Screening, sensitivity and specificity of a diagnostic test, R.O.C. curves, Bayes’ theorem

Prof. Giuseppe Verlato
Unit of Epidemiology & Medical Statistics, Department of Diagnostics & Public Health, University of Verona

Natural history of a disease (Rothman, 1981)

- **Primary prevention**
  - Induction (exposure) phase
  - Latency period
  - Survey on exposure to risk factors in the target population (smoke, alcohol, cholesterol)
  - Health education

- **Secondary prevention**
  - Preclinical stage
  - Onset of early signs
  - Population screening (Pap test, fecal occult blood testing)
  - Early diagnosis and treatment

- **Tertiary prevention**
  - Clinical stage
  - Onset of clinical signs and symptoms
  - Hospital survey
  - Diagnosis of full-blown disease, rehabilitation

- Recovery, chronic disease, death

Beginning of exposure, beginning of pathologic process, Onset of early signs, Onset of clinical signs and symptoms, Recovery, chronic disease, death

Time
Primary, secondary and tertiary prevention

Before the disease onset:
Primary prevention = preventing or eliminating exposure to risk factors (for example, anti-smoking or anti-alcohol campaigns).

The disease has already established, but it is still at an early stage and clinically undetectable:
Secondary prevention = detecting disease cases at an early stage through a screening (for example, Pap smear test for cervical cancer, mammography for breast cancer, fecal occult blood testing for colon cancer).

The disease has become fully evident:
Tertiary prevention = Proper treatment and rehabilitation to prevent or soften the negative impact of the disease (for instance, care and rehabilitation of people with myocardial infarction).

Screening

1) Administering a non-invasive and non-expensive test
2) to large population strata at risk for a certain disease
3) to detect affected individuals, before the disease itself becomes apparent from a clinical point of view.

The aim of a screening program is to detect the disease at an early stage, when chances of recovery are still high.
Ideal situation for a screening test

<table>
<thead>
<tr>
<th></th>
<th>diseased</th>
<th>healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ test</td>
<td>a</td>
<td>-----</td>
</tr>
<tr>
<td>- test</td>
<td>-----</td>
<td>d</td>
</tr>
</tbody>
</table>

In the real world

<table>
<thead>
<tr>
<th></th>
<th>diseased</th>
<th>healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ test</td>
<td>a</td>
<td>False positives</td>
</tr>
<tr>
<td>- test</td>
<td>False negatives</td>
<td>d</td>
</tr>
</tbody>
</table>
Occult blood testing in stools

+ test  disease

Colonscopy + biopsy

+ test  disease

Screening for syphilis infection

+ test  syphilis

1\textsuperscript{st} step: false negatives are removed by VDRL test.

2\textsuperscript{nd} step: false positives are removed by Nelson test.
Sensitivity = $p(T+/M+) = \frac{a}{a+c}$

Specificity = $p(T-/M-) = \frac{d}{b+d}$

<table>
<thead>
<tr>
<th></th>
<th>Population at high risk</th>
<th>General population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M+</td>
<td>M-</td>
</tr>
<tr>
<td></td>
<td>291</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>693</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>700</td>
</tr>
</tbody>
</table>

Prevalence : $P(M+) = \frac{300}{1000}= 0.30$

Sensitivity : $p(T+/M+) = \frac{291}{300} = 0.97$

Specificity : $p(T-/M-) = \frac{693}{700} = 0.99$

$V_+ = P(M+/T+) = \frac{291}{298} = 0.977$

$V_- = P(M/T-) = \frac{693}{702} = 0.987$

$V_+ = $ positive predictive value

$V_- = $ negative predictive value
**SCREENING**

Population at high risk

<table>
<thead>
<tr>
<th></th>
<th>M+</th>
<th>M-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>291</td>
<td>7</td>
</tr>
<tr>
<td>T-</td>
<td>9</td>
<td>693</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M+</th>
<th>M-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>300</td>
<td>700</td>
</tr>
<tr>
<td>T-</td>
<td>9</td>
<td>700</td>
</tr>
</tbody>
</table>

