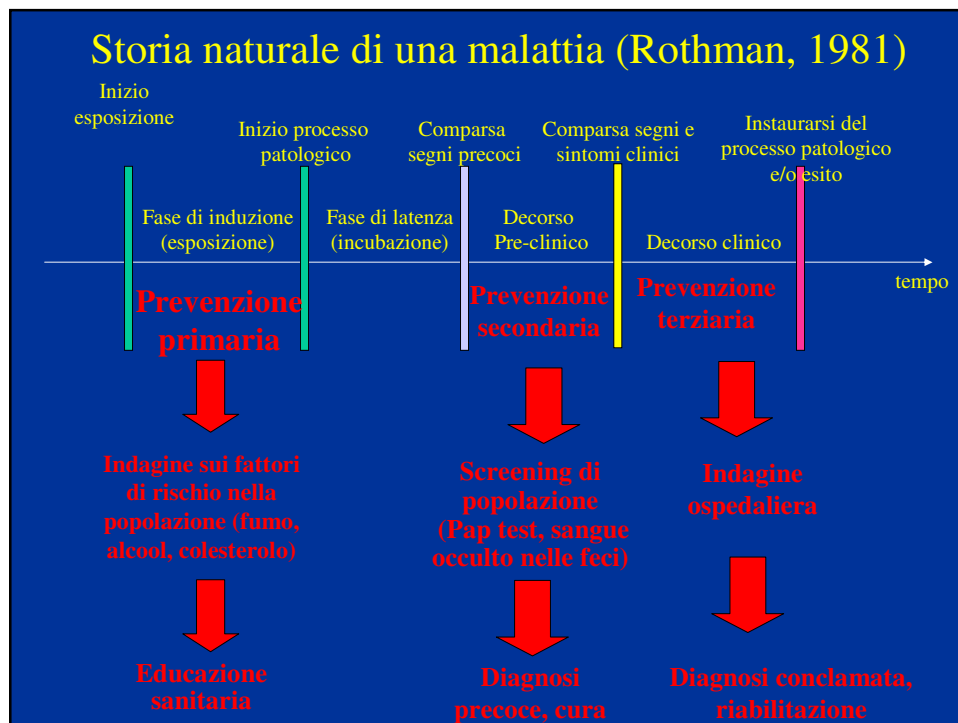


# Screening, sensibilità e specificità di un test diagnostico, curve R.O.C., teorema di Bayes

- Sezione di Epidemiologia e Statistica Medica, Università di Verona



## **Prevenzione primaria, secondaria e terziaria**

### **Prima che si instauri la malattia:**

Prevenzione primaria = Rimozione dei fattori di rischio (ad esempio, campagne contro il fumo o contro l'alcoolismo).

### **La malattia si è instaurata, ma non è ancora evidente dal punto di vista clinico:**

Prevenzione secondaria = Individuazione precoce dei casi tramite uno **screening** (ad esempio, Pap test per il tumore dell'utero, mammografia per il tumore del seno, sangue occulto nelle feci per il tumore del colon).

### **La malattia si è manifestata clinicamente:**

Prevenzione terziaria = Terapia appropriata e riabilitazione per prevenire o ridurre le conseguenze negative della malattia stessa (ad esempio, assistenza agli infartuati e riabilitazione).

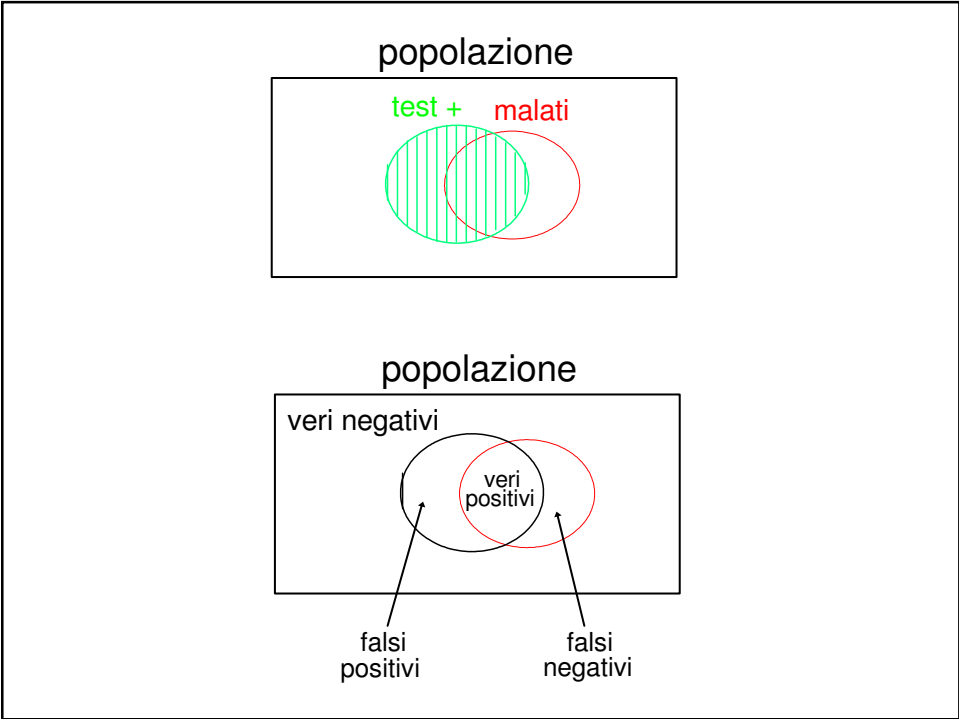
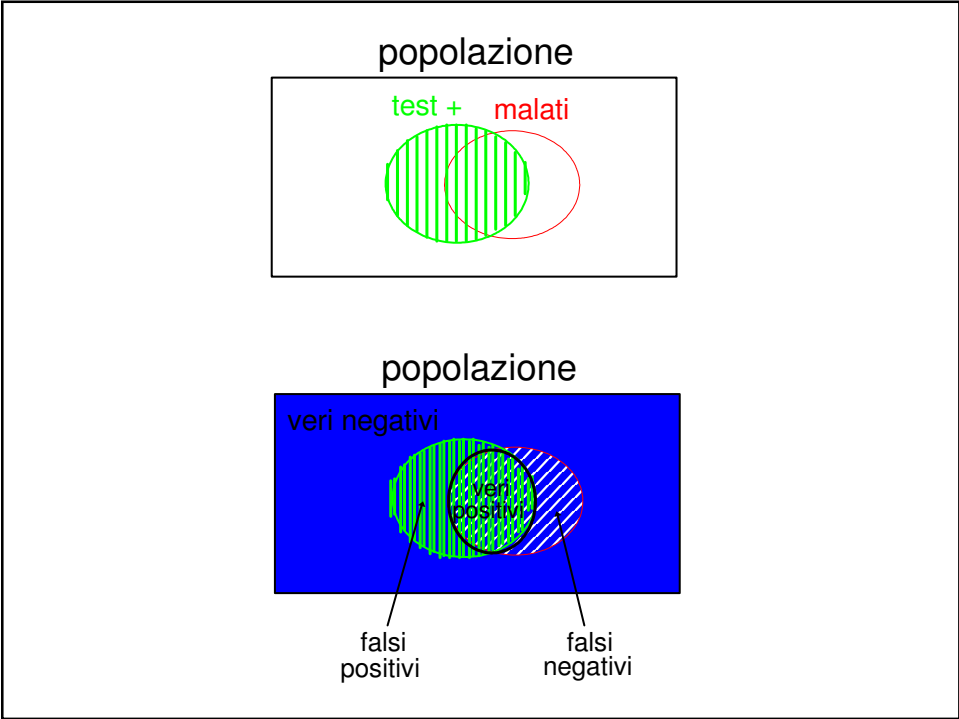
## **Screening**

