

# I TEST DIAGNOSTICI E L'ANALISI DELLA CURVA ROC

**Graziella D'Arrigo, Fabio Provenzano, Claudia Torino, Carmine Zoccali, Giovanni Tripepi**

CNR-IBIM, Unità di Ricerca di Epidemiologia Clinica e Fisiopatologia delle Malattie Renali e dell'Iperensione Arteriosa, Reggio Calabria

## INTRODUZIONE

La diagnosi, insieme alla prognosi e alla terapia, è uno dei tre processi decisionali della medicina clinica. I *test* diagnostici ideali (i cosiddetti *golden standard*) discriminano perfettamente i malati dai sani, cioè, una volta eseguito il *test*, gli individui vengono classificati con assoluta certezza come affetti o non affetti dalla malattia di interesse. Quando un determinato *test* diagnostico non discrimina in maniera netta i malati dai sani, cioè quando le distribuzioni dei risultati del *test* sono parzialmente sovrapposte negli individui affetti e non affetti da una specifica malattia, è necessario calcolare il grado di incertezza della classificazione. Se il risultato del *test* diagnostico di interesse è una variabile binaria (affetto/non affetto), è sufficiente calcolare la sensibilità, la specificità, il potere predittivo positivo, il potere predittivo negativo e l'accuratezza. Se invece il risultato del *test* è una variabile continua, è indispensabile utilizzare l'analisi della curva ROC (acronimo dei termini inglesi *Receiver Operating Characteristics*). La curva ROC è una tecnica statistica che misura l'accuratezza di un *test* diagnostico lungo tutto il *range* dei valori possibili. Poiché la curva ROC misura l'accordo tra il *test* di interesse e la presenza/assenza di una specifica malattia (così come identificata da un *golden standard*), essa rappresenta il metodo d'elezione per validare un *test* diagnostico. La curva ROC permette anche di identificare il valore soglia ottimale (il cosiddetto *best cut-off*), cioè il valore del *test* che massimizza la differenza tra i veri positivi (cioè la proporzione di individui che hanno un valore alterato del *test* tra tutti quelli realmente affetti dalla malattia) e i falsi positivi (cioè la proporzione di individui che pur avendo un valore alterato del *test* non sono affetti dalla malattia di interesse).

## KEY WORDS:

ROC curves,  
Sensitivity,  
Specificity,  
Diagnostic tests,  
Negative  
predictive value,  
Positive  
predictive value

## PAROLE CHIAVE:

Curve ROC,  
Sensibilità,  
Specificità,  
Test diagnostici,  
Valore predittivo  
negativo,  
Valore predittivo  
positivo

## SENSIBILITÀ, SPECIFICITÀ, POTERE PREDITTIVO POSITIVO E POTERE PREDITTIVO NEGATIVO

Il potere diagnostico di un *test* è di per sé un concetto multidimensionale, in quanto include la sensibilità, la specificità, il potere predittivo positivo, il potere predittivo negativo e l'accuratezza. Per comprendere il significato di questi indici consideriamo la Tabella I che mette in rapporto, in termini generali, i risultati di un ipotetico *test* diagnostico (che può essere positivo o negativo) con la presenza/assenza di una specifica malattia.

Utilizzando i dati riportati in ciascuna cella della Tabella I (denominata tavola di contingenza) è possibile calcolare i 5 indici che esprimono il potere diagnostico di un *test*, cioè:

- la **sensibilità**: la proporzione di pazienti con *test* positivo tra tutti quelli che hanno la malattia [ $a/(a+c)$ ], cioè la proporzione di veri positivi;
- la **specificità**: la proporzione di pazienti con *test* negativo tra tutti quelli che sono sani [ $d/(b+d)$ ], cioè la proporzione di veri negativi;

**TABELLA I - TABELLA 2 X 2 TRA I RISULTATI DI UN IPOTETICO TEST DIAGNOSTICO (POSITIVO/NEGATIVO) E LA PRESENZA/ASSENZA DI UNA SPECIFICA MALATTIA**

