Diagnostic test evaluation with perfect reference test

Paolo Eusebi

16/09/2021

Four phases in architecture of diagnostic research

- **Phase I** Do test results in patients with the target disorder differ from those in normal people?
- **Phase II** Are patients with certain test results more likely to have the target disorder than patients with other test results?
- **Phase III** Does the test result distinguish patients with and without the target disorder among patients in whom it is clinically reasonable to suspect that the disease is present?
- **Phase IV** Do patients who undergo this diagnostic test fare better (in their ultimate health outcomes) than similar patients who are not tested?

Measures of diagnostic accuracy

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

- TP = True Positives
- FP = False Positives
- TN = True Negatives
- FN = False Negatives

Measures of diagnostic accuracy

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

- Sensitivity = TP/D+
- Specificity = TN/D-
- PPV = TP/T+
- NPV = TN/T-

Measures of diagnostic accuracy

- Sensitivity and specificity do not depend on the disease prevalence.
- PPV and NPV depend on the sensitivity, specificity, and the disease prevalence.

$$PPV = \frac{Se \cdot p}{Se \cdot p + (1 - Sp) \cdot (1 - p)}$$

$$NPV = \frac{Sp \cdot (1 - p)}{(1 - Se) \cdot p + Sp \cdot (1 - p)}$$

Measures of diagnostic accuracy: Bayesian reasoning

Suppose a new HIV test has 95% Se and 98% Sp and is to be used in a population with a HIV prevalence of 1/1000.

We can simulate the diagnostic results in a population of 100,000 individuals.

	HIV+	HIV-
T+	95	1,998
T-	5	97,902

- PPV = 95/(95+1,998)= 0.0045 (Only 4.5%!)

Example from Spiegelhalter et al. *Bayesian approaches to clinical trials and health-care evaluation* John Wiley & Sons, 2004.

Measures of diagnostic accuracy: Bayesian reasoning

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

$$\theta = \theta_0(HIV+), \theta_1(HIV-)$$

- Data probability

$$P(Y) = P(\theta_0) \cdot P(Y|\theta_0) + P(\theta_1) \cdot P(Y|\theta_1) =$$

$$= 0.001 \cdot 0.95 + 0.999 \cdot 0.02 = 0.02093$$

- Prior

$$P(\theta_0) = 1/1000 = 0.001$$

- Likelihood (95% Se)

$$P(Y|\theta_0) = 0.95$$

- Posterior

Measures of diagnostic accuracy

- Frequencies

	D+	D-
T+	y[1]	y[3]
T-	y[2]	y[4]

- Probabilities

	D+	D-
T+	prob[1]	prob[3]
T-	prob[2]	prob[4]

Measures of diagnostic accuracy

	D+	D-
T+	prob[1]	prob[3]
T-	prob[2]	prob[4]

We can define parameters as function of the cell probabilities:

- p = prob[1]+prob[2]
- Se = prob[1]/(prob[1]+prob[2])
- Sp = prob[4]/(prob[3]+prob[4])

Or cell probabilities in terms of parameters:

- prob[1] = Se*p
- prob[2] = (1-Se)*p
- prob[3] = Sp*(1-p)
- prob[4] = Sp*(1-p)

Bayesian model

```
"model {

# likelihood
  y[1:4] ~ dmulti(prob[1:4], n)

  prob[1] <- p * Se
  prob[2] <- p * (1 - Se)
  prob[3] <- (1 - p) * (1 - Sp)
  prob[4] <- (1 - p) * Sp

# priors
  p ~ dbeta(1, 1)
  Se ~ dbeta(1,1)
  Sp ~ dbeta(1,1)
"
```

Bayesian model

Let's code!