

- *Quando si sottopone una popolazione ad una procedura diagnostica, non tutti i soggetti malati risulteranno positivi al test, così come non tutti i soggetti sani risulteranno negativi. Il test, cioè, fornirà inevitabilmente un certo numero di falsi negativi e di falsi positivi.*
- *Sensibilità e specificità sono due parametri correlati alla qualità del test impiegato e dipendenti dalla quota di falsi positivi e falsi negativi generati dal test.*
- *La predittività del test, ovvero la probabilità di individuare, sulla base di un singolo risultato, lo stato di malato/sano, oltre ad essere correlata alla qualità del test impiegato, dipende anche dalla frequenza della malattia.*

Affidabilità e validità di un test

Un test ideale (sia esso clinico, strumentale, di laboratorio) dovrebbe essere affidabile e valido.

Per affidabilità generalmente s'intende la capacità di un test di offrire sempre lo stesso risultato nel corso di misurazioni ripetute. Questa pertanto è una caratteristica intrinseca al test e dipendente dalla bontà dello strumento e/o dell'operatore.

Esiste però un altro parametro importante da valutare, rappresentato dalla validità: ovvero dalla capacità di un test di distinguere in una popolazione i soggetti sani da quelli malati.

Utilizzando un test ideale, pertanto, tutti i soggetti sani dovrebbero risultare negativi al test ed analogamente tutti i malati dovrebbero risultare positivi.

La situazione appena descritta, riportata su un semplice diagramma cartesiano sul cui asse delle ascisse è indicato il valore della variabile misurata dal test (ad esempio, pressione arteriosa o glicemia o calcemia, ecc.) e sull'asse delle ordinate il numero di soggetti individuati nelle singole misurazioni, darebbe luogo a due curve ben distinte: da un lato i sani, dall'altro i malati (figura 9.1).

Il valore della variabile, discriminante per assegnare un soggetto al gruppo dei sani o dei malati, viene chiamato valore di cut-off.

Ipotizzando che la variabile misurata sia la glicemia basale e la patologia indagata sia il diabete mellito, potremmo dire che (se questo test fosse un test ideale!) dato il valore di 110 mg/ml, tutti i diabetici risulterebbero avere una glicemia basale >110 e tutti i soggetti non diabetici un valore inferiore o uguale a 110.

Se ciò fosse vero (per la misurazione della glicemia basale così come per tutti gli altri test diagnostici), basterebbe una singola misurazione per far diagnosi di malattia!

Sfortunatamente in medicina la realtà è notevolmente differente...

Sensibilità e specificità

Sottoponendo una qualsiasi popolazione ad un test di screening, dato un certo valore di cut-off purtroppo avremo sempre un certo numero di soggetti sani che risulteranno positivi al test e, simmetricamente, un certo numero di soggetti malati che il test non riuscirà a identificare come tali, e pertanto saranno erroneamente classificati come «sani».

Affidabilità di un test: capacità del test di offrire sempre lo stesso risultato nel corso di misurazioni ripetute

Validità di un test: capacità del test di distinguere in una popolazione i soggetti sani da quelli malati

Valore di cut-off: è quel valore assunto dalla variabile misurata nel test al di sopra del quale il test viene definito *positivo* e al di sotto del quale il test si definisce *negativo*

Il test ideale... si avrebbe qualora, dato un certo valore di cut-off, tutti i soggetti sani risultassero negativi al test e tutti i soggetti malati positivi

Il test reale...
una popolazione sottoposta ad un test di screening si distribuisce sempre realizzando un certo livello di sovrapposizione fra la sottopopolazione dei sani e quella dei malati

Riportando sul solito grafico la situazione precedentemente illustrata, possiamo vedere come la curva dei sani e quella dei malati presentano in realtà un certo livello di sovrapposizione (figura 9.1).

Un «coda» della curva scura infatti si ritrova al di là del limite imposto dal cut-off: questi soggetti, pur essendo sani, pertanto saranno classificati come malati e costituiranno la quota dei *falsi positivi*.