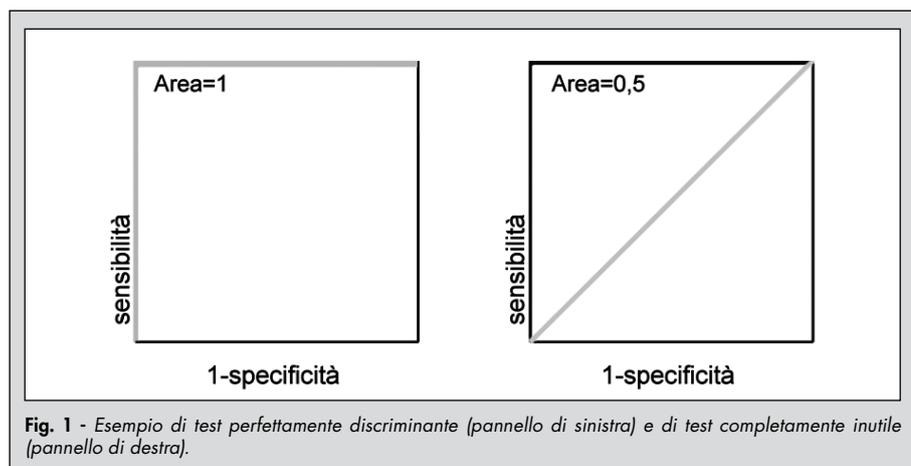
Risultato del test	Malattia		
	Presenza	Assenza	
Positivo	a	b	a+b
Negativo	c	d	c+d
Totale	a+c	b+d	N

- il **potere predittivo positivo**: la proporzione di pazienti malati tra tutti quelli che sono positivi al test [ $a/(a+b)$ ];
- il **potere predittivo negativo**: la proporzione di pazienti sani tra tutti quelli che sono negativi al test [ $d/(c+d)$ ];
- l'**accuratezza**: la proporzione di pazienti correttamente classificati [ $(a+d)/N$ ].

La sensibilità e la specificità sono misure indipendenti dalla prevalenza della malattia, cioè non sono influenzate dalla frequenza con cui una certa patologia è presente in uno specifico campione di individui. Viceversa, il potere predittivo negativo e positivo sono invece strettamente dipendenti dalla frequenza della malattia di interesse.

## LA CURVA ROC

Mentre la sensibilità e la specificità, il potere predittivo negativo e positivo classificano gli individui come affetti o non affetti da una specifica malattia sulla base di un predefinito valore del test (valore soglia), la curva ROC viene costruita considerando tutti i possibili valori del test e, per ognuno di questi, si calcola la proporzione di veri positivi (la sensibilità) e la proporzione di falsi positivi. La proporzione di falsi positivi si calcola con la formula *standard*:  $1 - \text{specificità}$ . Congiungendo i punti che mettono in rapporto la proporzione di veri positivi e di falsi positivi (le cosiddette coordinate) si ottiene una curva chiamata curva ROC. L'area sottostante alla curva ROC (AUC, acronimo dei termini inglesi "Area Under the Curve") è una misura di accuratezza



**Fig. 1 - Esempio di test perfettamente discriminante (pannello di sinistra) e di test completamente inutile (pannello di destra).**

diagnostica. Se un ipotetico nuovo *test* discriminasse perfettamente i malati dai sani, l'area della curva ROC avrebbe valore 1, cioè il 100% di accuratezza (Fig. 1, pannello di sinistra). Nel caso in cui il nuovo *test* non discriminasse per niente i malati dai sani, la curva ROC avrebbe un'area di 0.5 (o 50%) che coinciderebbe con l'area sottostante la diagonale del grafico (Fig. 1, pannello di destra). Nella realtà, si considera adeguato un *test* diagnostico con un'area sotto la curva  $\geq 80\%$ .

