

I TEST DIAGNOSTICI NELLA EPIDEMIOLOGIA: LINEAMENTI TEORICI, SVILUPPI METODOLOGICI ED ESEMPI APPLICATIVI

Massimiliano Giacalone

Università degli Studi di Bologna

e-mail: massimilia.giacalone@unibo.it

Giorgio Suvero

e-mail: suverogiorgio@gmail.com

Riassunto

Il presente contributo intende mostrare le tecniche statistiche che possono risultare di utile supporto ai clinici e agli operatori per la scelta del tipo di trattamento più idoneo da adottare nei singoli casi nell'interesse dei pazienti. Un certo numero di esempi concreti illustra le tecniche trattate.

Parole chiave: epidemiologia, test diagnostici, studi caso-controllo, posti letto, teorema di Bayes, curva ROC.

1. INTRODUZIONE

Un qualsiasi discorso di *statistica epidemiologica* richiede una definizione preventiva del termine *epidemiologia*. Il termine è costituito dalle parole greche “epi”, “demos” e “logos”, che significano “su”, “popolo” e “studio”. L'epidemiologia, come studio degli avvenimenti sulla popolazione, è legata storicamente allo studio delle epidemie, ma oggi i suoi campi di applicazione sono più vasti. L'O.M.S. definisce l'*epidemiologia* come “la descrizione della distribuzione e della dimensione dei problemi connessi alle malattie ed alle infermità nella popolazione umana e la identificazione dei fattori eziologici nella patogenesi delle malattie”. L'epidemiologia si avvale di metodologie statistiche non solo per riassumere le informazioni sulle popolazioni o gruppi di individui studiati, ma anche per le modalità di conduzione delle indagini.

Dunque, l'obiettivo primario della *statistica epidemiologica* è di indi-

viduare le cause e i fattori che determinano l'insorgenza e condizionano la diffusione della malattia, attraverso numeri, formule, tabelle e grafici; un obiettivo secondario è la valutazione della efficacia e della efficienza dei servizi sanitari, per individuare i reali bisogni di salute della popolazione, al fine di programmare in modo appropriato diagnosi e cure.

Le indagini epidemiologiche possono essere di tipo descrittivo, analitico, sperimentale.

Gli studi descrittivi si occupano della frequenza e della distribuzione delle malattie nella popolazione. Si avvalgono di dati statistici relativi alla morbosità e alla mortalità delle malattie in relazione alla loro distribuzione nel tempo, geografica e alle caratteristiche delle persone. Vengono quindi evidenziate le variabili legate al tempo, elemento importante nello studio epidemiologico poiché dà informazioni sulla durata della manifestazione della malattia; variabili legate allo spazio, luoghi diversi infatti possono essere rilevanti nel confronto tra i tassi di una malattia che oltretutto può essere diffusa in zone ristrette o molto ampie; le variabili individuali, di interesse risultano le caratteristiche sociali e demografiche come età, sesso, etnia, stato civile, condizioni socio-economiche e così via.

Gli studi analitici prendono in esame il rapporto di causalità che intercorre tra i fattori di rischio e la malattia; individuano le cause delle malattie e i fattori che ne favoriscono o ne ostacolano l'insorgenza e la diffusione. Questo tipo di studi può essere trasversale, di caso/controllo, di coorte.

Lo studio trasversale, noto anche come studio di prevalenza, ha come scopo quello di definire la frequenza di riscontro di una condizione morbosa o di una caratteristica nella popolazione. Può valutare lo stato di salute collettivo e i problemi sanitari di una popolazione ai fini della corretta programmazione sanitaria. Questo tipo di studio non si presta all'indagine su un fenomeno patologico raro. Il suo successo si basa sulle capacità organizzative e sulle risorse a disposizione del ricercatore, nonché sulla disponibilità della popolazione a partecipare all'indagine.