Analogamente la coda sinistra della curva dei malati cade al di sotto del livello di cut-off: tali soggetti, pur essendo malati, saranno erroneamente identificati come sani e formeranno il gruppo dei *falsi negativi*.

Un ulteriore modo di rappresentare la distribuzione di un'ipotetica popolazione in funzione della presenza/assenza di malattia e dei risultati di un test può essere dato dalla classica tabella 2x2, a noi già familiare.

Le colonne rappresentano la distinzione dei soggetti in malati e sani; nelle righe invece i pazienti sono distribuiti in funzione del risultato al test (tabella 9.1).

Tanto più basse saranno le quote di falsi positivi e falsi negativi, tanto più il test sarà valido.

Un primo parametro da valutare in un test è la *sensibilità*.

Per sensibilità si intende la capacità di un test di individuare in una popolazione i soggetti malati. Essa è data dalla proporzione dei soggetti realmente malati e positivi al test (veri positivi) rispetto all'intera popolazione dei malati.

Come si evince dalla definizione (e dalla formula riportata a lato), la sensibilità è condizionata negativamente dalla quota di falsi negativi: pertanto un test molto sensibile dovrà associarsi ad una quota molto bassa di falsi negativi (ovvero di soggetti

Tabella 9.1 – RAPPRESENTAZIONE SU UNA TABELLA 2X2 DELLA DISTRIBUZIONE DI UNA POPOLAZIONE IN BASE AI RISULTATI DI UN TEST

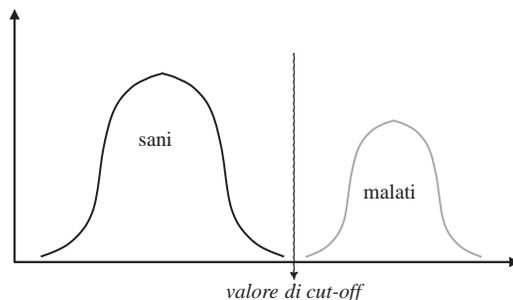
	M+	M-	
T+	VP	FP	T _P
T-	FN	VN	T _N
	T _{M+}	T _{M-}	N

M+= malati; M-= sani
T+= test positivo; T-= test negativo

Sensibilità:
capacità del test di individuare in una popolazione i soggetti malati

$$\frac{VP}{T_{M+}} = \frac{VP}{VP + FN}$$

VP= veri positivi
FN= falsi negativi
T_{M+}= totale malati



il test ideale...

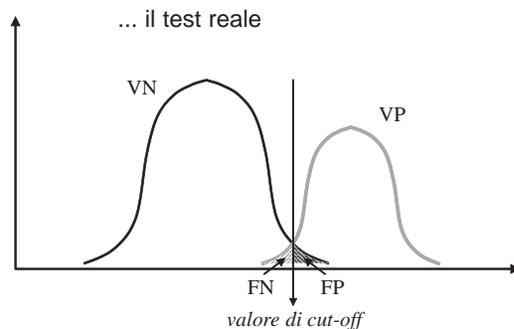


FIGURA 9.1

Distribuzione di una popolazione in funzione del risultato di un test di screening.