L'area sotto la curva può assumere valori compresi tra 0.5 e 1.0. Tanto maggiore è l'area sotto la curva (cioè tanto più la curva si avvicina al vertice del grafico) tanto maggiore è il potere discriminante del *test*. Per l'interpretazione dei valori dell'area sottostante la curva ROC è possibile riferirsi alla classificazione proposta da Swets (1):

- 1)  $AUC=0.5$  il *test* non è informativo;
- 2)  $0.5 < AUC \leq 0.7$  il *test* è poco accurato;
- 3)  $0.7 < AUC \leq 0.9$  il *test* è moderatamente accurato;
- 4)  $0.9 < AUC < 1.0$  il *test* è altamente accurato;
- 5)  $AUC=1$  *test* perfetto.

## ESEMPIO

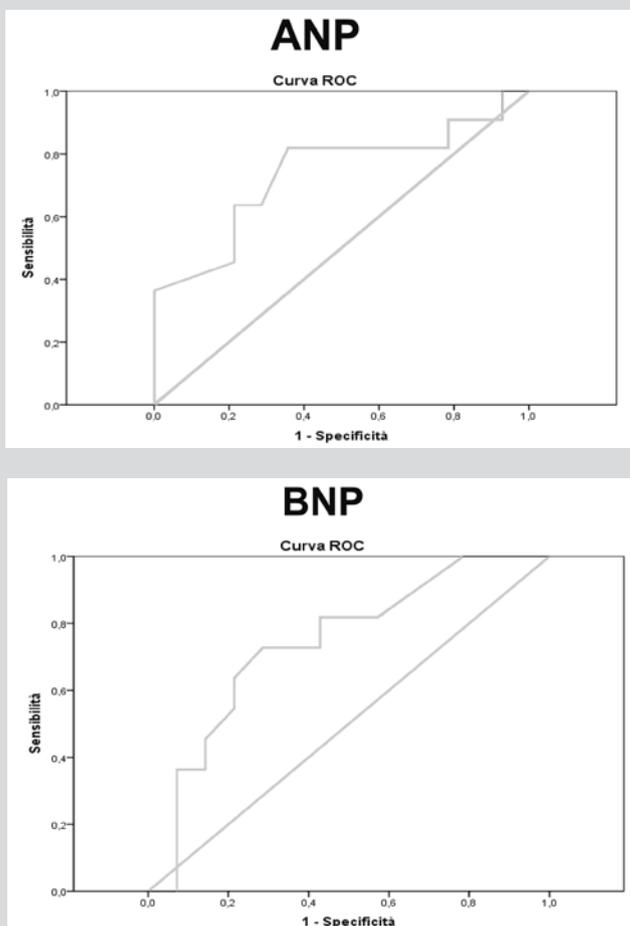
Consideriamo un ipotetico campione di 25 pazienti affetti da malattia renale terminale (Tab. II). Per ciascun paziente sono stati misurati due biomarcatori cardiaci (*atrial natriuretic peptide*, ANP, e *brain natriuretic peptide*, BNP) e la presenza/assenza di ipertrofia ventricolare sinistra è stata accertata tramite ecocardiografia (Tab. II). L'ANP è prodotto principalmente dall'atrio, mentre il BNP dal ventricolo sinistro e vi è evidenza in letteratura che questi due biomarcatori hanno un buon potere diagnostico per identificare

**TABELLA II - DATA-SET DEI 25 PAZIENTI IN DIALISI UTILIZZATI NELL'ESEMPIO DELLA CURVA ROC**

Identificativo paziente	Ipertrofia ventricolare sinistra (0=no; 1=si)	ANP (pg/mL)	BNP (pg/mL)
1	0	50	40
2	0	45	45
3	1	64	60
4	1	37	50
5	1	50	45
6	0	38	38
7	1	70	34
8	0	54	52
9	0	80	15
10	1	23	53
11	0	39	15
12	1	99	51
13	0	35	25
14	0	49	50
15	0	45	4
16	1	83	60
17	1	80	49
18	1	88	15
19	1	50	15
20	0	80	10
21	0	20	15
22	0	35	25
23	0	45	8
24	0	80	75
25	1	85	70