Gli studi retrospettivi o di caso/controllo indagano sul ruolo dei fattori di rischio ricostruendo la storia della malattia e l'esposizione al rischio, per questo motivo sono utili nell'analisi di malattie rare. Mettono a confronto due gruppi di soggetti: il gruppo dei casi che ha già sperimentato l'evento in studio e il gruppo di controllo che invece tale evento non ha avuto, ma che è il più possibile simile ai casi in ogni aspetto. Per accertare se la prevalenza di un evento clinico è considerevole si inizia, nei due gruppi, la ricerca della

trascorsa esposizione al fattore di rischio. Questo permetterà di valutare se eventuali differenze sono prodotte dalla realtà clinica o da eventi dovuti alla casualità. Inoltre, si parla di studio retrospettivo perché per la sua esecuzione si fa riferimento a dati raccolti prima della effettuazione della ricerca. Quindi lo scopo dello studio è la valutazione del rischio relativo, ovvero del rischio di contrarre una malattia per le persone esposte al fattore in esame, rispetto a quelle non esposte. Questo studio può essere eseguito relativamente facilmente ed a costi bassi. Inoltre, per valutare eventuali effetti causali non occorre attendere il periodo che esiste tra esposizione al fattore e insorgenza della malattia.

Partendo da queste premesse, abbiamo cercato di chiarire nell'articolo, anche con l'ausilio di semplici esempi il legame tra l'epidemiologia e le tecniche statistiche più idonee per la valutazione *ex-ante* ed *ex-post* riguardante la probabilità di sviluppare o meno una determinata malattia per i soggetti sottoposti a test di natura statistica.

L'analisi è strutturata in tre parti, che costituiscono l'oggetto dei paragrafi seguenti.

Nel primo, vengono fornite informazioni statistiche e valutazioni generali sui test diagnostici descrivendo situazioni reali ed ideali ed introducendo le misure più comunemente usate per discriminare sulla validità dei predetti test, quali la sensibilità, la specificità, l'efficacia e l'accuratezza.

Nel secondo vengono analizzate le indagini epidemiologiche, descrittive, analitiche, retrospettive e sperimentali chiarendo la differente struttura e le diverse modalità di tali indagini statistiche

Nel terzo vengono, infine, descritti alcuni esempi applicativi con l'impiego del teorema di Bayes, largamente utilizzato in questi studi per la determinazione dei due parametri VPP e VPN (valore predittivo positivo) e (valore predittivo negativo), fondamentali per verificare *ex- post* l'effettiva rispondenza alla realtà dei test impiegati *ex- ante*.

1. I TEST DIAGNOSTICI E LE VALUTAZIONI STATISTICHE

1.1 TEST DIAGNOSTICI

Per verificare un'ipotesi, in qualsiasi ambito scientifico, viene utilizzata una procedura concettualmente standard, nota come test statistico.

Un test diagnostico ricade in questo ambito. Questi test trovano appli-

cazione nel settore dell'epidemiologia ove rappresentano lo strumento di base nelle operazioni di screening, eseguite su popolazioni presuntivamente sane allo scopo di identificare precocemente la presenza di infezioni o di malattie subcliniche. Nell'attività diagnostica di routine, i test svolgono un ruolo determinante nel processo decisionale volto a confermare o escludere la presenza di una determinata malattia già sospettata in base ai dati clinici. Ai due suddetti vasti campi di azione se ne può aggiungere un terzo che spazia nell'opera di controllo degli alimenti di origine animale, dove i test vengono applicati per determinare sanità e salubrità (assenza di agenti patogeni, sostanze chimiche ecc.) dei suddetti prodotti.

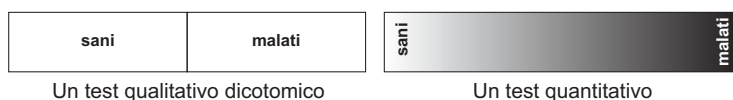
In base alla tipologia di responso fornito, i test possono essere classificati in due categorie. Alla prima appartengono i test "qualitativi" in quanto misurano l'esistenza (qualità) di un fenomeno e non la sua ampiezza (quantità) fornendo risultati del tipo positivo/negativo oppure sano/malato, un test di questo tipo con output binario (in due sole categorie), viene detto «nominale dicotomico».