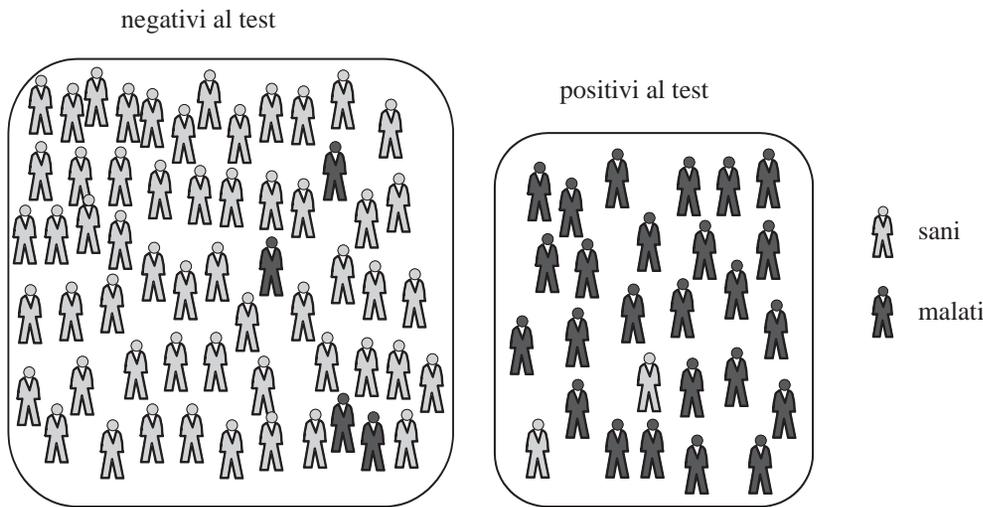


FIGURA 9.2

Distribuzione schematica di una popolazione in funzione dello stato di malattia e del risultato di un test di screening

malati che «sfuggono» all'identificazione attraverso il test).

Il calcolo della sensibilità tiene quindi in conto esclusivamente la popolazione dei malati (ovvero la prima colonna della tabella 2x2), in funzione dell'identificazione come positivi o negativi al test.

Essa è pertanto una caratteristica intrinseca al test stesso e non dipende affatto dalla prevalenza della malattia nella popolazione sottoposta al test.

Nell'esempio riportato nella figura 9.2 la sensibilità risulta pari all'86,2%, in quanto il test, su 29 soggetti malati, è riuscito ad identificarne 25.

Un secondo parametro, per certi versi specularmente al precedente, è dato dalla *specificità*.

Per specificità si intende la capacità di un test di identificare come negativi i soggetti sani. Come si può notare dalla formula riportata a lato, la specificità è influenzata in particolare dalla quota di falsi positivi: ovvero un test sarà tanto più specifico quanto più bassa risulterà la quota di falsi positivi (cioè di soggetti sani identificati dal test come malati).

Un test molto specifico, in definitiva, ci consente di limitare la possibilità che un soggetto sano risulti positivo al test.

Utilizzando ancora una volta lo schema della figura 9.2, per calcolare la specificità dovremo fare riferimento esclusivamente al gruppo dei sani ed alla loro distribuzione fra positivi e negativi al test.

Nell'esempio in figura la specificità sarà data dalla proporzione di sani identificati come tali dal test (55) ed il totale dei sani (57), quindi sarà pari al 96,5%.

Un test altamente specifico sarà dunque un test che produrrà una bassa quota di falsi positivi.

Se il test in questione fosse rappresentato dalla misurazione di una variabile continua (come la glicemia dell'esempio precedente), una maniera per aumentarne la specificità sarebbe quello di aumentare il limite di cut-off, ovvero il livello al di sopra del quale «etichettare» un soggetto come malato.

In ogni caso, ovviamente, la reale distribuzione della popolazione fra malati e sani in funzione della variabile misurata non cambierebbe (!), pertanto spostando la linea a destra avremmo una riduzione globale dei soggetti positivi al test con un conseguente aumento della quota di falsi negativi, cioè di soggetti realmente malati che andremmo ad identificare come sani.

Specificità:
capacità del test di individuare in una popolazione i soggetti sani

$$\frac{VN}{T_{M-}} = \frac{VN}{VN + FP}$$

VN= veri negativi
FP= falsi positivi
T_{M-}= totale sani

Essendo aumentata la quota di falsi negativi, diminuirebbe quindi la sensibilità (figura 9.3).

In maniera analoga, potremmo invece aumentare la sensibilità del test (se, per esempio, volessimo garantirci di poter riconoscere la quota più alta possibile di soggetti malati). In tal caso, inevitabilmente, abbassando il livello di cut-off, includeremmo nel gruppo dei positivi un certo numero di sani (la coda destra della curva dei sani) che rappresenterebbero i falsi positivi: diminuirebbe pertanto la specificità del test.

Sensibilità e specificità sono, quindi, due parametri reciprocamente dipendenti.

Per capire meglio questo concetto potremmo pensare al funzionamento dell'allarme antifurto di un'automobile.