1) Somministrazione di un test diagnostico poco costoso e poco invasivo

2) a larghi settori della popolazione a rischio per una determinata patologia

3) per identificare gli individui ammalati prima che la malattia si riveli dal punto di vista clinico.

Lo scopo dello screening è diagnosticare precocemente la malattia, quando è ancora curabile.



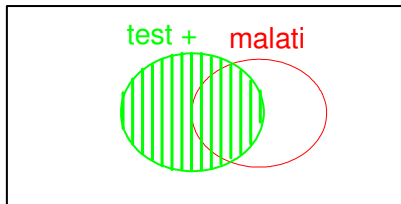
**Situazione ideale in un test di screening**

	<b>malati</b>	<b>sani</b>	
<b>Test +</b>	<b>a</b>	-----	
<b>Test -</b>	-----	<b>d</b>	

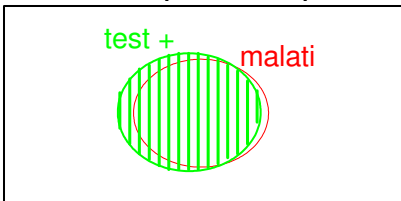
**Nella cruda realtà**

	<b>malati</b>	<b>sani</b>	
<b>Test +</b>	<b>a</b>	<b>Falsi positivi</b>	
<b>Test -</b>	<b>Falsi negativi</b>	<b>d</b>	

sangue occulto nelle feci



colonscopia + biopsia



Verlato, 1998

<b>Sensibilità = <math>p(T+/M+)</math> = probabilità che il test sia positivo nei malati</b>		
	<b>malati</b>	
<b>Test +</b>	<b>a</b>	
<b>Test -</b>	<b>c</b>	
	<b>a+c</b>	
		<b>Sens = <math>a/(a+c)</math></b>
<b>Specificità = <math>p(T-/M-)</math> = probabilità che il test sia negativo nei sani</b>		
		<b>sani</b>
<b>Test +</b>		<b>b</b>
<b>Test -</b>		<b>d</b>
		<b>b+d</b>
		<b>Spec = <math>d/(b+d)</math></b>

SCREENING					
	Popolazione a rischio			Popolazione generale	
	M+	M-		M+	M-
T+	291	7		2910	9970
T-	9	693		90	987030
	300	700		3000	997000
	1000			1000000	
Prevalenza = $P(M+)$ =	$300/1000 = 0.30$			_____	
Sensibilità = $P(T+/M+)$ =	$291/300 = 0.97$			_____	
Specificità = $P(T-/M-)$ =	$693/700 = 0.99$			_____	
$V+ = P(M+/T+)$ =	$291/298 = 0.977$			_____	
$V- = P(M-/T-)$ =	$693/702 = 0.987$			_____	
V+ = Valore predittivo dei positivi V- = Valore predittivo dei negativi					

## SCREENING

Popolazione a rischio			Popolazione generale			
	M+	M-		M+	M-	
T+	291	7	298	2910	9970	12880
T-	9	693	702	90	987030	987120
	300	700	1000	3000	997000	1000000

Prevalenza = P(M+) =	300/1000 = 0.30	<b>3000 / 1 000 000 = 0,003 = 0,3%</b>
Sensibilità = P(T+/M+) =	291/300 = 0.97	<b>2910 / 3000 = 0,97 = 97%</b>
Specificità = P(T-/M-) =	693/700 = 0.99	<b>987030 / 997000 = 0,99 = 99%</b>
V+ = P(M+/T+) =	291/298 = 0.977	<b>2910 / 12880 = 0,226 = 22,6%</b>
V- = P(M-/T-) =	693/702 = 0.987	<b>987030 / 987120 = 0,9999 = 99,99%</b>

