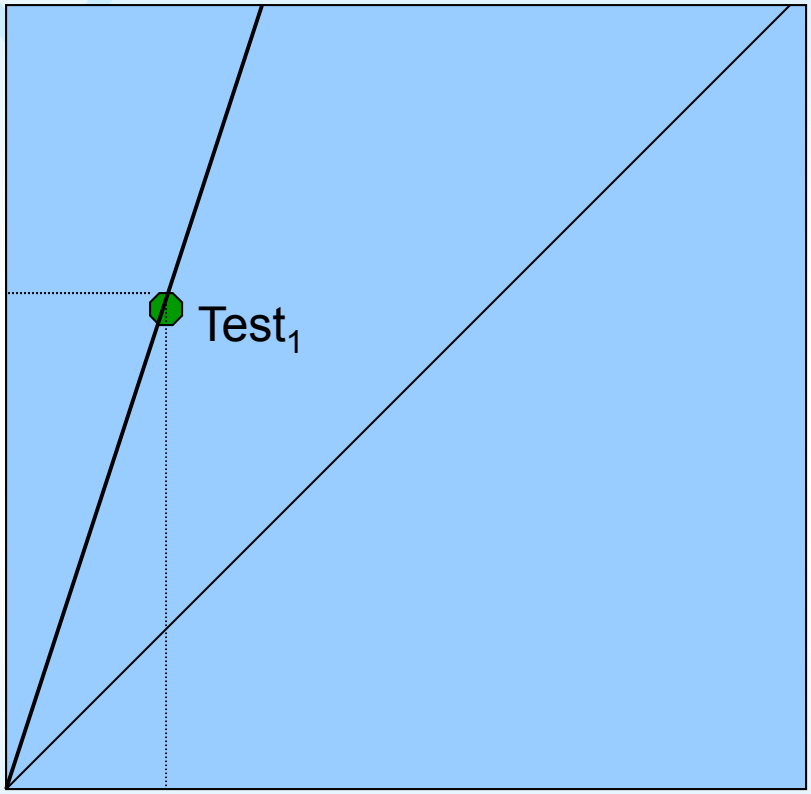
	T_1 test alone	**T_2 test alone**
sensitivity	**74.8% (113/151)**	**84.8% (128/151)**
specificity	**79.2% (281/355)**	**49.9% (177/355)**

How to compare these two tests?

Two tests should be compared using the likelihood ratios (not sensitivity, specificity).

For more details, Biggerstaff, B.J. (2000). Comparing diagnostic tests: a simple graphic using likelihood ratios. *Statistics in Medicine* 19: 649-663.

Same prevalence
$Test_1$: (Se_1, Sp_1, π) and $Test_2$: (Se_2, Sp_2, π)



$$\frac{R_1}{1-R_1} = PLR \times \frac{\pi}{1-\pi}$$

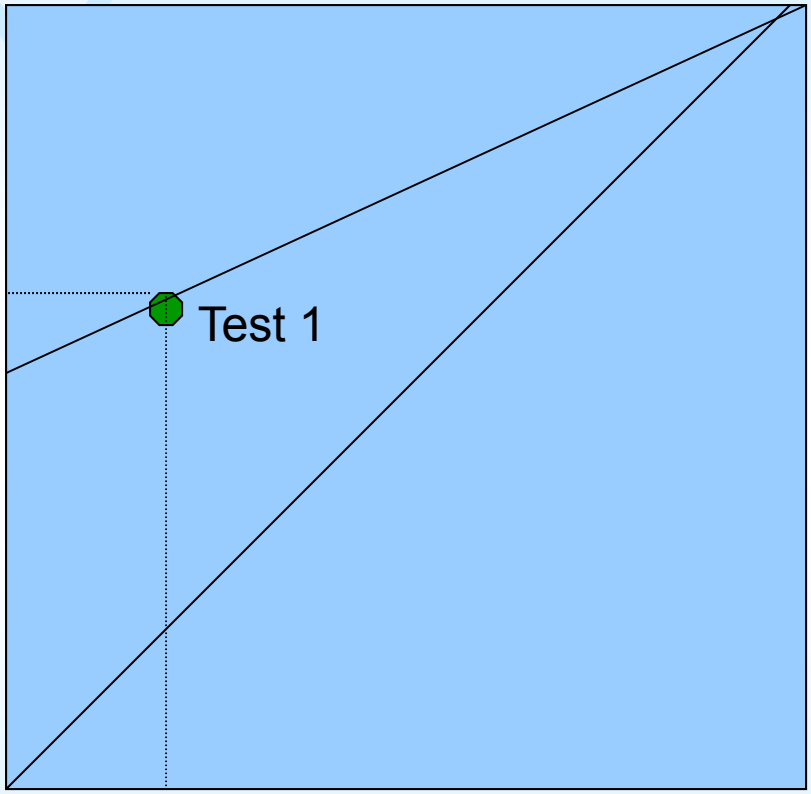
$PLR_1 = PLR_2 \Leftrightarrow PPV_1 = PPV_2$

