Diagnostic Tests:

Performance Characteristics and Selection Criteria

Rosanna W Peeling
Professor and Chair, Diagnostic Research
London School of Hygiene & Tropical Medicine
Director, International Diagnostics Centre

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

The International Diagnostics Centre

www.idc-dx.org

Learning Objectives

- Identify ways that diagnostic test performance can be expressed

- Distinguish between evaluation, validation and verification of test performance and operational characteristics

- Selecting a diagnostic test

Performance Characteristics

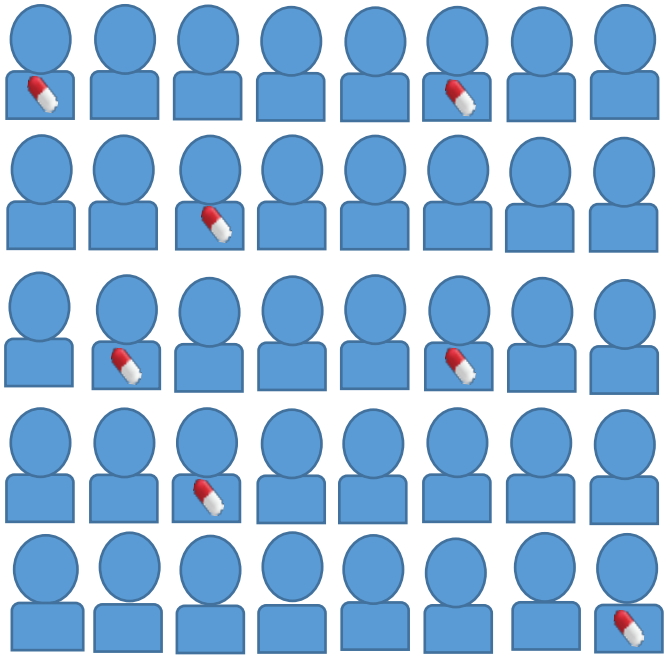
The basic performance characteristics of a test designed to distinguish infected from uninfected individuals are:

- **Sensitivity:** the probability that a truly infected individual will test positive

- **Specificity:** the probability that a truly uninfected individual will test negative

Sensitivity and specificity are usually determined against a **reference standard** test ("gold standard"), which represents the true value of what is being measured.

Evaluation of a diagnostic test is particularly challenging when there is no recognised reference standard test.



Definition: Sensitivity

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

$$\frac{\text{\# of Positives measured by the test evaluated}}{\text{\# of Positives measured by the ref. standard}} \times 100$$

		Reference Standard	
		+	-
Test outcome	+	a	b
	-	c	d

Sensitivity = a / a+c

Definition: Specificity

- Ability of a test to correctly identify an uninfected individual

- % who are not infected and have a negative test result:

$$\frac{\text{\# of negatives id. by test evaluated}}{\text{\# of negatives id by ref standard}} \times 100$$

		Reference Standard	
		+	-
Test outcome	+	a	b
	-	c	d
		Sensitivity: = a/a+c	Specificity: = d/b+d

Measures of Test Performance of Value to Clinical Decision Making

- **Positive predictive value** (PPV) The probability that those testing positive by the test are truly infected.

- **Negative predictive value** (NPV) The probability that those testing negative by the test are truly uninfected.

PPV and NPV depend not only on the sensitivity and specificity of the test, but also on the prevalence of infection in the population

Using the same test in a population with higher prevalence increases **positive predictive value**

.

PPV depends on Disease Prevalence

In a study of fever patients in Burkina Faso, malaria RDTs positive:

- In rainy season: 443/650 (68.2%)
- in dry season: 113/400 (28.3%)

In rainy season when malaria transmission is high, the rapid test PPV ranged from 38% for adults to 82% for infants, while the NPV ranged from 84% for infants to > 99% for adults

During the dry season when malaria transmission is low, the same test has a PPV of only 9% and NPV of 99.8% for malaria-attributable fever

Bisoffi *et al.,* Accuracy of a rapid diagnostic test on the diagnosis of malaria infection and of malaria - attributable fever during low and high transmission season in Burkina Faso *Malaria Journal* 2010, 9:192

Definition: Positive Predictive Value

- **Positive predictive value is the** probability that a patient with a **positive** test result actually has the disease
  - Measures the reliability of a positive test result
  - The ideal value of the PPV, with a perfect test, is 1 (100%), and the worst possible value would be 0.
- Calculation:

$$\frac{\text{\# of True Positives}}{\text{\# of True Positives + \# of False Positives}} \times 100$$