La seconda, numericamente più consistente, comprende i test di tipo "quantitativo" che forniscono risultati numerici misurabili su una scala numerica (variabili «continue»), come ad esempio i valori di densità ottica (D. O.) di un test ELISA misurati con lo spettrofotometro.

Il test ELISA (Enzyme - Linked Immuno Sorbent Assay) è test quantitativo che impiega un enzima coniugato ad un anticorpo per identificare e quantificare la presenza di anticorpi (o di antigeni) nel siero di sangue o in altri materiali. In caso di positività, l'enzima induce una variazione di colore tanto più intensa quanto più elevata è la presenza di anticorpi (o di antigeni) nel campione. La variazione di colore viene rilevata attraverso uno strumento (spettrofotometro) e si esprime con un valore numerico attraverso una unità di misura detta assorbanza (o densità ottica).

Per i test quantitativi (ed anche per quelli semi-quantitativi), sorge un problema di interpretazione: occorre stabilire un valore critico o soglia o cut-off, che rappresenta il limite di separazione tra "positività" e "negatività" del test. Ciò corrisponde generalmente alla separazione ammalato/sano.

Lo sketch che segue fornisce in forma grafica la differenza fra un test di tipo qualitativo ed uno di tipo quantitativo.



Ad esempio: al di sotto di quale numero di eritrociti/mm³ un animale può essere giudicato anemico? Al di sopra di quale densità ottica ottenuta con un test ELISA un animale è da ritenere malato?

Due sono i requisiti che un test, sia esso clinico, strumentale o di laboratorio, dovrebbe soddisfare: affidabilità e validità.

Per affidabilità generalmente s'intende la capacità di un test di offrire sempre lo stesso risultato nel corso di misurazioni ripetute. Questa è una caratteristica intrinseca del test e dipende dalla bontà dello strumento e/o dell'operatore. Ma i test sono comunemente utilizzati con l'obiettivo di riconoscere, relativamente a una qualche patologia, i soggetti malati da quelli sani.

La validità (o performance), è proprio la capacità diagnostica del test di distinguere accuratamente in una popolazione i soggetti sani da quelli malati. In questo senso, un test si definisce ideale se distingue perfettamente tutti i soggetti sani dai malati.

Il pannello superiore della figura 1 rappresenta la distribuzione di un test ideale nei sani e nei malati. In questo caso le due distribuzioni sono separate ed è facile individuare sull'asse delle ascisse il valore di cut-off capace di distinguere con precisione assoluta le due popolazioni: tutti i soggetti sani hanno valori del test inferiori al cut-off e risultano negativi al test ("veri negativi", VN), mentre tutti i soggetti malati hanno valori del test superiori al cut-off e risultano positivi al test ("veri positivi", VP).

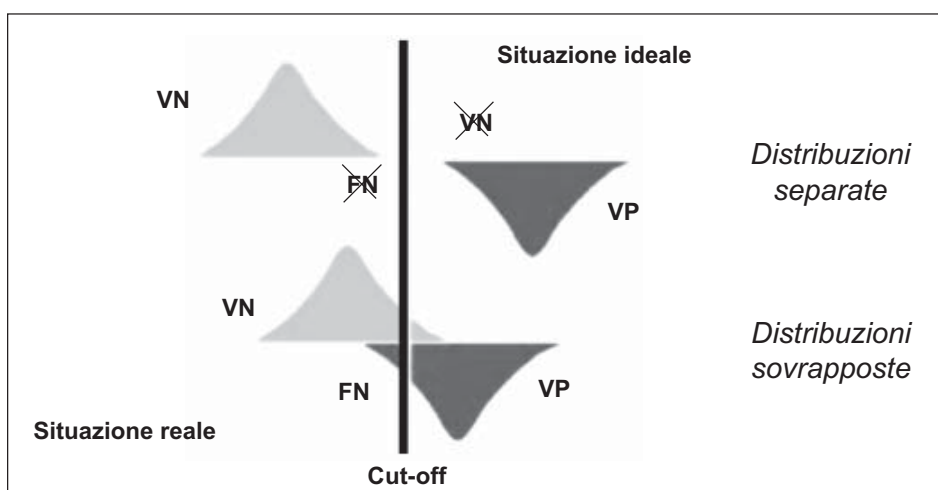


Fig.1: Test ideale, sulla prima riga e test reale sulla seconda.