Il proprietario può a suo piacimento aumentarne o diminuirne la sensibilità: valori molto alti di sensibilità farebbero scattare l'allarme molto spesso (magari al solo volare di una mosca...); si potrebbero cioè verificare molti falsi allarmi (... cioè falsi positivi), con conseguente scarsa specificità.

Al contrario, abbassando la sensibilità dell'allarme, ridurremmo certo i falsi positivi, aumenteremmo la specificità, ma po-

tremmo correre il rischio di avere falsi negativi: cioè l'allarme non suonerà, e il ladro porterà via la nostra auto...

I concetti finora esposti sono di estrema importanza soprattutto nella pratica clinica. È infatti molto interessante poter valutare come un test si comporti in una popolazione (in termini di sensibilità e specificità) in funzione di determinati valori di cut-off. A tale scopo sono state realizzate le cosiddette curve ROC. Non sono altro che la rappresentazione su di un grafico, che riporta in ascisse la percentuale di errore falso positivo (ovvero 1 meno la specificità) ed in ordinate la sensibilità, dei diversi valori misurati in una popolazione a differenti livelli di cut-off.

Immaginiamo di voler identificare i soggetti diabetici con una singola misurazione della glicemia basale. In condizioni limite, se stabilissimo come cut-off 400 mg/100 ml, avremmo tutta la popolazione identificata come sana, pertanto il tasso di errore falso positivo sarebbe 0... ma anche la sensibilità! (Nessun malato infatti sarebbe identificabile.) Segneremmo pertanto un puntino in basso a sinistra (figura 9.4). Al contrario, se il cut-off fosse fissato a 40 mg/100 ml, allora tutta la popolazione sa-

Sensibilità e specificità sono parametri fra loro inversamente proporzionali e dipendenti esclusivamente dalla tipologia di test

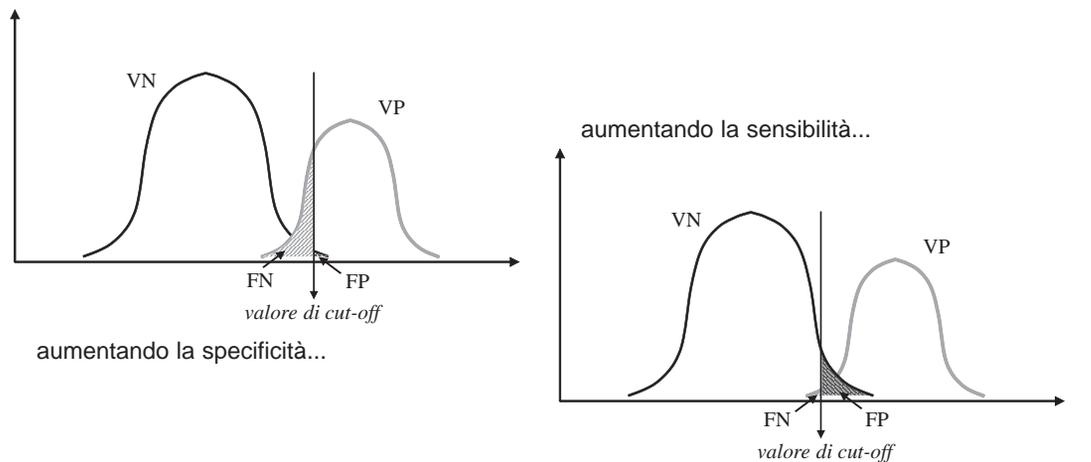


FIGURA 9.3

Specificità e sensibilità in funzione di variazioni del cut-off.

rebbe identificata come malata: certamente riusciremmo ad includere tutti i malati (sensibilità= 100%), ma il tasso di errore falso positivo sarebbe anche del 100%! In questo caso segneremmo un puntino in alto a destra. Per costruire la curva avremo bisogno di misurazioni nella popolazione con livelli intermedi di cut-off: ad ogni livello corrisponderà una coppia «sensibilità/errore falso positivo», quindi un puntino sul grafico. Ovviamente la linea retta ritrae una situazione poco auspicabile, in cui la sensibilità è sempre pari al valore del tasso di falsi positivi: rappresenta cioè la linea di «nessun beneficio». Tanto più la curva misurata si scosta dalla linea di nessun beneficio, tanto essa sarà migliore, in quanto ci permetterà di identificare un valore di cut-off (ovvero un puntino) che ci dia il massimo della sensibilità con il tasso di errore falso positivo più basso possibile.