**TABELLA III** - COORDINATE DELLA CURVA ROC PER L'ANP PER L'IDENTIFICAZIONE DEI PAZIENTI CON IPERTROFIA VENTRICOLARE

Positivo se maggiore o uguale a:	Sensibilità (veri positivi)	1 - Specificità (falsi positivi)	Youden's index (veri positivi - falsi positivi)
19.0 pg/mL	1.00	1.00	0.00
21.5 pg/mL	1.00	0.93	0.07
29.0 pg/mL	0.91	0.93	-0.02
36.0 pg/mL	0.91	0.79	0.12
37.5 pg/mL	0.82	0.79	0.03
38.5 pg/mL	0.82	0.71	0.11
42.0 pg/mL	0.82	0.64	0.18
47.0 pg/mL	0.82	0.43	0.39
<b>49.5 pg/mL</b>	<b>0.82</b>	<b>0.36</b>	<b>0.46 (best cut-off)</b>
52.0 pg/mL	0.64	0.29	0.35
59.0 pg/mL	0.64	0.21	0.43
67.0 pg/mL	0.54	0.21	0.33
75.0 pg/mL	0.46	0.21	0.25
81.5 pg/mL	0.36	0.00	0.36
84.0 pg/mL	0.27	0.00	0.27
86.5 pg/mL	0.18	0.00	0.18
93.5 pg/mL	0.09	0.00	0.09
100.0 pg/mL	0.00	0.00	0.00

**Fig. 2** - Curva ROC di ANP e BNP per l'identificazione dell'ipertrofia ventricolare sinistra.

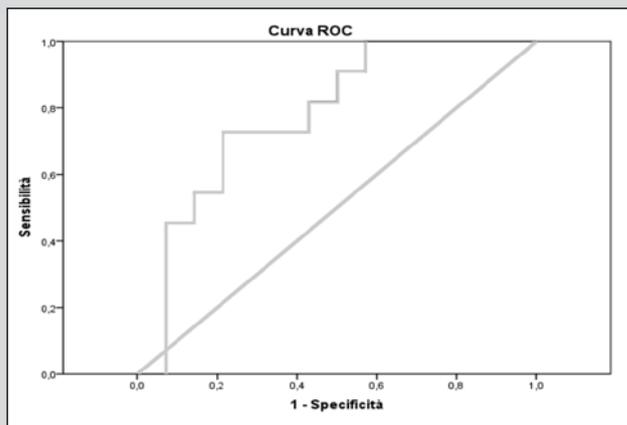


Fig. 3 - Curva ROC ottenuta dall'uso simultaneo di ANP e BNP.

l'ipertrofia ventricolare sinistra nei pazienti in dialisi (2).

Per disegnare il grafico della curva ROC è necessario calcolare la sensibilità, la specificità e la proporzione di falsi positivi (1-specificità) relative a una serie di valori soglia di ANP e BNP. La scelta dei valori soglia viene effettuata direttamente dal *software* attraverso uno specifico algoritmo che tiene conto dei valori pari merito e della numerosità dei pazienti. Per brevità, nella Tabella III sono riportate le coordinate della curva ROC per il solo ANP.

Per esempio, un *cut-off* di ANP pari a 59 pg/mL ha una sensibilità del 64% (cioè  $7/11=0.64$ ), per l'identificazione dei pazienti con ipertrofia ventricolare sinistra, e una specificità del 79% ( $11/14=0.79$ ) per escludere questa alterazione. Una specificità del 79% implica una percentuale di falsi positivi del 21% [cioè:  $(1-0.79)*100=21\%$ ]. Riportando nel grafico tutte le possibili coppie di veri positivi e falsi positivi, corrispondenti a ciascun valore soglia, si ottiene la curva ROC. La Figura 2 riporta le due curve ROC riferite al potere discriminante di ANP e BNP per l'ipertrofia ventricolare sinistra. Per l'ANP, l'AUC è 0.74 e per il BNP è 0.75. Le due curve ROC non differiscono statisticamente ( $P=NS$ ).