In realtà accade che, quando si sottopone una popolazione ad una procedura diagnostica, per un certo valore di cut-off, non tutti i soggetti malati risulteranno positivi al test, così come non tutti i soggetti sani risulteranno negativi. Questo problema genera incertezza nell'interpretazione di un test perché, esiste una zona di sovrapposizione fra le distribuzioni dei risultati del test sani/ammalati. (Figura 1).

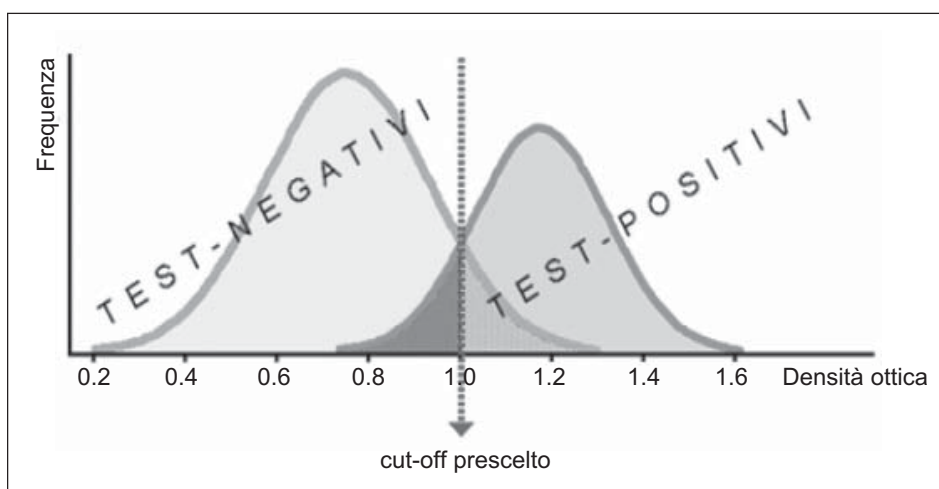


Fig. 2: Riproposizione del pannello inferiore della figura 1.

Si avrà sempre un certo numero di soggetti sani che risulteranno positivi al test (“falsi positivi”, FP) e, simmetricamente, un certo numero di soggetti malati che il test non riuscirà ad identificare come tali, e pertanto saranno erroneamente classificati come “sani” (“falsi negativi”, FN).

In pratica è impossibile individuare sull'asse delle ascisse un valore di cut-off che consenta una classificazione perfetta, ossia tale da azzerare sia i falsi positivi che i falsi negativi. Quindi, un test diagnostico è tanto più accurato quanto più è piccola la probabilità che esso produca dei “falsi”.

É possibile rappresentare questo tipo di situazione utilizzando una tabella di contingenza di tipo 2x2, che confronta l'output del test in esame con il vero stato dei soggetti (Tabella 1).

Il vero stato dei soggetti può già essere noto in partenza (ad esempio, nel caso di una malattia infettiva, saggiando gruppi di individui sicuramente esenti dall'infezione, i cosiddetti *Specific Pathogen Free* ed altri sottoposti ad infezione sperimentale) oppure può essere stabilito per mezzo di un test di riferimento dotato di attendibilità maggiore.

E' il cosiddetto "test aureo" o "golden test" o "gold standard", possibilmente basato su un principio biologico diverso rispetto al test in esame.