PLR=Se/(1-Sp)

PLR is a tangent of the line (0,0) –(Se_1, Sp_1)

The larger PLR, the higher PPV.

Same prevalence
Test 1: (Se_1, Sp_1, π) and Test 2: (Se_2, Sp_2, π)



$$\frac{R_0}{1-R_0} = NLR \times \frac{\pi}{1-\pi}$$

$NLR_1 = NLR_2 \Leftrightarrow NPV_1 = NPV_2$

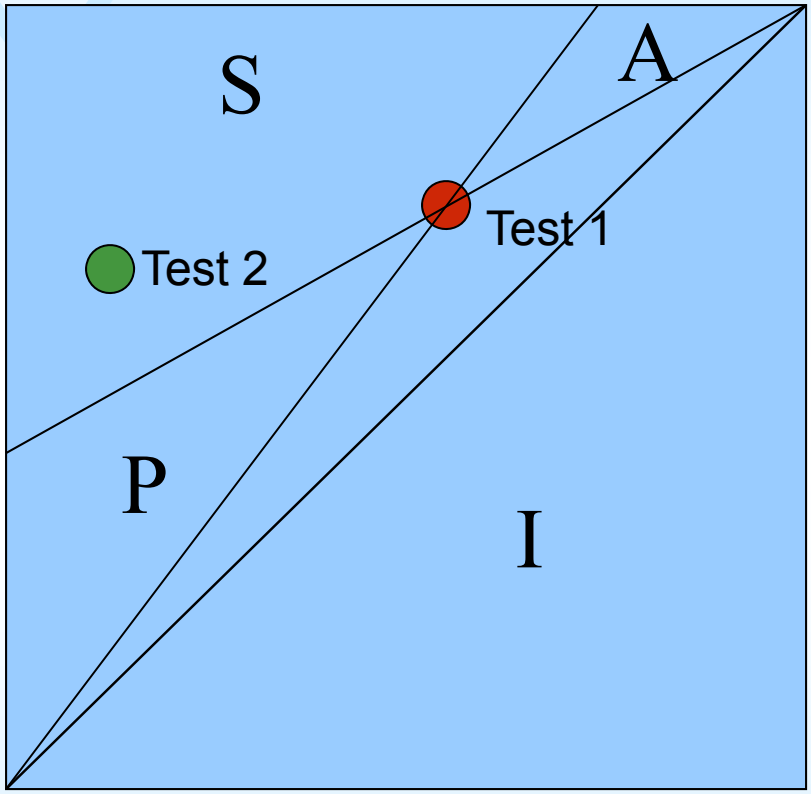
NLR=(1-Se)/Sp

NLR is a tangent of the line (1,1) –(Se_1, Sp_1)

⇔

The smaller NLR, the lower 1-NPV (higher NPV).

Regions:



S – superior overall
(PPV_2>PPV_1 and NPV_2>NPV_1)

I – inferior overall
(PPV_2<PPV_1 and NPV_2<NPV_1)

A – superior for confirming absence of disease
(PPV_2<PPV_1 and NPV_2>NPV_1)