General population

<table>
<thead>
<tr>
<th></th>
<th>M+</th>
<th>M-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>2910</td>
<td>12880</td>
</tr>
<tr>
<td>T-</td>
<td>90</td>
<td>987030</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>1000000</td>
</tr>
</tbody>
</table>

Prevalence \( P(M+) = \frac{300}{1000} = 0.30 \)

Sensitivity \( P(T+/M+) = \frac{291}{300} = 0.97 \)

Specificity \( P(T-/M-) = \frac{693}{700} = 0.99 \)

Positive predictive value (\( V+ \)) = \( p(M+/T+) = \frac{291}{298} = 0.977 \)

Negative predictive value (\( V- \)) = \( p(M-/T-) = \frac{693}{702} = 0.987 \)

\[ V+ = \frac{a}{a+b} \]

\[ V- = \frac{d}{c+d} \]

**Population at high risk**

**General population**

<table>
<thead>
<tr>
<th></th>
<th>M+</th>
<th>M-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>2910</td>
<td>12880</td>
</tr>
<tr>
<td>T-</td>
<td>90</td>
<td>987030</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>1000000</td>
</tr>
</tbody>
</table>

Prevalence \( P(M+) = \frac{300}{1000} = 0.30 \)

Sensitivity \( P(T+/M+) = \frac{291}{300} = 0.97 \)

Specificity \( P(T-/M-) = \frac{693}{700} = 0.99 \)

Positive predictive value (\( V+ \)) = \( p(M+/T+) = \frac{291}{298} = 0.977 \)

Negative predictive value (\( V- \)) = \( p(M-/T-) = \frac{693}{702} = 0.987 \)

\[ V+ = \frac{a}{a+b} \]

\[ V- = \frac{d}{c+d} \]
EXAMPLE: SCREENING for BREAST CANCER

Giorgi et al [2006] summarized the results of screening programs for breast cancer, performed in Italy in 2003-04: 7.8% of women undergoing their 1st mammography were referred for further examinations, while the percentage of women diagnosed with breast cancer was 0.65% in the overall population participating in screening programs [Giorgi et al, 2006].

Hence the positive predictive value of mammography was 0.65% / 7.8% = 0.083: in other words 1 in 12 women, referred for invasive diagnostic procedures, did have a malignancy. Positive predictive value is always rather low in screening programs on the general population.

Of course, it is fully acceptable that 11 healthy women could uselessly undergo invasive procedures, in order to detect and eliminate a malignancy at an early stage. However, “this value needs to be reasonably low, in order to limit the negative psychological impact (anxiety), the invasive procedure (cytology, core, or surgical biopsies), which may be required, as well as costs” (questo valore deve essere ragionevolmente basso, per limitare l’impatto psicologico negativo (ansietà), le procedure invasive indicate (citologia, prelievo dal centro del nodulo. o biopsie chirurgiche), come pure i costi) [Giorgi et al, 2006].


Other measures of diagnostic accuracy, mainly used in the clinical setting

Positive likelihood ratio (LR+)

Ratio between the probability of a POSITIVE test given the PRESENCE of the disease and the probability of a POSITIVE test given the ABSENCE of the disease:

\[ LR^+ = \frac{P(T+/M+)}{P(T+/M-)} = \frac{sensitivity}{1 - specificity} \]

Negative likelihood ratio (LR-)

Ratio between the probability of a NEGATIVE test given the PRESENCE of the disease and the probability of a NEGATIVE test given the ABSENCE of the disease:

\[ LR^- = \frac{P(T-/M+)}{P(T-/M-)} = \frac{1 - sensitivity}{specificity} \]
Cut-off for LR+ and LR-

When LR+ is greater than 5, a positive test will confidently confirm the presence of the disease
When LR- is lower than 0.2, a negative test will confidently exclude the disease

Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994 Mar 2;271(9):703-7.

Sometimes the diagnostic test is based on a CONTINUOUS variable. For instance:

- The fasting plasma glucose diagnostic threshold for diabetes is 7.0 mmol/l (126 mg/dl).
- The blood pressure threshold for defining hypertension is 140/90 mmHg.