In genere i *golden test* presentano alcuni svantaggi (es. difficile somministrazione, rischio per il soggetto da testare, costo elevato, ecc.) che li rendono inapplicabili di routine o in operazioni di screening.

Tab. - 1 Rappresentazione su una tabella 2x2 della distribuzione di una popolazione in base ai risultati di un test.

M+ = Malati; M- = sani ;

T+ = test positivo; T- = test negativo, Tm+ = totale dei malati ,

Tm- = totale dei sani,

Tp = totale dei soggetti con test positivo , Tn = totale dei soggetti con test negativo.

Vero Stato

		M+	M-	Totale
Risultato del test	T+	VP	FP	Tp
	T-	FN	VN	Tn
Totale		Tm+	Tm-	N

1.2 VALUTAZIONE DEI TEST DIAGNOSTICI

Per valutare un test, nell'ipotesi più semplice si assume che il *golden test* fornisca risultati perfettamente corrispondenti alla verità.

Sono possibili quattro risultati a seconda della posizione del valore del cut-off:

Se il risultato di un test è positivo e anche il valore vero è positivo, viene chiamato "vero positivo" (VP);

Se invece il risultato del test è positivo ma il valore vero è negativo, viene chiamato "falso positivo" (FP);

Contrariamente, si ottiene un "vero negativo" (VN) quando entrambi, il risultato e il valore vero, sono negativi;

Si ha un "falso negativo" (FN), invece quando il risultato è negativo ma il valore vero è positivo.

La validità del test può essere misurata tramite la proporzione di falsi positivi e negativi: tanto più basse saranno le loro quote, tanto più il test sarà valido.

L'Efficienza chiamata anche accuratezza, fornisce la probabilità di

ottenere un risultato corrispondente alla realtà, poiché mette in rapporto gli esami corretti (veri positivi e veri negativi) con il totale degli esami eseguiti. Quindi, viene calcolata nel seguente modo:

$$\text{Efficienza (o accuratezza)} = \frac{VP + VN}{VP + FP + VN + FN}$$

1.3. CARATTERISTICHE DI UN TEST DIAGNOSTICO

Le misure più comunemente usate per valutare la performance di un test sono la sensibilità e la specificità.

Per sensibilità si intende la capacità (espressa in percentuale) di un test di individuare in una popolazione i soggetti malati: è la probabilità che il risultato di un test sia positivo per una persona veramente malata. Essa è data dalla proporzione dei soggetti realmente malati e positivi al test (veri positivi) rispetto all'intera popolazione dei malati. Viene calcolata con la formula:

$$Se = \frac{VP}{VP + FN}$$

Se si ottiene un valore ≤ 5 il test è poco sensibile (probabilità bassa), se $\geq 0,8$, è ritenuto un test con buona sensibilità. Minore la sensibilità, maggiori i falsi positivi.

La sensibilità è condizionata negativamente dai falsi negativi: pertanto un test molto sensibile dovrà associarsi ad una quota molto bassa di falsi negativi, ovvero di soggetti malati che "sfuggono" all'identificazione attraverso il test.

Un secondo parametro, è dato dalla specificità.

La specificità è la capacità (espressa in percentuale) di un test di identificare come negativi i soggetti nella realtà sani: è la probabilità che il risultato di un test sia negativo per una persona non malata:

$$Sp = \frac{VN}{FP + VN}$$

Come per la *sensibilità*, anche i valori della *specificità* sono ritenuti scadenti se $\leq 0,5$, buoni, se $\geq 0,8$. In questo caso, minore è la specificità maggiore saranno i falsi negativi.

La specificità è influenzata dai 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.

Per calcolare la specificità dovremo fare riferimento esclusivamente al gruppo dei sani ed alla loro distribuzione fra positivi e negativi al test (ovvero la seconda colonna della tab. 1). Un test altamente specifico sarà dunque un test che produrrà una bassa quota di falsi positivi.

È facile verificare che sensibilità e specificità sono due parametri dipendenti, correlati inversamente in funzione del valore di cut-off (Fig. 3).