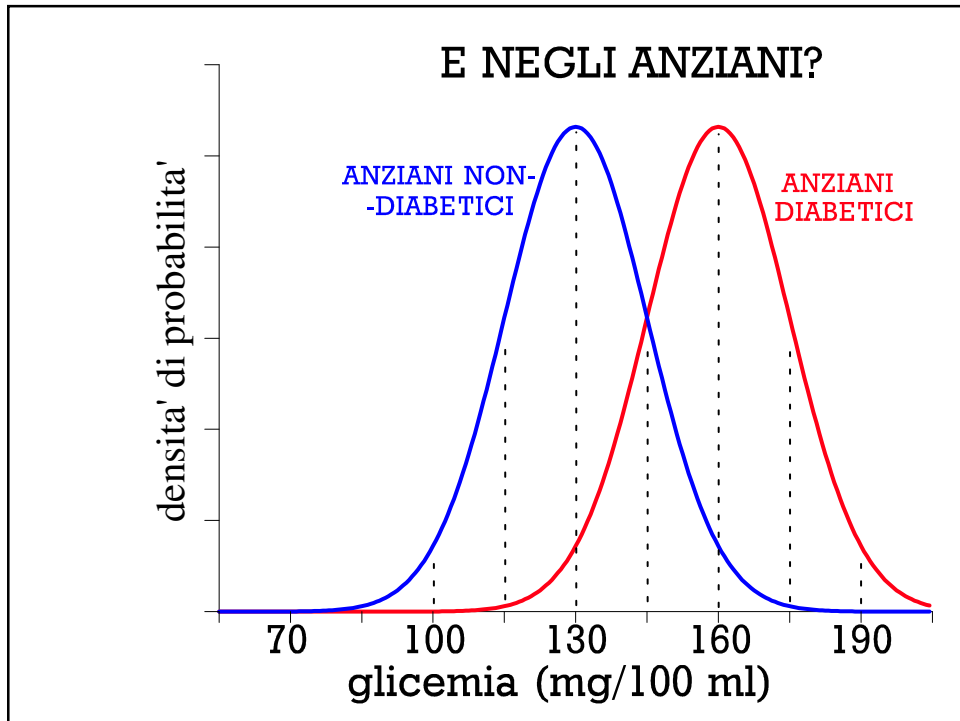
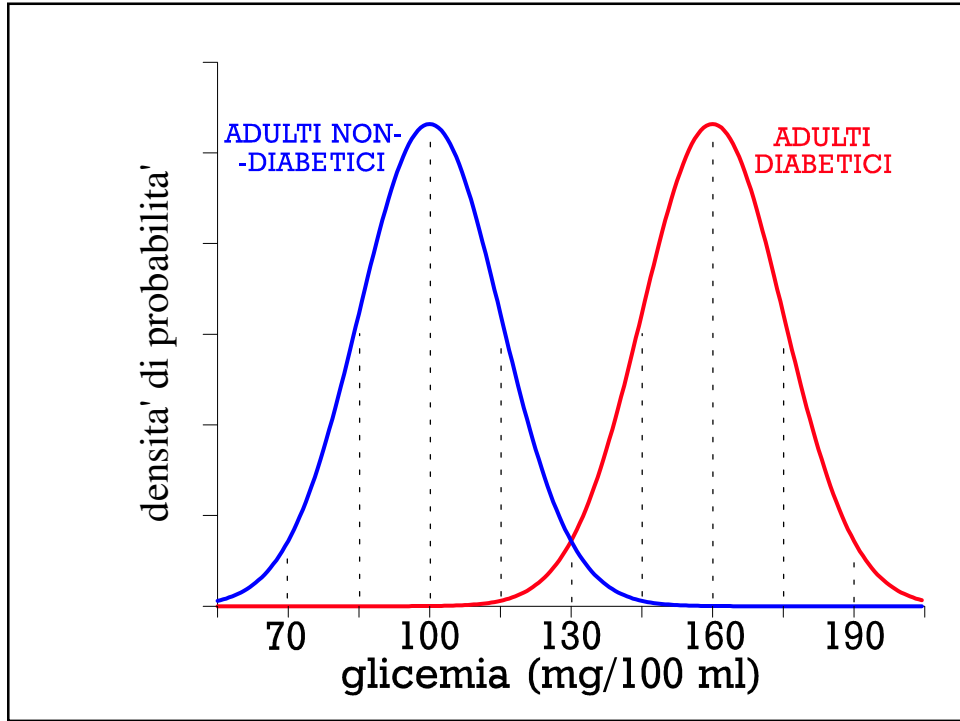
V+ = Valore predittivo dei positivi  
V- = Valore predittivo dei negativi

