Locksley Messam
Conor McAloon
Eleftherios Meletis

# Diagnostic test evaluation with Bayesian latent class models

UCD DUBLIN

COST
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY

COST is supported by the EU Framework
Programme Horizon 2020

HARMONY
Novel tools for test evaluation and
disease prevalence estimation

1

---

UCD DUBLIN

# Methods of Test Validation

|      | D+ | D- |
|------|----|----|
| T+   | TP | FP |
| T-   | FN | FN |

COST
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY
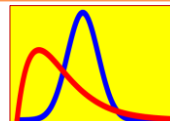
COST is supported by the EU Framework
Programme Horizon 2020

HARMONY
Novel tools for test evaluation and
disease prevalence estimation

**Locksley L. McV. Messam  Dr. Med. Vet., PhD.**
Lecturer in Veterinary Epidemiology
University College Dublin, School of Veterinary Medicine,
Section: Herd Health and Animal Husbandry

2

1

# Topics

- Test validation (Meaning)

- Sampling of Population(s)

- Gold Standard Approaches

- Non-Gold Standard Approaches

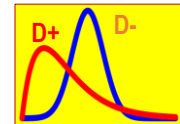|     | D+ | D- |
|-----|----|----|
| T+  | TP | FP |
| T-  | FN | FN |

# Test Validation

|     | D+ | D- |
|-----|----|----|
| T+  | TP | FP |
| T-  | FN | TN |

"The process of establishing the sensitivity and specificity of a diagnostic test"
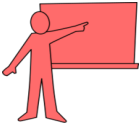
# Requirements

## Validity

- Internally valid – Unbiased for the sample selected (No retesting, blinding to true status)
- Externally valid – Unbiased for the target population (consistency between study pop. and target pop., re-validate if target pop changes)

## Precise

- As little random error as possible (ie narrow intervals)

5

# Analytic Approaches

## Gold Standard (Reference Test)

- Suitable test necessary (Se = Sp =100%)
- Definition - Independent of test to be validated.
- Single or multiple tests (Series or Parallel)

## Non-Gold Standard methods

- No reference test

6

3

# Analytic Approaches (Gold Standard)

## Cross-sectional sampling: Complete verification

- Random sampling (ideal)
- Each individual tested with Gold Standard
- Each individual tested with Diagnostic test
- Equal verification for T+ and T- (no work up bias)
- Sequence of testing unimportant
- Blinding essential (Ignorance of true disease status)
- Sample prevalence unbiased estimate of TP

7

# Analytic Approaches (Gold Standard)

## Cross-sectional sampling: Complete verification

E.g. Trypanosoma in cattle (*Greiner and Gardner, PVM 45: 1-2 (2000)*)

- 183 cattle, randomly sampled
- Reference test – DNA detection
- Validation of ELISA
- Se = 0.64; 95% CI: 0.52 - 0.74
- Sp = 0.65; 95% CI: 0.55 - 0.74
- P = 0.44; 95% CI: 0.36 – 051

|     | D+ | D-  |     |
|-----|----|-----|-----|
| T+  | 51 | 36  | 87  |
| T-  | 29 | 67  | 96  |
|     | 80 | 103 | **183** |

8

4

# Analytic Approaches (Gold Standard)

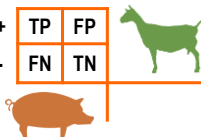## Cross-sectional sampling: Partial verification

- Gold standard risky, invasive, expensive
- Each individual tested with Diagnostic test (to be validated)
- Some individuals tested with Gold Standard
- Sequence of testing: Diagnostic test first
- Blinding not always possible (Ignorance of true disease status)
- Randomly choose among T+ and T- (Verification independent of true disease status).

|     | D+     | D-     | ?      |           |
|-----|--------|--------|--------|-----------|
| T+  | A (18) | B (5)  | E (92) | n1 = 115  |
| T-  | C (2)  | D (15) | F ( 68) | n2 = 85   |

$$R_1 = \frac{(A + B)}{n_1}, \ R_2 = \frac{(C + D)}{n_2}$$

$$Se = \frac{A/R_1}{A/R_1 + C/R_2}, \ Sp = \frac{D/R_2}{D/R_2 + B/R_1}$$

$$PPV = \frac{A}{A + B}, \ NPV = \frac{D}{C + D}$$

---

|     | D+ | D- |
|-----|----|----|
| T+  | TP | FP |
| T-  | FN | TN |

# Test Validation
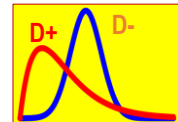
## Convenience Sampling:

Samples of known disease status not collected to represent a known population (e.g., experimental infections).

- Misclassification unlikely
- Artificially high challenges doses
- SPF animals (little cross-reacting antibodies)
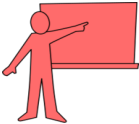- Overestimation of Se and Sp!!