In altre parole, l'adozione di una soglia che offre un'elevata sensibilità comporta una perdita di specificità e viceversa.

Se il test in questione fosse rappresentato dalla misurazione di una variabile continua, un modo per aumentare la specificità o la sensibilità sarebbe quello di aumentare o diminuire il limite di cut-off.

Spostando la soglia a destra, si avrebbe una riduzione globale dei soggetti positivi al test con un conseguente aumento della quota di falsi negativi, cioè di soggetti realmente malati che vengono identificati come sani. (aumenta la specificità e si riduce la sensibilità).

In maniera analoga, spostando la soglia a sinistra, si avrebbe una riduzione globale dei soggetti negativi al test con conseguente aumento della quota di falsi positivi. (aumenta la sensibilità e si riduce specificità)

Un metodo empirico comunemente utilizzato per la scelta del cut-off consiste nel fissare a priori il valore desiderato di specificità (generalmente > 0.9) e quindi nel calcolare la corrispondente sensibilità del test nella suddetta condizione.

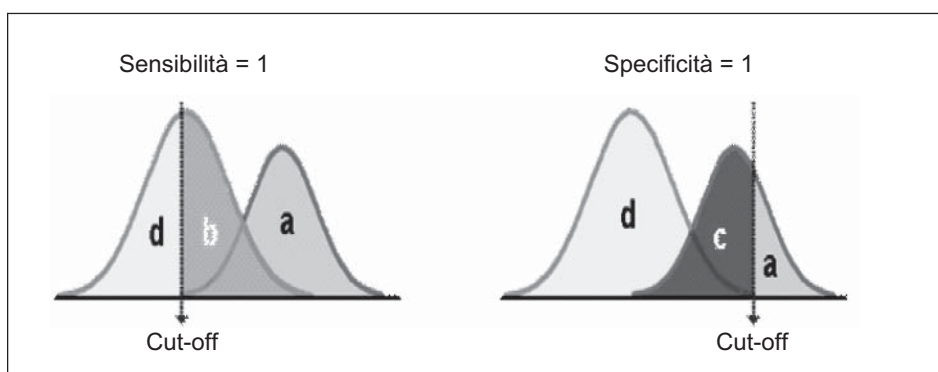


Fig.3: Relazione tra Sensibilità e Specificità

Questo approccio genera tuttavia due effetti collaterali negativi.

Il primo è rappresentato dall'evenienza che il test in questione possa produrre risultati complessivamente migliori attraverso l'adozione di un cut-off diverso da quello scelto.

Il secondo è legato all'impossibilità di effettuare un raffronto affidabile fra la performance di due o più test valutati in base ad un singolo valore di cut-off.

Alle difficoltà ora accennate, è da aggiungere un ulteriore elemento che ostacola sia la scelta del cut-off ottimale per un singolo test che il raffronto tra le performance di test diversi.

Tale elemento è costituito dalla prevalenza della malattia nella popolazione studiata.

1.4. PREVALENZA DELLA MALATTIA E PREDITTIVITÀ

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

Finora, infatti, ci siamo occupati di sensibilità e specificità.

Queste grandezze forniscono la probabilità di identificare 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).

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). Ad esempio: se i soggetti identificati come positivi al test sono 27, e quelli realmente malati sono 25, il VPP sarà dato dal numero di soggetti realmente malati (veri positivi), cioè 25, diviso il numero di 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 modo speculare possiamo calcolare il valore predittivo negativo (VPN), come la quota di veri negativi sul totale dei negativi.

Se, continuando l'esempio precedente, i veri negativi sono 55 su un totale di negativi di 59, che include anche quattro falsi negativi, il 93,2% indicherà la probabilità che un soggetto risultato negativo al test ha di essere nella realtà sano.

La predittività di un test, al contrario di specificità e sensibilità, non è una caratteristica intrinseca del test, ma può dipendere dalla frequenza della malattia nella popolazione.