**Valore predittivo dei positivi (V+) =  $p(M+/T+) =$  probabilità che chi ha il test positivo sia malato**

	malati	sani	
<b>Test +</b>	<b>a</b>	<b>b</b>	<b>V+ = a/(a+b)</b>
<b>Test -</b>			

**Valore predittivo dei negativi (V-) =  $p(M-/T-) =$  probabilità che chi ha il test negativo sia sano**

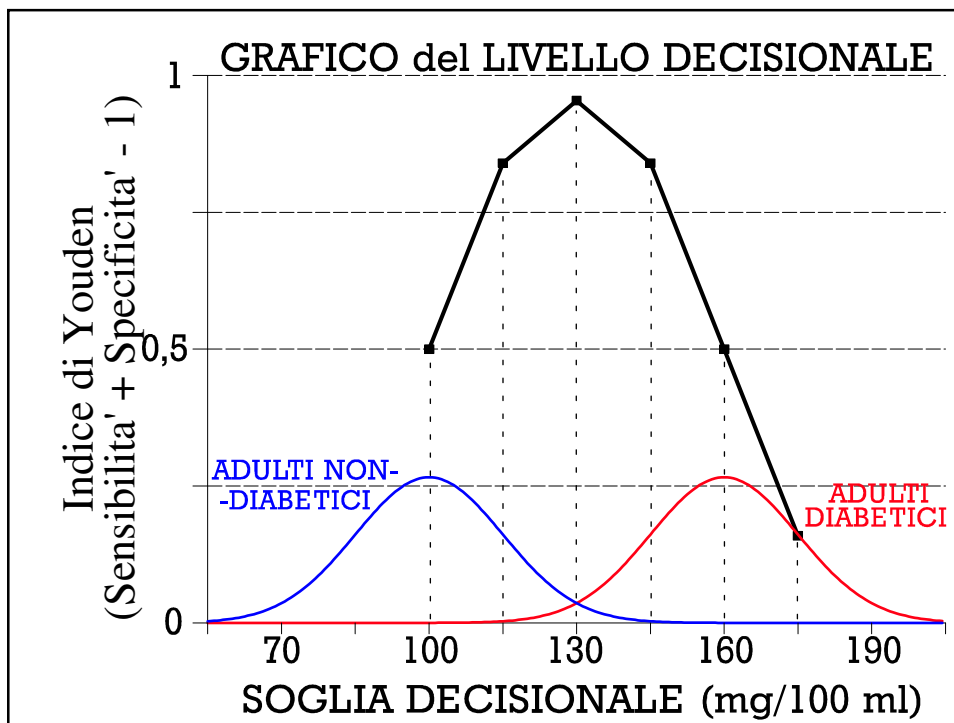
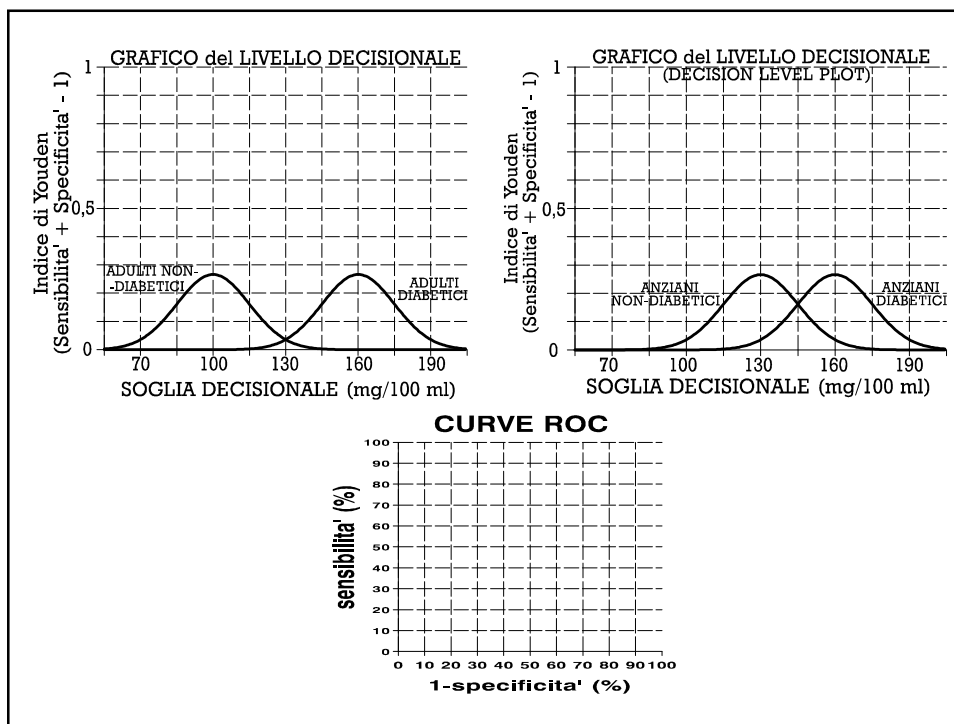
	malati	sani	
<b>Test +</b>			<b>V- = d/(c+d)</b>
<b>Test -</b>	<b>c</b>	<b>d</b>	

