

Valutazione di un test

(specificità, sensibilità, valore predittivo, likelihood ratio)

Ogni test diagnostico binario, cioè positivo o negativo, ha una quota di errore: rispettivamente falsi positivi e falsi negativi.

Tradizionalmente, per la valutazione di un test in medicina, si usano gli indici “*specificità*” e “*sensibilità*”.

I due termini ***specificità*** e ***sensibilità***, sono indici statistici che indicano la proporzione correttamente classificata rispettivamente di negativi e positivi. (1)



Aspetti complementari

*Questa famosa illusione ottica, conosciuta come coppa di Rubin, presenta due **immagini complementari**: in bianco è possibile vedere la coppa, in nero due volti di profilo.*

Per valutare la validità di un test si preferisce far riferimento alla quota di errore piuttosto che a quella di corretta classificazione.

La quota di errore è espressa indirettamente dai **complementi** di *specificità* e *sensibilità* che indicano rispettivamente i **falsi positivi** e **falsi negativi**:

complemento specificità = **1-Sp** = % falsi positivi

complemento sensibilità = **1-Sn** = % falsi negativi.

La conoscenza dei margini di errore è indispensabile per l'appropriata utilizzazione di un test ma non è sufficiente in quanto non tiene conto della probabilità a priori, cioè della prevalenza della malattia nel campione esaminato.

Tabella di contingenza

I risultati di un classificatore binario, come un test diagnostico o uno studio clinico, sono inquadrati in una **tabella di contingenza**, 2 x 2.

Da questa tabella sono ricavati gli indici: *specificità*, *sensibilità*, *valore predittivo*, *likelihood ratio*.

La comprensione di una tabella di contingenza, nonostante la sua semplicità, spesso non è così immediata, ma è fondamentale per la valutazione di un test diagnostico.

I valori di ingresso, con i risultati del test, sono solo quattro, inseriti nel riquadro giallo.

Gli indici calcolati sono riportati nei riquadri rossi.

Legenda

D+ = malati; D- = sani

T+ = test positivo; T- = test negativo

Vp = Veri positivi; Fp = Falsi positivi; Vn = Veri negativi; Fn = Falsi negativi.

Sn = *sensibilità*; Sp = *specificità*; Vpp = *valore predittivo positivo*; Vpn = *valore predittivo negativo*.

prevalenza = probabilità pre test

LR = likelihood ratio o fattore Bayes

	D+	D-	tot	val predittivo	LR
T +	Vp	Fp <i>errore α</i>	tot test+	Vpp: $Vp / (Vp+Fp)$	LR + $Sn / (1-Sp)$
T -	Fn <i>errore β</i>	Vn	tot test -	Vpn: $Vn / (Vn+Fn)$	LR - $Sp / (1-Sn)$
tot	tot malati	tot sani	totale	prevalenza: tot malati /totale	
	Sn: $Vp / (Vp+Fn)$	Sp: $Vn / (Fp+Vn)$			

Gli indici *specificità*, *sensibilità* e *valore predittivo*, utilizzano **denominatori differenti**:

Specificità e sensibilità: proporzioni rispettivamente **tra individui sani e malati**.

Valori predittivi: proporzioni rispettivamente **tra test positivi e negativi**.

Likelihood ratio: rapporto tra due proporzioni: $\%Vp / \%Fp$; $\%Vn / \%Fn$.

Specificità

Validità test positivi

Definizione: proporzione V_n/sani ($V_n + F_p$)

Cioè test negativi tra i sani.

Il complemento a 1 ($1 - Sp$) indica: **errore I° tipo** o errore α : **Falsi positivi**

es. : Troponina hs, specificità 60% = V_n 60%, F_p 40%

cioè : se test negativo, correttamente classificati 60% ; se test positivo, errore 40%

L'informazione principale, fornita dal termine specificità, non è tanto quella dei correttamente classificati come negativi, ma quella complementare che indica la proporzione di errore di falsi positivi.

In pratica il termine *specificità* esprime indirettamente la validità di un **test positivo**.

Alta specificità = ridotto errore di falsi positivi: permette di individuare.

Il concetto è espresso con l'acronimo : **SpPIn** Specificity Positive rule In (5,6)

Sensibilità

Validità test negativi

Definizione: proporzione V_p/malati ($V_p + F_n$)

Cioè test positivi tra i malati.