L'uso delle curve ROC è oggi assai diffuso in svariati settori, ma il loro approfondimento esula dalle finalità del nostro corso. Rimandiamo pertanto ad altri testi per i necessari approfondimenti.

Predittività di un test

Un concetto, invece, su cui val la pena di riflettere (anche per le sue implicazioni in sanità pubblica), è quello di predittività.

Finora, infatti, abbiamo trattato di parametri che, in un certo senso, sono definibili a priori: sensibilità e specificità sono caratteristiche intrinseche di un test. Esse ci informano su qual è la probabilità di reclutare soggetti malati o sani da una certa popolazione di partenza (di malati o di sani), mentre nulla ci dicono sulla probabilità che abbiamo, di fronte ad un singolo risultato positivo, che quel soggetto sia realmente malato.

Per rispondere a questo interrogativo dobbiamo poter calcolare un nuovo parametro: il valore predittivo positivo (VPP).

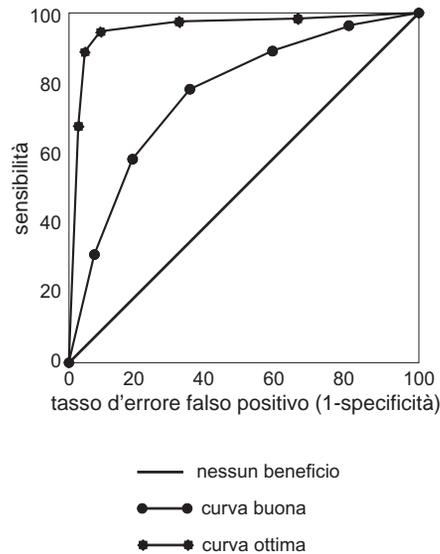


FIGURA 9.4

Curve ROC. Esempi di curve risultanti da misurazioni su popolazioni differenti.

Esso esprime proprio la probabilità che ha un soggetto, risultato positivo al test, di essere realmente malato.

Il VPP si calcola come quota di soggetti veri positivi sul totale dei positivi (veri e falsi positivi).

Ritornando allo schema della figura 9.2, valutando tutti i soggetti identificati come positivi al test, il VPP sarà dato dal numero di soggetti realmente malati (veri positivi), cioè 25, su tutti quelli risultati positivi (veri e falsi), cioè 27.

Il valore risultato (92,6%) ci indica la probabilità per un soggetto con un test positivo di essere realmente malato.

In maniera del tutto speculare possiamo calcolare il valore predittivo negativo, come la quota di veri negativi sul totale dei negativi.

Nell'esempio in figura esso sarà pari a 55 (veri negativi) sul totale dei negativi (che include anche quattro falsi negativi). Il 93,2% indicherà la probabilità che ha un soggetto risultato negativo al test di essere effettivamente sano.

La predittività di un test, al contrario di specificità e sensibilità, non è una caratteristica intrinseca del test, ma può dipende-

Valore predittivo positivo:

esprime la probabilità che un soggetto risultato positivo ad un test sia effettivamente malato

$$\frac{VP}{T_p} = \frac{VP}{VP + FP}$$

VP= veri positivi
FP= falsi positivi
T_p= totale positivi

Valore predittivo negativo:

esprime la probabilità che un soggetto risultato negativo ad un test sia effettivamente sano

$$\frac{VN}{T_N} = \frac{VN}{VN + FN}$$

VN= veri negativi
FN= falsi negativi
T_N= totale negativi

La predittività di un test
 è proporzionale alla
 prevalenza della malattia
 nella popolazione
 sottoposta al test

re dalla frequenza della malattia nella popolazione.