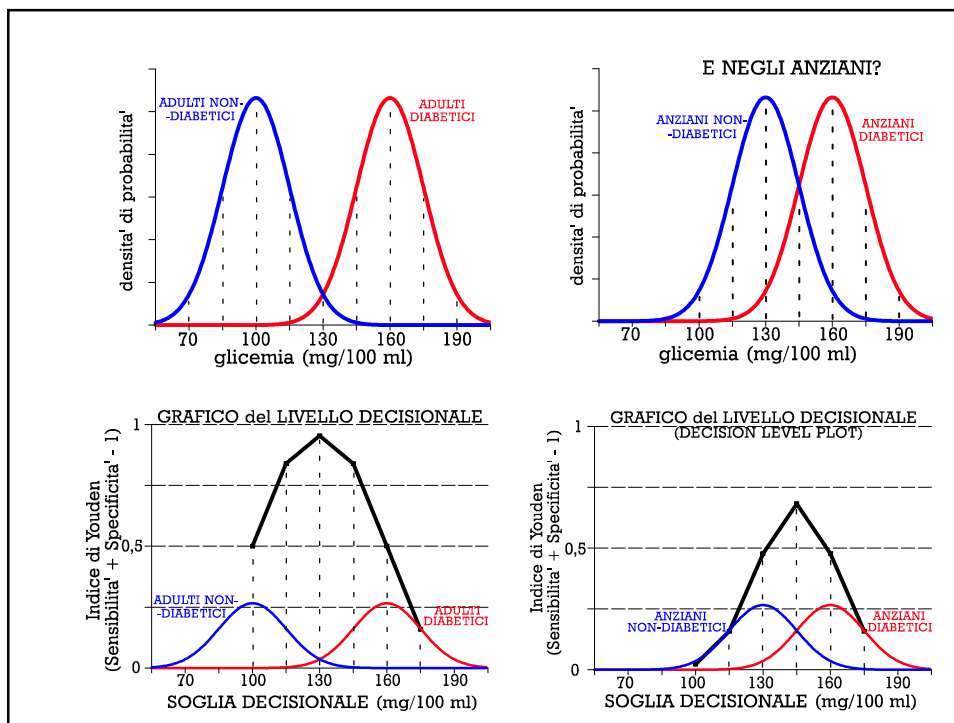
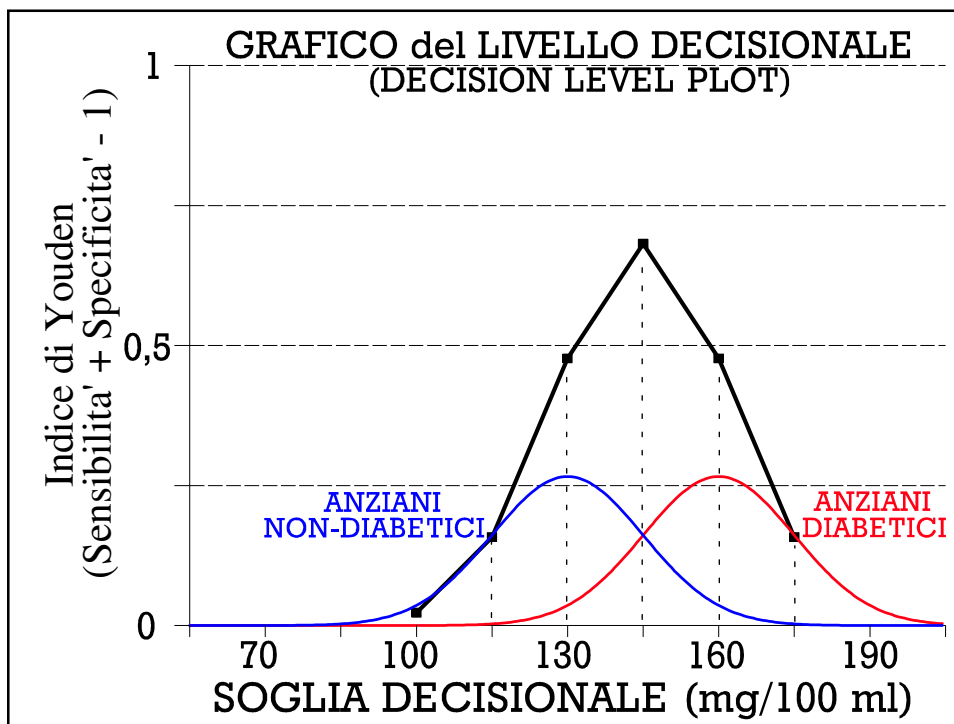


