Evaluation of Screening and Diagnostic Tests

1

Ask

Marilyn®

BY MARILYN VOS SAVANT

A particularly interesting and important question today is that of testing for drugs. Suppose it is assumed that about 5% of the general population uses drugs. You employ a test that is 95% accurate, which we'll say means that if the individual is a user, the test will be positive 95% of the time, and if the individual is a nonuser, the test will be negative 95% of the time. A person is selected at random and is given the test. It's positive. What does such a result suggest? Would you conclude that the individual is a drug user? What is the probability that the person is a drug user?

2

Diagnostic Tests

- Important part of medical decision making
- In practice, many tests are used to obtain diagnoses
- There is a need to evaluate these tests (for accuracy and overall value)

3

Example

- Pap smear – screening test for cancer of the cervix
  - When a Pap smear is positive, abnormal cells are present approximately 95% of the time (positive predictive value).
  - When a Pap smear is negative, there can still be abnormal cells missed by the test. The test may miss abnormal cells 30% of the time.

4

Example

- Prostate cancer screening
  - Two primary tests
    - digital rectal examination (DRE)
    - prostate specific antigen (PSA) test
      - Positive Predictive Value = 20% for men > 50

5

Performance Characteristics

- **How do we evaluate a screening or diagnostic test?**
  - **How well does the test identify patients with a disease?**
  - **How well does the test identify patients without a disease?**
  - **Once a screening test result is obtained, what is the likelihood that it is correct?**

6

Why do we screen?

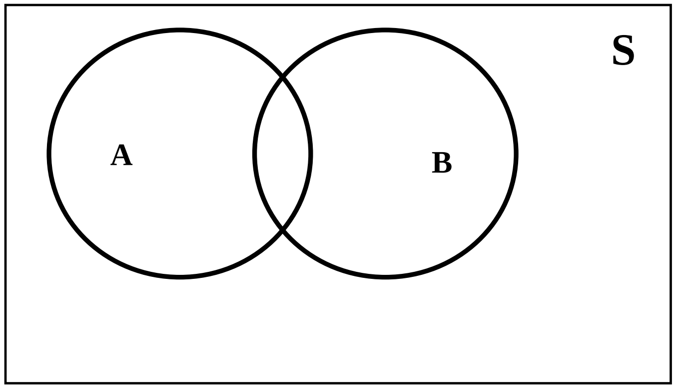
- Reduced cost
- **Resource efficiency**
- **Speed and timeliness of results**
- **Simplicity**
- **Adaptability to a variety of clinical settings**
- **Lack of invasiveness**
- **Acceptability in the population being evaluated**
- **Practicality to administer to many people**

7

Evaluation is Relative

- **The performance of a screening test should be taken in context of these factors.**
- Consideration of alternatives (e.g., other screening tests or performing a complete exam) should also be considered.

8



S=Sample space (totality of all events)

9

Probability

- The concept: If a trial (or experiment) is independently repeated a large number of times (N) and an outcome (A) occurs n times, then:
  - P(A) = n/N
  - Interpretation: if the trial is repeated again in the future, the likelihood that A will be the outcome again is n/N.

10

Introductory Probability

- Venn Diagrams
  - Draw them. Useful for thinking about probability
- Definitions
  - Intersection ($A \cap B$)
  - Union ($A \cup B$)
  - Complement (A^C)
  - ∅ A *null* event: Cannot happen
  - Mutually exclusive events implies no intersection
    - Example: ($A \cap A^C$) = ∅ **by definition**

11

Probability

- In general: $0 \le P(A) \le 1$
- Sample space: P(S)=1
- Impossible (null event): P(A)=0
- Sure thing: P(A)=1

12

Probability

- For any two events:
  - $P(A \cup B)=P(A)+P(B)-P(A \cap B)$
- If A and B are mutually exclusive, then $P(A \cap B)=0$, and thus
  - $P(A \cup B)=P(A)+P(B)$
- $P(A^C)=1-P(A)$

13

Conditional Probability

- The probability of an event (B) occurring conditional (or given) that event A occurs is:
  - $P(B|A)=[P(A \cap B)] / P(A)$
- Two event (A and B) are ***independent*** if $P(B|A)=P(B)$ (i.e, knowledge of A does not affect the probability of B)
  - Thus in this case, $P(A \cap B)=P(A)P(B)$
  - Example: drawing cards w/ and w/o replacement

14

Conditional Probability

- Sensitivity, specificity, and predictive values (which are all used to evaluate screening/diagnostic tests) are all conditional probabilities.
- Note: a p-value is a conditional probability too.