Ipotizziamo di avere una popolazione rappresentata da 400 soggetti, in cui si registra una prevalenza del 5% di malati (figura 9.5).

Proviamo ora a condurre uno screening su questa popolazione utilizzando un ipotetico test che presenti valori di sensibilità e specificità entrambi pari al 90%.

Da ciò deriva che dei 20 soggetti malati 18 saranno identificati dal test come positivi. Dei 380 soggetti sani, ne saranno identificati come negativi 342: i restanti 38 risulteranno falsamente positivi al test.

La popolazione dei positivi, pertanto, sarà composta dai 18 malati (VP) e dai 38 sani (FP).

Dalla popolazione di origine, pertanto, sono stati individuati come positivi complessivamente 56 soggetti. Di questi, però, solo 18 sono realmente malati.

In definitiva, la probabilità che un soggetto positivo sia realmente malato (cioè il VPP) è estremamente bassa: 32,1%.

Ora ipotizziamo di ripetere la stessa esperienza in una popolazione che presenta una prevalenza di malattia superiore (20%) (figura 9.6).

Utilizzando lo stesso test (sensibilità e specificità pari al 90%), riuscirei a identificare come positivi 72 degli 80 soggetti ma-

lati, e dei restanti 320 soggetti sani, 288 risulterebbero negativi e 32 positivi (falsi positivi).

La popolazione dei positivi risulterebbe pertanto composta da 72 malati e 32 sani. Il valore predittivo positivo del test, in questo caso, salirebbe al 69,2%.

Questa osservazione presenta numerosi risvolti pratici. Ritornando al paragone dell'allarme antifurto, potremmo dire che, a parità di sensibilità, la probabilità che ad un allarme corrisponda effettivamente un tentativo di furto (e che non si tratti di un falso allarme!) dipende dal contesto; se ci troviamo in una città dove i furti sono rari, è più probabile che si tratti di un falso allarme, se ci troviamo in una città ad alto tasso di furti, allora è il caso di preoccuparsi al primo suonar di sirena...

Quanto esposto è particolarmente importante qualora si decidesse di avviare una campagna di screening di massa.

Un test di screening oltre a possedere alcune caratteristiche particolari (di facile esecuzione, poco costoso, accettabile dall'utente, ecc.), deve possedere senza dubbio una buona sensibilità.

Ma, secondo quanto finora osservato, dobbiamo sempre essere consapevoli che la predittività del test sarà sempre proporzionale alla prevalenza della malattia nella popolazione sottoposta a screening; per aumentar-

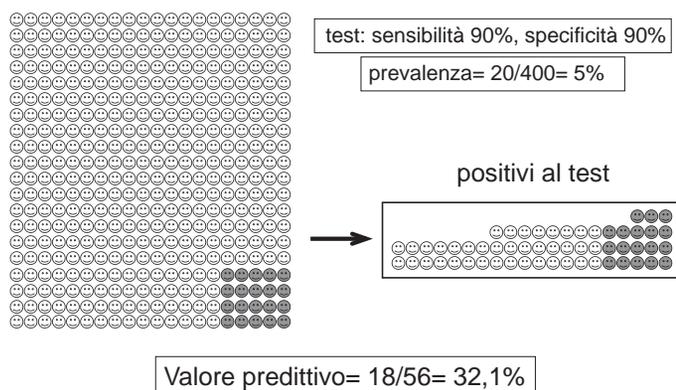


FIGURA 9.5

Predittività di un test di screening in una popolazione a bassa prevalenza.