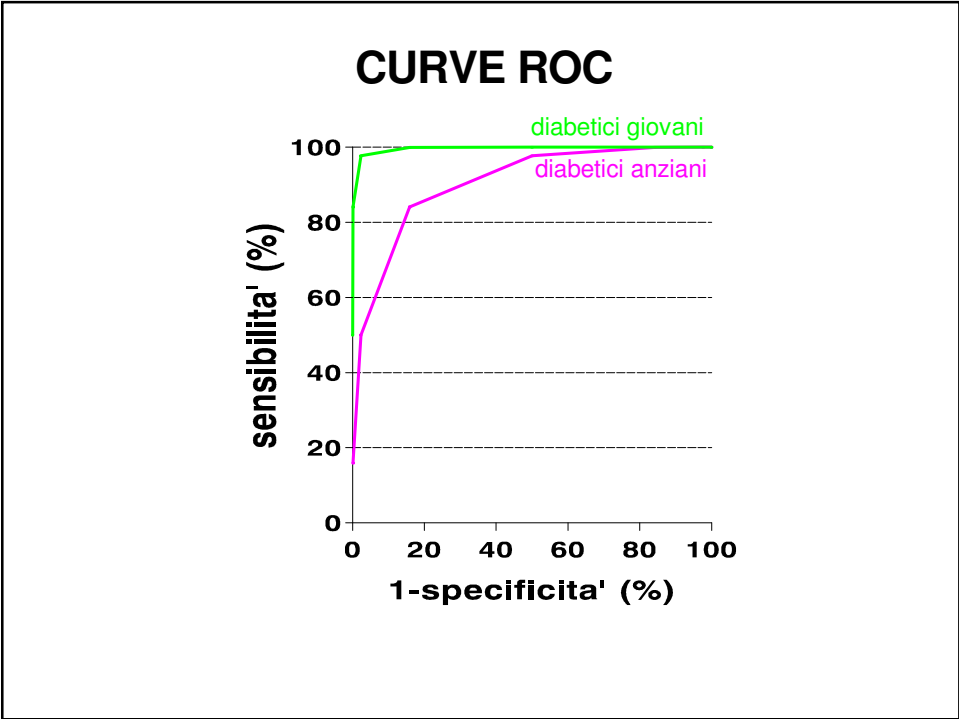
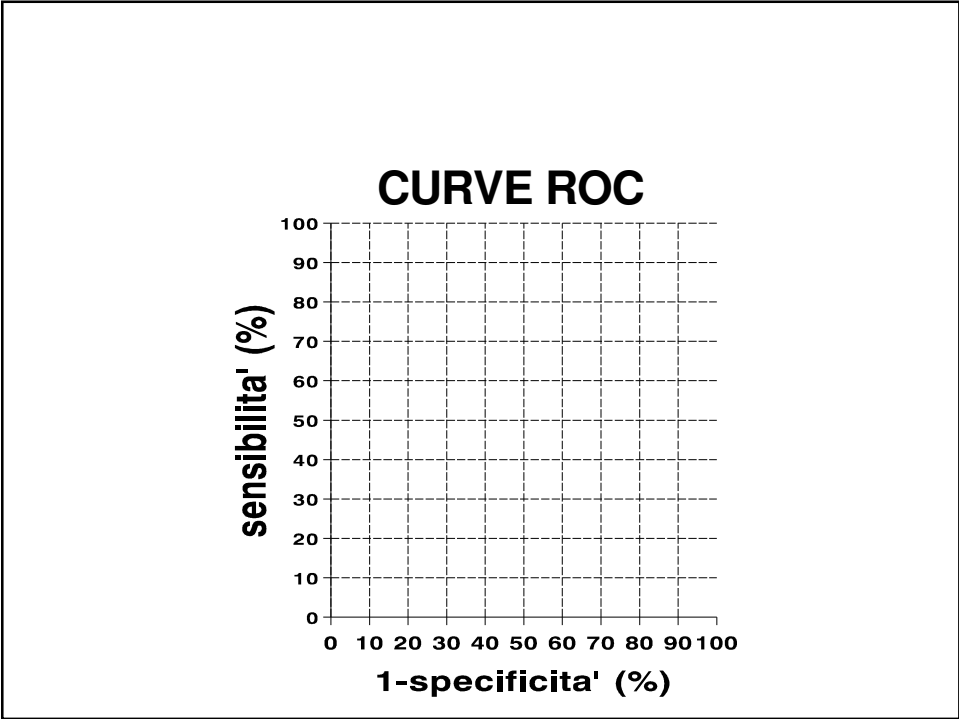
I ESEMPIO (PAZIENTI DIABETICI)				II ESEMPIO (PAZIENTI DIABETICI ANZIANI)		
specificità	1-specificità	sensibilità	LIVELLO DECISIONALE	specificità	1-specificità	sensibilità
			100 mg/dl			
			115 mg/dl			
			130 mg/dl			
			145 mg/dl			
			160 mg/dl			
---	---	---	175 mg/dl			
utilizzati per rappresentare il LIVELLO DECISIONALE e le CURVE ROC				utilizzati per rappresentare il LIVELLO DECISIONALE e le CURVE ROC		

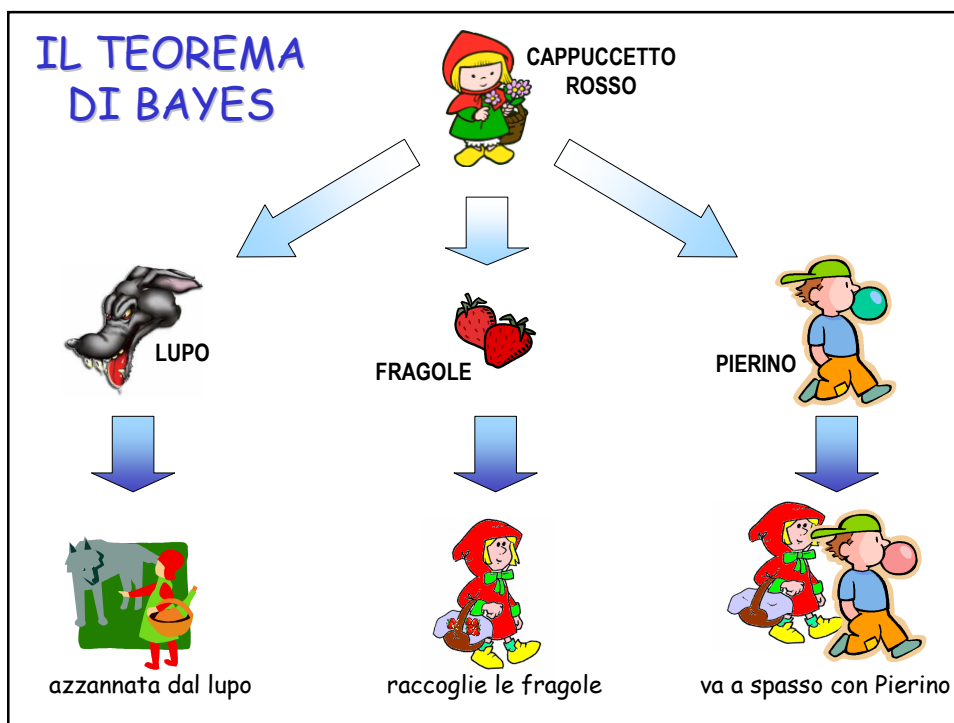
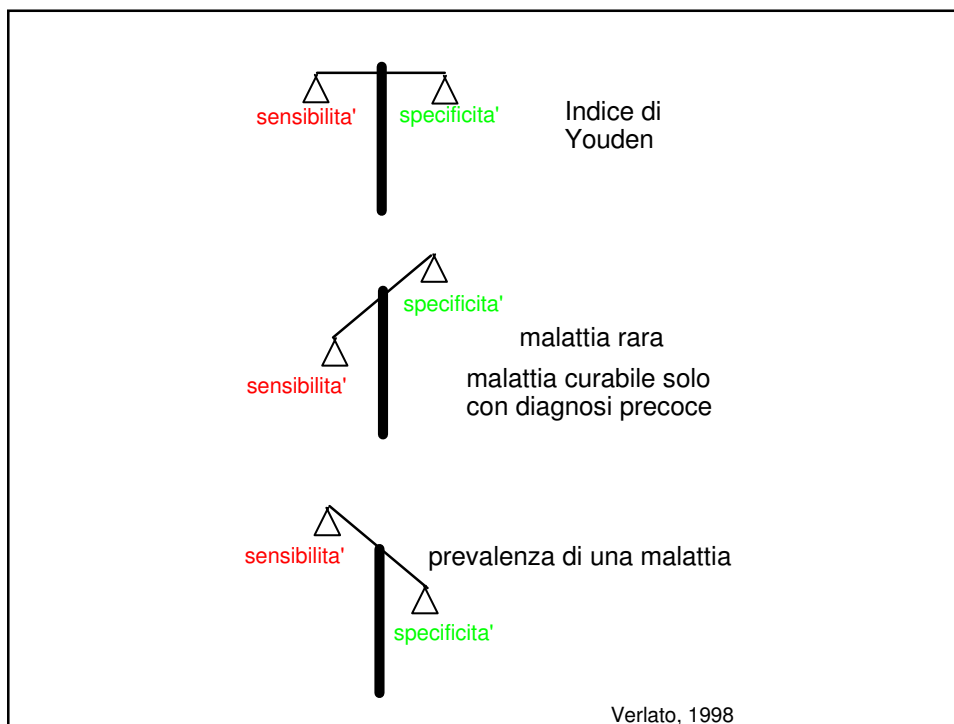
I ESEMPIO (PAZIENTI DIABETICI)				II ESEMPIO (PAZIENTI DIABETICI ANZIANI)		
specificità	1-specificità	sensibilità	LIVELLO DECISIONALE	specificità	1-specificità	sensibilità
50.0 %	50.0 %	99.997 %	100 mg/dl	2.3 %	97.7 %	99.997 %
84.1 %	15.9 %	99.9 %	115 mg/dl	15.9 %	84.1 %	99.9 %
97.7 %	2.3 %	97.7 %	130 mg/dl	50.0 %	50.0 %	97.7 %
99.9 %	0.1 %	84.1 %	145 mg/dl	84.1 %	15.9 %	84.1 %
99.997 %	0.003 %	50.0 %	160 mg/dl	97.7 %	2.3 %	50.0 %
---	---	---	175 mg/dl	99.9 %	0.1 %	15.9 %
utilizzati per rappresentare il LIVELLO DECISIONALE e le CURVE ROC				utilizzati per rappresentare il LIVELLO DECISIONALE e le CURVE ROC		