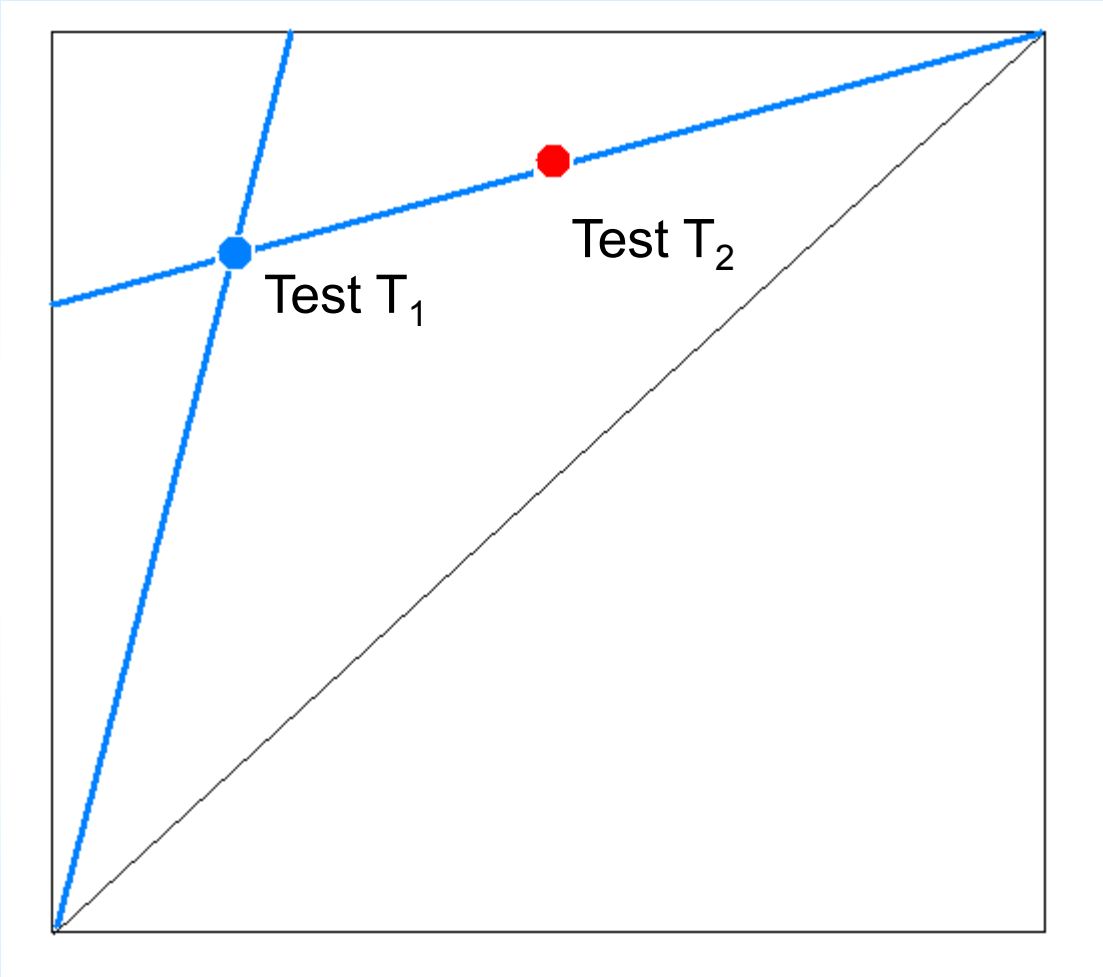
P – superior for confirming presence of disease
(PPV_2>PPV_1 and NPV_2<NPV_1)

Example: T_1 – pre-surgical assessment (pos, neg)
T_2 – qualitative test (pos, neg)

Can T_2 be used instead of T_1?

	T_1 test alone	T_2 test alone
sensitivity	74.8% (113/151)	84.8% (128/151)
specificity	79.2% (281/355)	49.9% (177/355)
PLR = se/(1-sp)	**3.59** **95% CI: 2.887 to 4.50**	**1.69** **95% CI: 1.49 to 1.92**
NLR = (1-se)/sp	**0.32** 95% CI: 0.24 to 0.42	**0.31** 95% CI: 0.21 to 0.45
Prevalence = 29.8%		
PPV	**60.4% (113/187)**	**41.8% (128/306)**
1-NPV	11.9% (38/319)	11.5% (23/200)

Relationship between test T_1 and test T_2



Can T_2 be used instead of T_1? **NO!!!**

- ***Likelihood Ratios and 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 four biomarkers (four SNPs).

These biomarkers are related to disease D.