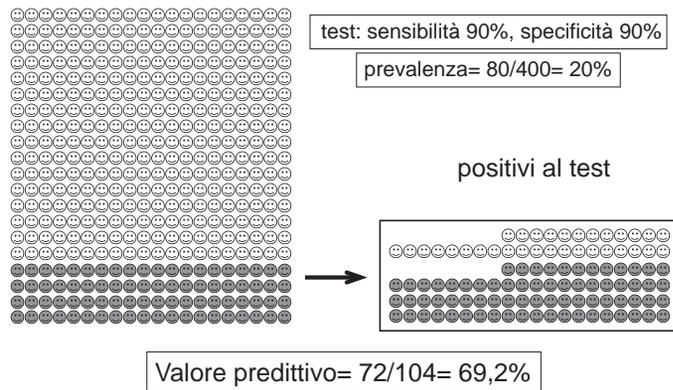


FIGURA 9.6

Predittività di un test di screening in una popolazione a prevalenza più elevata.

la, pertanto, sarà bene scegliere accuratamente la popolazione su cui avviare lo screening, per evitare di dover fare i conti con una quota troppo elevata di falsi positivi.

In ogni caso, solitamente ad un primo test di screening conviene far seguire un secondo test (dotato generalmente di maggiore specificità) cosiddetto «di conferma», che avrà proprio lo scopo di identificare (e quindi escludere) i falsi positivi nel gruppo dei soggetti risultati positivi al primo test. La predittività del secondo test sarà sempre molto elevata, in quanto esso verrà eseguito su una popolazione fortemente selezionata dal primo test e, quindi, ad elevata prevalenza.

Come abbiamo visto, quindi, i valori predittivi di un test sono influenzati pesantemente dalla prevalenza della condizione in esame. Un test con un valore predittivo positivo molto vicino al 100% sarà comunque poco utile se la prevalenza della condizione che vogliamo studiare è molto bassa. In altre parole quanto più la prevalenza della condizione in esame è elevata, tanto migliore sarà la performance di un test con un elevato valore predittivo.

La conseguenza diretta di questa osservazione è che lo stesso test diagnostico potrà funzionare in modo diverso secondo la popolazione che viene ad esso sottoposto. La prevalenza della condizione in esame, in effetti, può essere molto variabile secondo

lo scenario in cui operiamo, anche per la stessa malattia. Se noi vogliamo applicare un test di screening alla popolazione generale, la probabilità di incontrare una determinata condizione patologica sarà uguale alla prevalenza. Se noi invece vogliamo applicare il test diagnostico ai pazienti che afferiscono ad un ambulatorio specialistico, la prevalenza di questa popolazione sarà notevolmente maggiore di quella della popolazione generale.

Questa «prevalenza», o meglio, la probabilità di incontrare un paziente con una certa malattia si definisce come «probabilità pre-test». Tale probabilità può variare secondo la prevalenza nella popolazione generale, il gruppo di età, il sesso, la presenza di sintomi clinici, e, appunto, lo scenario nel quale il paziente viene osservato.

Il likelihood ratio o rapporto di verosimiglianza esprime di quante volte la probabilità di una determinata diagnosi di malattia è modificata per effetto del test.

Il rapporto di verosimiglianza positivo (LR+) si calcola come segue:

LR+ = proporzione di veri positivi/proporzione di falsi positivi

ovvero:

sensibilità/1-specificità.

Il likelihood ratio

è dato dal rapporto fra la proporzione di veri e di falsi positivi

$\frac{\text{proporzione di VP}}{\text{proporzione di FP}}$

Il **likelihood ratio** è dato dal rapporto fra la proporzione di falsi e di veri negativi

proporzione di FN
proporzione di VN

Il rapporto di verosimiglianza negativo (LR-) si calcola invece:

LR- = proporzione di falsi negativi/proporzione di veri negativi

ovvero:

1-sensibilità/specificità.

In tutti e due i casi il rapporto di verosimiglianza esprime un valore che moltiplicato per la probabilità pre-test ci permetterà di calcolare la probabilità post-test in caso di risultato positivo (LR+) o negativo (LR-).

Per il rapporto di verosimiglianza positivo, valori superiori a 10 indicano che il test è molto efficace nell'aumentare la nostra probabilità pre-test. Allo stesso modo, per il rapporto di verosimiglianza negativo, valori minori di 0,1 sono da considerare tipici di test particolarmente attendibili.

L'uso di questo parametro permette di eseguire valutazioni della performance di un test diagnostico del tutto indipendenti dalla prevalenza della condizione in esame e di verificarne l'utilità secondo la propria realtà specifica.