		Reference Standard	
		+	-
Test outcome	+	a	b
	-	c	d

PPV = a/a+b

NPV = d/c+d

Sensitivity: = a/a+c

Specificity: = d/b+d

Definition: Negative Predictive Value

- % with a negative test result who are not infected
  - Measures the reliability of a negative test result
  - The ideal value of the NPV, with a perfect test, is 1 (100%), and the worst possible value would be 0.
- Calculation

$$\frac{\text{\# of True Negatives}}{\text{\# of True Negatives + \# of False Negatives}} \times 100$$

		Reference Standard	
		+	-
Test outcome	+	a	b
	-	c	d

PPV = a/a+b

NPV = d/c+d

Sensitivity: = a/a+c

Specificity: = d/b+d

Types of Diagnostic Evaluations

Lab evaluation (use well characterised panels)

- A test's ability to measure an analyte accurately and reliably:
  - Analytical sensitivity & specificity
  - Reproducibility

Field Evaluation (use consecutive patient samples)

- A diagnostic test's ability to detect or predict the disease/condition associated with an analyte – clinical decision making
  - Clinical sensitivity & specificity, positive and negative predictive values
  - Robustness and feasibility

Utility Studies (use mathematical models)

- The balance of risks and benefits associated with the use of a test for the patient and for the control programme
  - Health outcomes
  - Cost-effectiveness
  - Implementation research (e.g. changing patient pathways, linkage to care, data connectivity)

Sample Size

Increasing sample size reduces uncertainty around estimates of sensitivity and specificity. Uncertainty is summarized by confidence intervals (CI)

The narrower the confidence interval, the greater the precision of the estimate:

A 95% CI is often used – you can be 95% confident that the interval contains the true values of sensitivity and specificity

Number of infected (non-infected) subjects required (as defined by ref. test)	Estimated Test Sensitivity (or Specificity)				
	50%	60%	70%	80%	90%
	95% confidence interval around the estimated sensitivity				
50	+/- 14%	+/- 14%	+/- 13%	+/- 11%	+/- 8%
100	+/- 10%	+/- 10%	+/- 9%	+/- 8%	+/- 6%
150	+/- 8%	+/- 8%	+/- 7%	+/- 6%	+/- 5%
200	+/- 7%	+/- 7%	+/- 6%	+/- 6%	+/- 4%
500	+/- 4%	+/- 4%	+/- 4%	+/- 4%	+/- 3%
1000	+/- 3%	+/- 3%	+/- 3%	+/- 2%	+/- 2%

Evaluation, Validation and Verification

- **Evaluation**:
  - Evaluation is the process of determining merit, worth, or significance
  - sample size varies depending on type of evaluation – lab, clinic, utility

- **Validation**:
  - usually applies to laboratory developed assays e.g. primers and probes based on published protocols for the detection of SARS-CoV-2 RNA should be validated with reference strains and negative controls before use on patient samples

- **Verification:** ~40 samples
  - test kits procured should be checked against performance claims in the product insert before use on patient samples
    - Precision: assay variability (10 samples for each parameter)
    - Accuracy: concordance of + and – (10 negative and 20 positive samples)
    - Reference and reportable ranges: for quantitative assays

Reproducibility and Repeatability

- **Reproducibility and Repeatability are measures of precision**

- The **reproducibility** of the test is a measurement of the closeness of agreement between test results when the conditions for testing or measurement changes
  - reproducibility may be measured between operators, between different test sites, and using different instruments and different kit lots.
  - can be performed using a minimum panel of 10 samples (include positive and negative controls) with 3 different operators; at 3 sites, and the samples evaluated on three different days at each site.

- The **repeatability** of the test results refers to measuring the closeness of test results when no conditions of measurement changes, e.g. 3 samples repeated 6 times by the same operator

Selecting a Diagnostic Test

1. **Define the test's purpose – why, what, where, who?**
2. **Review the market**
3. **Review regulatory approval by international and national bodies**
4. **Determine the test's optimal diagnostic accuracy**
5. **Determine the test's diagnostic accuracy in practice**
6. **Monitor the test in routine use**

Kosack et al. Bull World Health Organ 2017;95:639–645. http://dx.doi.org/10.2471/BLT.16.187468

Example: COVID-19

- If a test for COVID-19 is 98% specific, in 100 tests there will be 2 false positive results

- If the prevalence of COVID in the general population is 50%:
  100 tests will give **50 true positive** and **2 false positive** results (PPV = 96%)

- If the prevalence of COVID-19 is 1%:
  100 tests will give **1 true positive** and **2 false positive** results (PPV = 33%)

Summary

- Sensitivity and specificity are inherent characteristics of a diagnostic test
- Positive and negative predictive values vary by prevalence in the population
- Test performance estimates can be affected by the reference standard used, sample size and appropriate use
- Evaluation, validation and verification of test performance should be carried out in appropriate circumstances
- Criteria for selecting a diagnostic test should take into account how the test will be used, in what settings and by whom, accuracy of test performance and feasibility and ability to assure quality of tests and testing


Thank you