Ad esempio ipotizziamo di avere una popolazione rappresentata da 400 soggetti, in cui si registra una prevalenza del 5% di malati.

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%.

Supponiamo ora di ripetere la stessa esperienza in una popolazione che presenta una prevalenza di malattia superiore (20%).

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

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

La predittività del test sarà sempre proporzionale alla prevalenza della malattia nella popolazione sottoposta a screening; per aumentarla, 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.

N.B.



Il valore **predittivo positivo** e il valore **predittivo negativo** dipendono da:

- prevalenza della malattia nella popolazione
(= proporzione di soggetti malati)
- sensibilità e specificità dello strumento di screening

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 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.

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, ("caratteristiche operative del ricevitore").

1.5. CURVA ROC

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 misura l'accordo tra il test di interesse e la presenza/assenza di una specifica malattia (così come identificata da un golden standard).

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).

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.

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 (Fig. 4). L'area sottostante alla curva ROC (AUC, acronimo dei termini inglesi "Area Under the Curve") è una misura di accuratezza diagnostica.

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 (1998):

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

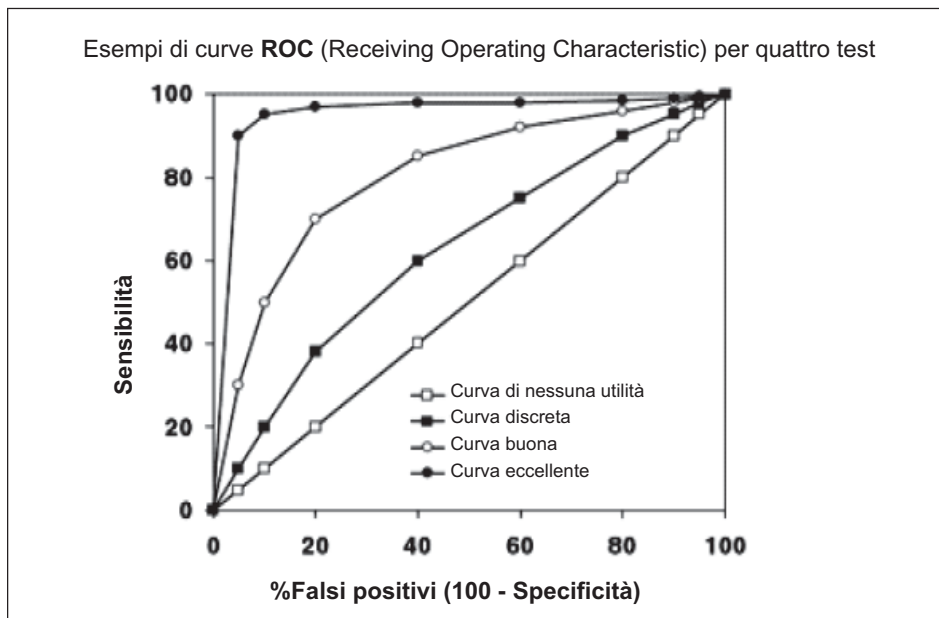


Fig. 4: La curva ROC

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. Le curve ROC possono essere confrontate fra loro con l'uso di un appropriato test statistico disponibile in quasi tutti i software in commercio.