DECISION LEVEL PLOT → to choose the optimal cut-off
R.O.C. CURVE → to evaluate the overall performance of the test over the entire range of possible cut-offs
Two well-separated distributions

Non-diabetic adults

Diabetic adults

Probabilty density

glycaemia (mg/100 ml)

What about the elderly?

Two partly superimposed distributions

Elderly without diabetes

Elderly with diabetes

Probabilty density

glycaemia (mg/100 ml)
### 1st EXAMPLE:
**ADULTS with or without diabetes**

<table>
<thead>
<tr>
<th>specificity</th>
<th>1-specificity</th>
<th>sensitivity</th>
<th>CUT-OFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.0 %</td>
<td>50.0 %</td>
<td>99.997 %</td>
<td>100 mg/dl</td>
</tr>
<tr>
<td>84.1 %</td>
<td>15.9 %</td>
<td>99.9 %</td>
<td>115 mg/dl</td>
</tr>
<tr>
<td>97.7 %</td>
<td>2.3 %</td>
<td>97.7 %</td>
<td>130 mg/dl</td>
</tr>
<tr>
<td>99.9 %</td>
<td>0.1 %</td>
<td>84.1 %</td>
<td>145 mg/dl</td>
</tr>
<tr>
<td>99.997 %</td>
<td>0.003 %</td>
<td>50.0 %</td>
<td>160 mg/dl</td>
</tr>
</tbody>
</table>

|                  |               |             |         |
| ---              | ---           | ---         | 175 mg/dl |

### 2nd EXAMPLE:
**ELDERLY with or without diabetes**

<table>
<thead>
<tr>
<th>specificity</th>
<th>1-specificity</th>
<th>sensitivity</th>
<th>CUT-OFF</th>
<th>specificity</th>
<th>1-specificity</th>
<th>sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3 %</td>
<td>97.7 %</td>
<td>99.997 %</td>
<td>100 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.9 %</td>
<td>84.1 %</td>
<td>99.9 %</td>
<td>115 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.0 %</td>
<td>50.0 %</td>
<td>99.9 %</td>
<td>130 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84.1 %</td>
<td>15.9 %</td>
<td>84.1 %</td>
<td>145 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.7 %</td>
<td>2.3 %</td>
<td>50.0 %</td>
<td>160 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.9 %</td>
<td>0.1 %</td>
<td>84.1 %</td>
<td>175 mg/dl</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**R.O.C. = Receiver Operating Characteristic**
DECISION LEVEL PLOT

Youden index
(sensitivity + specificity - 1)

Elderly without diabetes
Elderly with diabetes

Decision level (mg/100 ml)

Probability density
glycaemia

Elderly without diabetes
Elderly with diabetes

Non-diabetic adults
adults with diabetes

DECISION LEVEL PLOT

Youden index
(sensitivity + specificity - 1)

Non-diabetic adults
adults with diabetes

Elderly without diabetes
Elderly with diabetes

DECISION LEVEL PLOT

Youden index
(sensitivity + specificity - 1)

Non-diabetic adults
adults with diabetes

Elderly without diabetes
Elderly with diabetes

E NEGLI ANZIANI?
R.O.C. (Receiver Operating Characteristic) Curves

- Sensitivity (%)
- Specificity (%)

1 - specificity (%)

Adults with/without diabetes
Elderly with/without diabetes

Verlato, 1998

- Sensitivity
- Specificity
- Youden index

Rare disease, or disease curable only if diagnosed early

Estimating the prevalence of a disease

Verlato, 1998
BAYES’ THEOREM and its application to DIFFERENTIAL DIAGNOSIS

1st EXAMPLE: from the world of fairy tales

2nd EXAMPLE: clinical examples
Assumptions:

1. \( P(\text{Wolf } \cap \text{ Strawberries}) = P(\text{Wolf } \cap \text{ Peter}) = P(\text{Peter } \cap \text{ Strawberries}) = 0 \)

2. \( P(\text{bitten by wolf}) + P(\text{picking strawberries}) + P(\text{wandering with Peter}) = 1 \)
**Bayes’ theorem**  
*(Thomas Bayes 1702 - 1761)*

We know the **effect**, we have a list of **possible causes** and we want to assign to each cause the probability to have produced the effect.