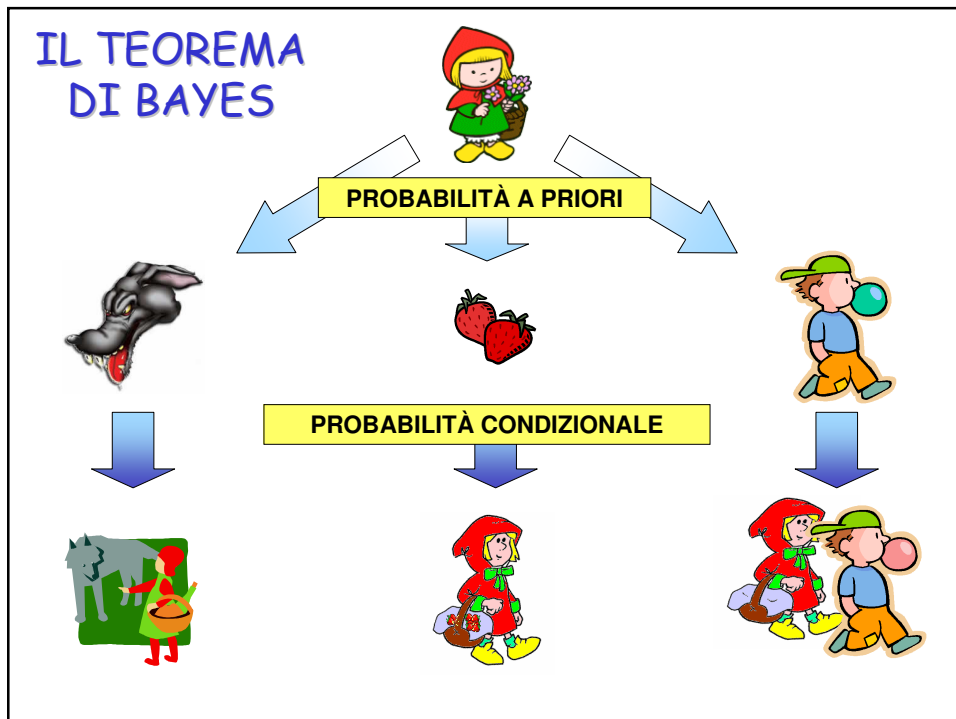




## Assunzioni:

1.  $P(\text{Lupo} \cap \text{Fragole}) = P(\text{Lupo} \cap \text{Pierino}) = P(\text{Pierino} \cap \text{Fragole}) = 0$

2.  $P(\text{azzannata dal lupo}) + P(\text{raccolta fragole}) + P(\text{a spasso con Pier.}) = 1$



## Teorema di Bayes

(Thomas Bayes 1702 - 1761)

Si conosce il **risultato** di un *esperimento* e si vuole conoscere la probabilità che esso sia dovuto ad una determinata **'causa'**