# Analytic Approaches (Non-Gold Standard)

## Motivation

- Gold standards: Costly, invasive, rare.
- Latent Class Methods – True disease status is unknown

## Approaches

- Frequentist: Hui – Walter *(Hui, Walter (1980), Biometrics 36: 167-71)*
- Bayesian

11

# Analytic Approaches (Bayesian Non-Gold Standard)

**Approach to parameter estimation (3 elements):**

1. Probability distribution for the parameter (characterizes uncertainty).

2. Evidence in the form of data (Likelihood function).

3. Posterior Distribution

Linked by Bayes theorem

12

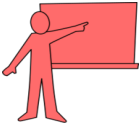## Analytic Approaches
## (Bayesian Non-Gold Standard)

**Bayes theorem – (Inverse probabilities)**

Calculating probability of an antecedent event (A) based on occurrence of subsequent (S) events.

$$P(A|S) = \frac{P(S|A) \times P(A)}{P(S)}$$
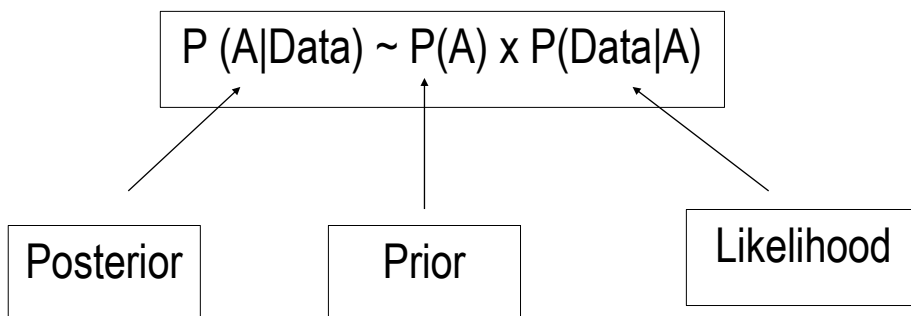
$$P(A|Data) \sim P(A) \times P(Data|A)$$

13

## Analytic Approaches
## (Bayesian Non-Gold Standard)

**Bayes Theorem**

$$P(A|Data) \sim P(A) \times P(Data|A)$$

Posterior

Prior

Likelihood

14

7

# Analytic Approaches
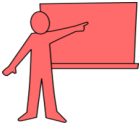## (Bayesian Non-Gold Standard)

**Prior [P(A)]**

- Researcher's uncertainty (or knowledge) independent of data
- From experts
- Historical or current (but independent) data
- Exact value, range, probability distribution (mathematical function)

**Likelihood Function [P(Data|A)]**

- Expresses probability of the data given the parameter (data treated as given, while parameter varies)

15

# Analytic Approaches
## (Bayesian Non-Gold Standard)

**Posterior [P(A|Data)]**

- Probability distribution

- Results: Medians, modes and outer percentiles

- Probability Intervals: Range of values that contain parameter with a given probability
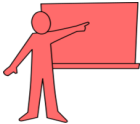  - Distinct from confidence intervals!

16

# Analytic Approaches
## (Bayesian Non-Gold Standard)

**Fundamental Principles**

• Probability distribution for all parameters (uncertainties)

• Obey laws of probability

# Analytic Approaches
## Bayesian Non-Gold Standard

**Options:**

**One test - One Population**

3 parameters to be estimated (Se, Sp, P), df = 1 (fixed sample size, number testing positive) $\rightarrow$ Problematic

**Two tests – One Population**

5 parameters to be estimated (2 Se, 2 Sp, P), df = 3 (fixed sample size, only 3 cells can vary) $\rightarrow$ (Still) Problematic

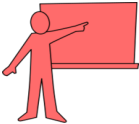## Analytic Approaches
## (Bayesian Non-Gold Standard)

**Conditional (in)Dependence:**

Tests are **conditionally independent** when Se or Sp of one test is **not dependent** on results of the other test with respect to both infected and non-infected individuals

- Partially

- Completely

19

## Analytic Approaches
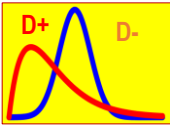## (Bayesian Non-Gold Standard)

**Options:**

**Two tests – Two populations (conditionally independent)**
- Se for both tests are equal, Sp for both tests are equal
- 6 parameters to be estimated, df = 6

**Two tests – Two populations (conditionally dependent)**
- Se for both tests are equal, Sp for both tests are equal
- 8 parameters to be estimated, df = 6
- **Need good prior information**

20

# Methods of Test Validation

|     | D+ | D- |
|-----|-----|-----|
| T+  | TP | FP |
| T-  | FN | TN |

?