In medicine a patient reports a **symptom** to a doctor, who has to finding the **disease** causing this symptom among a list of possible diseases.

---

**Clinical application of Bayes’ theorem**  
**Clinical case: Hematuria in 25 years-old man**

<table>
<thead>
<tr>
<th></th>
<th>Kidney stone</th>
<th>Glomerulonephritis</th>
<th>Cancer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a priori probability− p(D)</td>
<td>0.1%</td>
<td>0.5%</td>
<td>0.01%</td>
<td>------</td>
</tr>
<tr>
<td>conditional prob. - p(S/D)</td>
<td>50%</td>
<td>80%</td>
<td>60%</td>
<td>------</td>
</tr>
<tr>
<td>probability product</td>
<td>5/10,000</td>
<td>40/10,000</td>
<td>0.6/10,000</td>
<td>45.6/10,000</td>
</tr>
<tr>
<td>posterior probability (D/S)</td>
<td>5/45.6</td>
<td>40/45.6</td>
<td>0.6/45.6</td>
<td>45.6/45.6</td>
</tr>
<tr>
<td></td>
<td>11.0%</td>
<td>87.7%</td>
<td>1.3%</td>
<td>100%</td>
</tr>
</tbody>
</table>

\[
p(D_1/S) = \frac{p(D_1) * p(S/D_1)}{p(D_1) * p(S/D_1) + p(D_2) * p(S/D_2) + p(D_3) * p(S/D_3)}
\]

**ASSUMPTIONS**

1) There are only three diseases (kidney stone, glomerulonephritis, cancer) causing hematuria

2) The three diseases are mutually exclusive
Bayes’ formula

It conveniently displays the single steps of diagnostic procedure, showing how probabilities initially attributed to different causes (diseases) are subsequently modified by newly collected information (symptoms).

\[
P(D_i \mid S) = \frac{P(D_i) \cdot P(S \mid D_i)}{\sum_{i=1}^{k} P(D_i) \cdot P(S \mid D_i)}
\]

where:

- \(D_1, \ldots, D_i, \ldots, D_k\) ⇒ possible causes of the symptom under study
- \(S\) ⇒ symptom/sign under study

**P(D_i)** is the *a priori probability* of the cause \(D_i\): it can be viewed as the probability that a physician assigns to a given disease BEFORE visiting the patient, according to disease occurrence (incidence/prevalence).

**P(S\mid D_i)** is the *conditional probability*: the probability of the symptom given the disease \(D_i\).

**P(D_i\mid S)** is the *posterior probability*: it measures the probability that the event \(S\), already occurred, could be attributed to the cause \(D_i\), among a finite set \(k\) of possible causes. In the clinical setting it represents the new probability that the physician assigns to the disease AFTER having visited the patient.

Computation of posterior probabilities ⇒ DIFFERENTIAL DIAGNOSIS
Clinical application of Bayes’ theorem
Clinical case: Hemoptysis in a 40 years-old man

<table>
<thead>
<tr>
<th></th>
<th>TBC</th>
<th>Lung cancer</th>
<th>Pneumonia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a priori probability – p(D)</td>
<td>0.01%</td>
<td>0.1%</td>
<td>1%</td>
<td>------</td>
</tr>
<tr>
<td>conditional prob. - p(S/D)</td>
<td>80%</td>
<td>40%</td>
<td>2%</td>
<td>------</td>
</tr>
<tr>
<td>probability product</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>posterior probability (D/S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ASSUMPTIONS

1) There are only three diseases (TBC, lung cancer, pneumonia) causing hemoptysis
2) The three diseases are mutually exclusive
Application of the Bayes’ theorem in clinical practice

Bayes’ theorem has not been extensively applied in clinical practice, as its assumptions are not met.

1. It is seldom possible to identify all possible diseases which could cause a certain symptom/sign.

2. Diseases often simultaneously occur in the same subjects (comorbidities, multimorbidities, syndromes).