Il complemento a 1 ($1 - Sn$) indica: **errore II° tipo** o errore β : **Falsi negativi**

es.: Troponina hs, sensibilità 95% = V_p 95%, F_n 5%

cioè : se test positivo, correttamente classificati 95% ; se test negativo, errore 5%

L'informazione principale, fornita dal termine sensibilità, non è tanto quella dei correttamente classificati come positivi, ma quella complementare che indica la proporzione di errore di falsi negativi.

In pratica il termine *sensibilità* esprime indirettamente la validità di un **test negativo**.

Alta sensibilità = ridotto errore di falsi negativi: permette di escludere.

Il concetto è espresso con l'acronimo : **SnNOut** Sensitivity Negative rule Out (5,6)

Valore predittivo, Likelihood ratio

Definizione (2)

- **Valore predittivo positivo** = proporzione tra Vp / test positivi (Vp+Fn)
- **Valore predittivo negativo** = proporzione tra Vn / test negativi (Vn+Fn)

Specificità e sensibilità sono caratteristiche intrinseche di un test, riferite alla presenza o assenza di malattia e **non dipendono dalla prevalenza della malattia**.

Il **valore predittivo** invece è direttamente proporzionale alla prevalenza della malattia nel campione esaminato: una maggiore prevalenza comporta un maggior *valore predittivo*.

La **prevalenza** della malattia è costituita dalla proporzione tra il totale dei malati e il totale dei soggetti. Rappresenta la **probabilità pre test**.

Il **likelihood ratio** (LR) o rapporto di verosimiglianza, noto come **fattore Bayes**, si ottiene dal rapporto tra le proporzioni di Vp/Fp o di Vn/Fn, cioè:

$$LR + = Sn / (1 - Sp)$$

$$LR - = Sp / (1 - Sn) .$$

Il *valore predittivo* può anche essere calcolato in base al **teorema di Bayes** (8,9):

probabilità pre test x LR = **probabilità post test**, cioè valore predittivo.

Graficamente il likelihood ratio può essere rappresentato con le cosiddette curve ROC (**R**eceiver **O**perating **C**haracteristic) riportando su diagramma cartesiano:

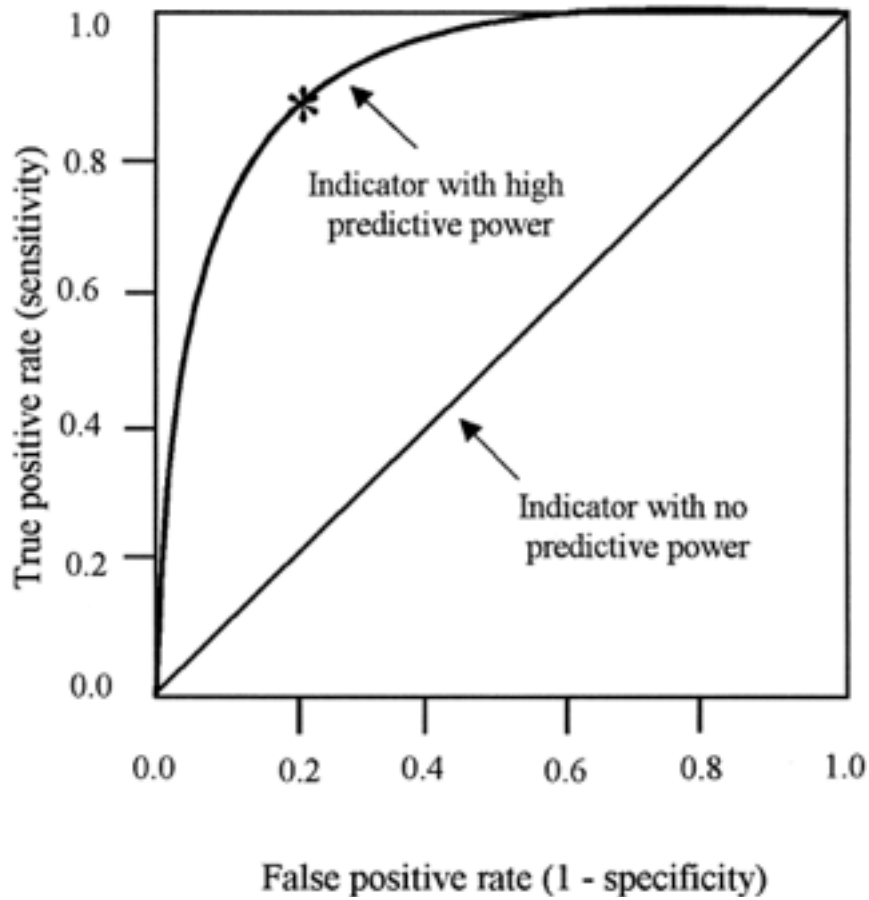
Sn in funzione di 1-Sp.