Each biomarker has 3 possible results

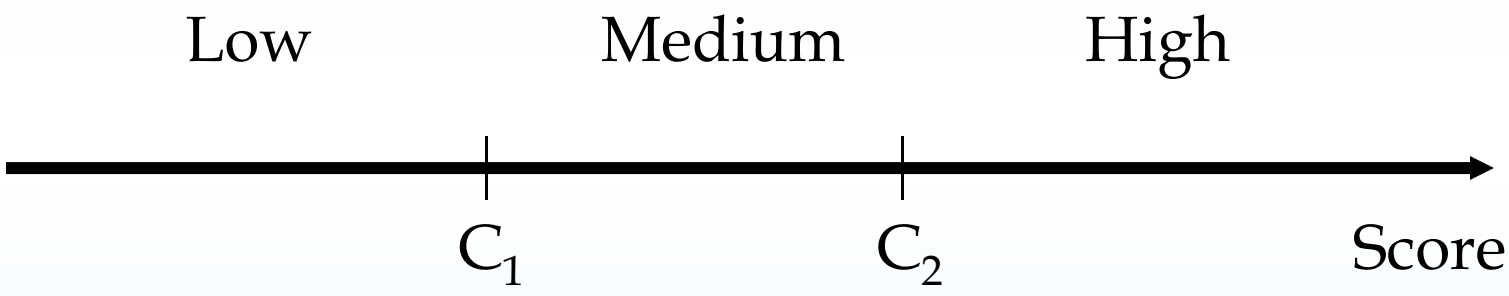
(aa, aA, AA).

Then test has 81 possible results: 81=3x3x3x3.

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),
(No HPV)

Test Results	Colposcopy/Biopsy		Total
	CIN2+	Not-CIN2+	
HPV 16/18	46	314	360
Other HPV types	20	380	400
No HPV	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)**

We can calculate 3 likelihood ratios LR(T=X)

R_X depends on LR(T=X) and prevalence

$$\frac{R_X}{1-R_X} = LR(T=X) \times \frac{\pi}{1-\pi}$$

Test Results	Colposcopy/Biopsy		LR
	CIN2+	Not-CIN2+	
HPV 16/18	46 65.7%	314 25.5%	**2.6** (65.7%/25.5%)
Other HPV types	20 28.6%	380 30.9%	**0.93** (28.6%/30.9%)
No HPV	4 5.7%	536 43.6%	**0.13** (5.7%/43.6%)
Total	70 100%	1230 100%	

Performance of the test with three outcomes is described: 1) pre-test probability; 2) three LRs; 3) three frequencies (percents) of results.

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%			

Relative Risks:

- There are many combinations of risks for calculations of relative risks;
- Relative risk relatively to π => it is related to LR
- It is convenient to calculate relative risk relatively to "HR neg" => it is related to OR

If prevalence (pre-test risk) is <u>low</u>, then

- ❖ Risk of Disease for Result/Pre-test risk ≈ LR (Result)
- ❖ Relative Risk (RR) ≈ Odds ratio

Example: 4 SNPs $LR(A_i, B_j, C_k, D_l)$

Consider the test for the target condition with 4 markers (SNPs); ORs for **individual** markers are obtained.

		SNP_1	SNP_2	SNP_3	SNP_4
LR	$Result_3$		1.27 BB		1.55 DD
	$Result_2$	1.05 Aa			
	$Result_1$			0.77 cc	

Multiplicative Model: an assumption that all four SNPs are independent (no interactions) - this assumption may be not correct.

$$LR(A_i, B_j, C_k, D_l) = 1.05 \times 1.27 \times 0.77 \times 1.55 = 1.59$$

Summary

- Clinical performance of the qualitative test with two outcomes can be described by a pair of sensitivity and specificity, or by PPV, NPV and prevalence.

- Risks and relative risks measure probabilities of events in a way that is interpretable and consistent with how people think.

- In addition, clinical performance of a medical test can be described by likelihood ratios.

Summary

Advantages:

- LRs do not depend on the prevalence;
- Absolute risks depend on the corresponding LR and prevalence;

- LRs are useful for comparing two qualitative tests with binary outcomes;
- LRs are useful for describing performance of the tests with multiple outcomes.

Because they do not depend on the pre-test risk, LRs and ORs can be calculated even in the case-control studies.

It is easy to adjust an OR for other variables (logistic regression)

Thank you!

Marina.Kondratovich@fda.hhs.gov