Un'area di AUC di 0.74 (cioè del 74%) indica che in un ipotetico esperimento che consiste nello scegliere in 100 diverse prove, in modo *random*, una coppia di pazienti di cui uno con ipertrofia ventricolare sinistra e uno senza, nel 74% dei casi i livelli di ANP sono più alti negli individui con ipertrofia ventricolare sinistra rispetto a quelli senza questa alterazione. L'ANP e il BNP hanno un buon potere discriminante per identificare l'ipertrofia ventricolare sinistra nei pazienti con malattia renale terminale in quanto gli intervalli di confidenza al 95% delle due curve ROC (BNP, IC al 95%: 0.54-0.94; ANP, IC al 95%: 0.59-0.96) non includono 0.5 (la soglia di indifferenza diagnostica).

Se l'obiettivo del ricercatore è quello di calcolare il potere discriminante dell'uso simultaneo di ANP e BNP per l'identificazione dell'ipertrofia ventricolare sinistra, è necessario costruire un modello logistico multiplo (3). In questo modello la variabile dipendente è la presenza/assenza dell'ipertrofia ventricolare sinistra e le variabili indipendenti sono i livelli circolanti di ANP e BNP. Attraverso questa analisi è possibile calcolare, a livello individuale, la probabilità di avere l'ipertrofia ventricolare sinistra a partire dai valori di ANP e BNP e di testare con l'analisi della curva ROC il potere discriminante della probabilità così calcolata. Nel nostro caso, l'uso simultaneo di entrambi i biomarcatori produce un'AUC di 0.78 (IC al 95%: 0.59-0.96) che non differisce

da quelle relative ai soli BNP e ANP. Pertanto, l'uso contemporaneo dei due biomarcatori non ha un potere discriminante aggiuntivo rispetto ai singoli biomarcatori (Fig. 3). Attraverso la curva ROC è possibile identificare il *best cut-off*, cioè il valore del *test* che massimizza la differenza tra veri positivi e falsi positivi [test di Youden (4)]. Nel nostro caso, il *best cut-off* per l'ANP è 49.5 pg/mL che si associa a una sensibilità dell'82% e a una percentuale di falsi positivi del 36% (Tab. III). Come è evidente dalla Tabella III, questo valore è quello che massimizza la differenza tra veri positivi e falsi positivi (test di Youden=0.46 o 46%) per l'identificazione dell'ipertrofia ventricolare sinistra.

## CONCLUSIONI

La curva ROC è una tecnica statistica che misura l'accuratezza di un *test* diagnostico lungo tutto il *range* dei valori possibili. La curva ROC permette anche di identificare il valore soglia ottimale (il cosiddetto *best cut-off*), cioè il valore del *test* che massimizza la differenza tra i veri positivi e i falsi positivi. Per ottenere validi risultati attraverso l'uso delle curve ROC è indispensabile che la presenza/assenza di una specifica malattia sia accertata tramite un *golden standard*. L'area sotto la curva ROC è una misura del potere discriminante del *test*. Le curve ROC possono essere confrontate fra loro con l'uso di un appropriato *test* statistico disponibile in quasi tutti i *software* in commercio. Per calcolare il potere discriminante di una serie di biomarcatori è necessario utilizzare la regressione logistica multipla.

## DICHIARAZIONE DI CONFLITTO DI INTERESSI

Gli Autori dichiarano di non avere conflitto di interessi.

## Indirizzo degli Autori:

Dr. Giovanni Tripepi  
CNR-IBIM, Istituto di Biomedicina  
Epidemiologia Clinica e Fisiopatologia  
delle Malattie Renali e dell'Ipertensione Arteriosa  
Via Vallone Petrarà 55/57  
89124 Reggio Calabria  
e-mail: gtripepi@ibim.cnr.it

---

## BIBLIOGRAFIA

1. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1998; 240: 1285-93.
2. Mallamaci F, Zoccali C, Tripepi G, et al. Diagnostic potential of cardiac natriuretic peptides in dialysis patients. *Kidney Int* 2001; 59: 1559-66.
3. Provenzano F, D'Arrigo G, Zoccali C, Tripepi G. La regressione logistica nella ricerca clinica. *G Ital Nefrol* 2011; 28: 210-3.
4. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3: 32-5.