Il parametro numerico di riferimento è l'area sottesa alla curva (AUC):

valore massimo = 1, valore per differenza nulla = 0.5

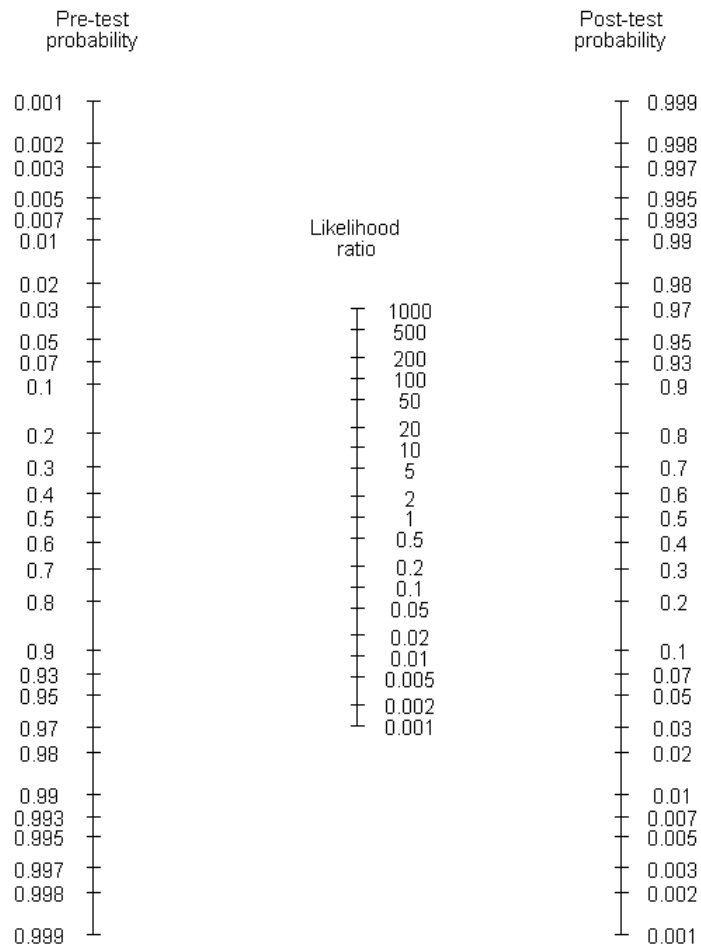
L'analisi mediante le curve ROC è attualmente il metodo di riferimento per il confronto tra test.(3)

ROC curves



(doi: 10.1128/AEM.69.11.6405-6411.2003 Appl. Environ. Microbiol. November 2003 vol. 69 no. 11 6405-6411)

NOMOGRAMMA DI FAGAN



Partendo dalla colonna sulla sinistra, che mostra la probabilità pre-test, intersecare il valore di verosimiglianza che è indicato dalla colonna centrale. L'intersezione con la colonna sulla destra indica la probabilità post-test.

(15)

Table 12 Characteristics of tests commonly used to diagnose the presence of coronary artery disease

	Diagnosis of CAD	
	Sensitivity (%)	Specificity (%)
Exercise ECG ^{a, 91, 94, 95}	45–50	85–90
Exercise stress echocardiography ⁹⁶	80–85	80–88
Exercise stress SPECT ⁹⁶⁻⁹⁹	73–92	63–87
Dobutamine stress echocardiography ⁹⁶	79–83	82–86
Dobutamine stress MRI ^{b,100}	79–88	81–91
Vasodilator stress echocardiography ⁹⁶	72–79	92–95
Vasodilator stress SPECT ^{96, 99}	90–91	75–84
Vasodilator stress MRI ^{b,98, 100-102}	67–94	61–85
Coronary CTA ^{c,103-105}	95–99	64–83
Vasodilator stress PET ^{97, 99, 106}	81–97	74–91

CAD = coronary artery disease; CTA = computed tomography angiography; ECG = electrocardiogram; MRI = magnetic resonance imaging; PET = positron emission tomography; SPECT = single photon emission computed tomography.

^a Results without/with minimal referral bias.