15

Evaluation of a Screening Test

- Give a group of people both tests (the screening test and the “gold standard” test (e.g., a full and complete exam that will diagnose disease without error) and then cross-classify the results.

16

		True Disease	
		+	-
Screening Test	+	a	c (false positive)
	-	b (false negative)	d

Performance Characteristics

- **Sensitivity = $P(T^+|D^+)$**
  = a/(a+b)
- **Specificity = $P(T^-|D^-)$**
  = d/(c+d)
- A perfect test would have b and c equal to 0.

Performance Characteristics

- **False positive rate = $P(T^+|D^-)$**
  = c/(c+d)
  = 1 – Specificity
- **False negative rate = $P(T^-|D^+)$**
  = b/(a+b)
  = 1 – Sensitivity

Predictive Values

- More informative from the patient or physician perspective
- Special applications of Bayes Theorem

Predictive Values

- **Positive Predictive Value = $P(D^+|T^+)$ =**

$$\frac{P(D^+)\,P(T^+ \mid D^+)}{P(D^+)\,P(T^+ \mid D^+) + P(D^-)\,P(T^+ \mid D^-)}$$

= a/(a+c) *if the prevalence of disease in the general population is the same as the prevalence of disease observed in the study*

21

Predictive Values

- **Negative Predictive Value = $P(D^-|T^-)$ =**

$$\frac{P(D^-)\,P(T^- \mid D^-)}{P(D^-)\,P(T^- \mid D^-) + P(D^+)\,P(T^- \mid D^+)}$$

= d/(b+d) *if the prevalence of disease in the general population is the same as the prevalence of disease observed in the study*

22

Answering Marilyn's Question

- Insert Marilyn_Example

23

Example: X-ray Screening for Tuberculosis

- Tuberculosis
  - infectious bacterial disease usually affecting the lungs
  - X-rays often used as a screening tool
  - PCR test can be performed to check for the presence of TB bacteria in the blood or other body fluids.

24

Example: X-ray Screening for Tuberculosis

- Insert TB_Example

25

Example: BPNS in NARC 007

- NARC 007 – Evaluation of the Brief Peripheral Neuropathy Screen (BPNS) vs. a comprehensive neurologic evaluation using the Total Neuropathy Score (TNS) for distal sensory-predominant neuropathy (DSPN)
- Insert neuropathy_example

26

ROC Curves

- **In some scenarios, the screening or diagnostic test is measured on a continuous scale and a "cutpoint" is used to categorize a patient as diseased or non-diseased.**

27

ROC Curves

- **For example, a test for a disease may be graded on a scale from 0 to 100. If the subject scores above 70, then the subject is characterized as diseased (i.e., a score below 70 indicates non-diseased).**

28

ROC Curves

- **In order to improve sensitivity, we may consider moving the cutpoint from 70 to 60. This will increase the number of "positives", and thus increase sensitivity.**

29

ROC Curves

- **Note, however there is a cost for this. The false positive rate (i.e., the probability that the test is positive when the subject is truly non-diseased) also increases. Thus specificity decreases.**
- **Thus, in general, as sensitivity increases, specificity decreases.**

30

ROC Curves

- **An ROC curve plots sensitivity versus (1-specificity) so that the trade-off can be visualized as the cutpoint varies. The original 2X2 table represents a single point on the ROC plot.**

31

Example: BNCS in NARC 007

- NARC 007 – Evaluation of the Brief Neuro-Cognitive Screen (BNCS) vs. a comprehensive neuropsychological exam for neurocognitive impairment
- Insert ROC_example

32