In medicina raramente esiste una relazione univoca tra **causa** (patologia) ed **effetti** (sintomi - *outcome*)

Spesso si hanno difficoltà ad individuare tutte le *possibili cause* che contribuiscono a 'determinare' un *outcome* e quindi le probabilità *a priori* ad esse associate

## Formula di Bayes

Permette di formalizzare i singoli passaggi del *procedimento diagnostico* e di modificare l'attribuzione iniziale di probabilità a delle **cause** (malattie) sulla base delle **conoscenze acquisite** (sintomi)

$$P(H_i | A) = \frac{P(H_i) \cdot P(A | H_i)}{\sum_{i=1}^k P(H_i) \cdot P(A | H_i)}$$

dove:

$H_1 \dots H_i \dots H_k \Rightarrow$  *possibili cause del sintomo in questione*

$A \Rightarrow$  *sintomo/malessere in studio*

$P(H_i)$  è nota come la **probabilità a priori** della causa  $H_i$ : può essere intesa come la **probabilità** che un medico può attribuire ad una *determinata patologia*, PRIMA di guardare il malato sulla base della diffusione della malattia (incidenza/prevalenza)

$P(A|H_i)$  è la **probabilità condizionale**: la probabilità del sintomo data la malattia  $H_i$ . Il sintomo è patognomonico della malattia?

$P(H_i|A)$  è la **probabilità a posteriori**: misura la probabilità che l'evento A, già realizzatosi, sia dovuto alla causa  $H_i$ , tra una serie finita k di cause possibili (i.e. rappresenta la probabilità che il medico attribuisce alla malattia, dopo aver acquisito le informazioni sui sintomi del paziente)

Calcolo delle probabilità a posteriori  $\Rightarrow$  DIAGNOSTICA DIFFERENZIALE

### Applicazione del teorema di Bayes Caso clinico: Ematuria in maschio di 25 anni

	Calcolo	Glomerulonefrite	Neoplasia	Totale
probabilità a priori – p(M)	0,1%	0,5%	0,01%	-----
prob. condizionale - p(S/M)	50%	80%	60%	-----
prodotto di probabilità	5 / 10000	40 / 10000	0,6 / 10000	45,6/100000
prob. a posteriori – p(M/S)	5 / 45,6 11,0%	40 / 45,6 87,7%	0,6 / 45,6 1,3%	45,6 / 45,6 100%

$$p(M_1/S) = \frac{p(M_1) * p(S/M_1)}{p(M_1) * p(S/M_1) + p(M_2) * p(S/M_2) + p(M_3) * p(S/M_3)}$$

**Assunzioni**

- 1) L'ematuria può avere solo 3 cause (calcolo, glomerulonefrite, neoplasia)
- 2) Le 3 cause sono mutuamente esclusive

### ***Esempio illustrativo***

*Un ginecologo riscontra anomalie all'utero di una paziente:*

- displasia al collo dell'utero ( $H_1$ )
  - stadio iniziale di tumore al collo dell'utero ( $H_2$ )
- } **Possibili cause**

*Dalla letteratura medica e per esperienza il medico sa che:*

$$P(H_1) = 0.85, P(H_2) = 0.15$$

*Il medico chiede se la paziente ha avuto perdite continue (**A**) nelle ultime settimane. Si sa che:*

$$P(A|H_1) = 0.10$$

$$P(A|H_2) = 0.80$$

$$P(H_1 | A) = \frac{0.85 \cdot 0.10}{0.85 \cdot 0.10 + 0.15 \cdot 0.80} = 0.40$$

$$P(H_2 | A) = \frac{0.15 \cdot 0.80}{0.85 \cdot 0.10 + 0.15 \cdot 0.80} = 0.60$$

12/12/2005

### **APPLICAZIONE del TEOREMA di BAYES CASO CLINICO: Emoftoe in maschio di 40 anni**

	TBC	Cancro polmonare	Bronco-polmonite	Totale
Prob. a priori Prevalenza= $p(M+)$	0.01 %	0.1 %	1.0 %	---
Prob. condizionale $p(S+/M+)$	80.0 %	40.0 %	2.0 %	---
Prodotto delle prob.				
Prob. a posteriori $p(M+/S+)$				
Prob. a posteriori $p(M+/S+)$				

ASSUNZIONI



## APPLICAZIONE del TEOREMA di BAYES CASO CLINICO: Emoftoe in maschio di 40 anni

	TBC	Cancro polmonare	Bronco-polmonite	Totale
Prob. a priori Prevalenza= $p(M+)$	0.01 %	0.1 %	1.0 %	---
Prob. condizionale $p(S+/M+)$	80.0 %	40.0 %	2.0 %	---
Prodotto delle prob.	0.8/10000	4.0/10000	2.0/10000	6.8/10000
Prob. a posteriori $p(M+/S+)$	0.8/6.8	4.0/6.8	2.0/6.8	6.8/6.8
Prob. a posteriori $p(M+/S+)$	11.8 %	58.8 %	29.4 %	100.0 %

### ASSUNZIONI

- 1) L'emoftoe puo' avere solo 3 cause (TBC, Ca polmonare, Broncopolmonite)
- 2) Le 3 cause sono mutuamente esclusive

## Perdite di sangue dal corpo umano

### **Sangue dall'apparato digerente:**

ematemesi = sangue vomitato

melena = sangue digerito (nero) nelle feci

rettorragia = sangue non digerito (rosso vivo) nelle feci

### **Sangue dall'albero respiratorio:**

emoftoe o emottisi

### **Sangue nell'urina:**

ematuria

### **Sangue dalle vie genitali femminili:**

menorragia = perdita mestruale abbondante

metrorragia = perdita non nel periodo delle mestruazioni