^b Results obtained in populations with medium-to-high prevalence of disease without compensation for referral bias.

^c Results obtained in populations with low-to-medium prevalence of disease.

Tabella tratta dalle **linee guida della Società Europea di Cardiologia**, pubblicate nell'agosto 2013, riguardo alla coronaropatia ostruttiva stabile (16).

Sono riportati i valori di **sensibilità** e **specificità** dei principali test provocativi di ischemia miocardica.

Table 13 Clinical pre-test probabilities^a in patients with stable chest pain symptoms¹⁰⁸

Age	Typical angina		Atypical angina		Non-anginal pain	
	Men	Women	Men	Women	Men	Women
30–39	59	28	29	10	18	5
40–49	69	37	38	14	25	8
50–59	77	47	49	20	34	12
60–69	84	58	59	28	44	17
70–79	89	68	69	37	54	24
>80	93	76	78	47	65	32

ECG = electrocardiogram; PTP = pre-test probability; SCAD = stable coronary artery disease.

^a Probabilities of obstructive coronary disease shown reflect the estimates for patients aged 35, 45, 55, 65, 75 and 85 years.

- Groups in white boxes have a PTP < 15% and hence can be managed without further testing.
- Groups in blue boxes have a PTP of 15–65%. They could have an exercise ECG if feasible as the initial test. However, if local expertise and availability permit a non-invasive imaging based test for ischaemia this would be preferable given the superior diagnostic capabilities of such tests. In young patients radiation issues should be considered.
- Groups in light red boxes have PTPs between 66–85% and hence should have a non-invasive imaging functional test for making a diagnosis of SCAD.
- In groups in dark red boxes the PTP is >85% and one can assume that SCAD is present. They need risk stratification only.

Tabella tratta dalle **linee guida della Società Europea di Cardiologia**, pubblicate nell'agosto 2013, riguardo alla coronaropatia ostruttiva stabile.(16)

Sono riportati i valori di **probabilità pre test** in base ad età, genere e tipo di angina.

Nelle caselle bianche sono evidenziati i valori di PTP inferiori a 15%, in questi casi non sono necessari altri test.

Nelle caselle scure sono evidenziati valori di PTP superiori a 85%, anche in questi casi non sono necessari ulteriori test diagnostici.

References

1. Altman DG, Bland JM. Diagnostic tests 1: [sensitivity and specificity](#). *BMJ* 1994; [308:1552](#).
2. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994;309:102.
3. Altman DG, Bland JM. Diagnostic tests 3: [ROC plots](#)
4. Deeks JJ, Altman DG. Diagnostic tests 4: [Likelihood Ratios](#)
5. Centre for Evidence-Based Medicine, University Health Network, Mount Sinai Hospital Toronto. Sensitivity & specificity (SnNouts and SpPins) www.cebm.utoronto.ca/glossary/spsn.htm
6. [Ruling a diagnosis in or out with "SpPIn" and "SnNOut": a note of caution](#). Pewsner D, Battaglia M, Minder C, Marx A, Bucher HC, Egger M. *BMJ*. 2004 Jul 24; 329(7459):209-13.
7. [The \(in\)validity of sensitivity and specificity](#). Guggenmoos-Holzmann , van Houwelingen HC. *Stat Med*. 2000 Jul 15;19(13):1783-92.
8. [Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease](#). Diamond GA, Forrester JS. *N Engl J Med*. 1979 Jun 14;300(24):1350-8.
9. [Bayesian methods for evidence evaluation: are we there yet?](#) Goodman SN. *Circulation*. 2013 Jun 18;127(24):2367-9. Epub 2013 May 14
10. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318:1728–33. [\[PubMed\]](#) [\[Cross Ref\]](#)
11. Yusuf S, Collins R, Peto R (1984) Why do we need some large, simple randomized trials? *Stat Med* 3: 409–422.
12. Ioannidis JP. Mega-trials for blockbusters. *JAMA*. 2013;309(3):239-240 [PubMed](#) | [Link to Article](#)
13. Kaul S; Diamond G. Trial and Error How to Avoid Commonly Encountered Limitations of Published Clinical Trials *J Am Coll Cardiol* 2010;55:415–27
14. Ioannidis J. Why most published research findings are false. *PLoS Med* 2005; 2:696-701

15. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975; 293:257.
[\[PubMed\]](#)
16. [2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology.](#) *Eur Heart J.* 2013 Oct;34(38):2949-3003