2. LE INDAGINI EPIDEMIOLOGICHE DESCRITTIVE, ANALITICHE, RETROSPETTIVE E SPERIMENTALI

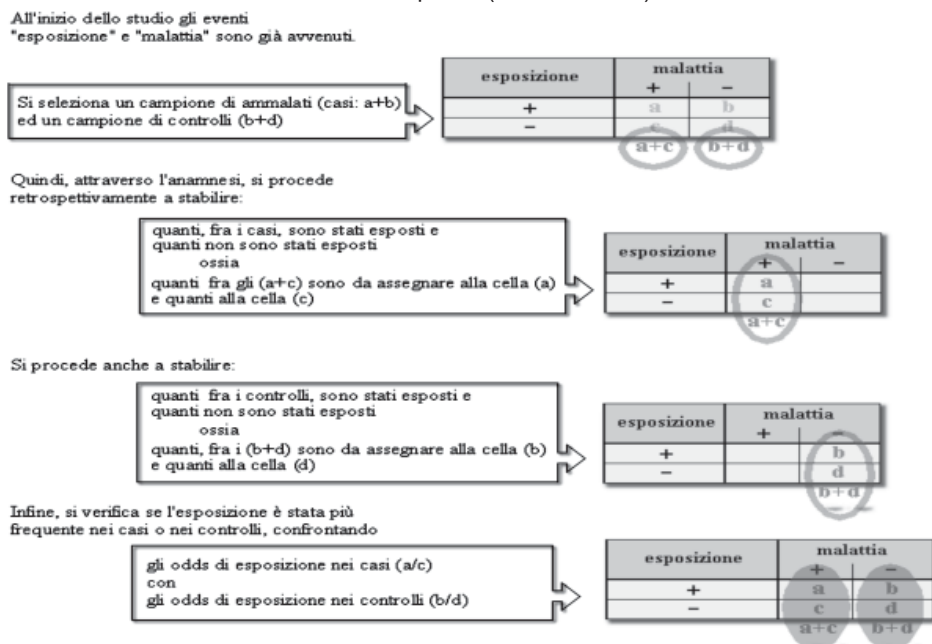
2.1 STUDI RETROSPETTIVI (O STUDI CASO-CONTROLLO).

Negli studi retrospettivi, lo sperimentatore inizia raccogliendo i cosiddetti “casi”, ossia gli individui che presentano la malattia in studio. Nella tabella i casi sono rappresentati dal totale degli individui ($a + c$). Viene anche scelto un adatto gruppo di paragone (o di controllo) che comprenderà individui sani ($b + d$). A questo punto, attraverso una accurata anamnesi su tutti i soggetti in studio, si stabilisce come gli ammalati ($a + c$) debbano essere assegnati alle celle a e c . Analogamente si stabilisce quanti, fra i controlli, debbano essere assegnati alle celle b e d .

La tabella risulta ora completata, e si può impostare l'analisi, confrontando gli odds di esposizione nei casi (a/c) con gli odds di esposizione nei controlli (b/d) (confronto fra colonne).

La struttura di uno studio retrospettivo è riassunta nello schema che segue.

Studi retrospettivi (caso-controllo)



Come già detto, gli studi retrospettivi sono basati su gruppi costituiti da individui che, già all'inizio dell'esperienza, sono noti come “casi” o

“controlli”; per questo gli studi di questo tipo sono detti anche “studi caso/controllo”.

Uno studio retrospettivo ha il vantaggio di fornire un risultato relativamente rapido, in quanto all’inizio dello studio il tempo necessario all’accadimento degli eventi è già trascorso. Un altro punto a favore degli studi retrospettivi, rispetto a quelli prospettivi, è la applicabilità ad indagini su malattie rare, per le quali i casi possono essere raccolti retrospettivamente anche da ospedali.

é però da notare che, proprio per la loro stessa natura, gli studi retrospettivi forniscono - in linea di massima - risultati meno affidabili rispetto agli studi prospettivi. Un altro elemento a sfavore degli studi retrospettivi riguarda l’eventualità che si voglia studiare una malattia di breve durata e ad esito generalmente letale. In questa situazione, il “bias” potrebbe derivare dalla selezione dei casi, i quali sarebbero costituiti dai pochi soggetti sopravvissuti, certamente non rappresentativi della popolazione dei malati. Si potrebbe addirittura verificare il paradosso di considerare come causa di una malattia un fattore che in realtà è protettivo: proprio quello stesso fattore che ha consentito la sopravvivenza dei casi e che senz’altro risulterà fortemente associato ad essi.

2.2 STUDI PROSPETTIVI (O STUDI DI COORTE)

Uno studio prospettivo inizia selezionando due gruppi, entrambi costituiti da soggetti sani: un gruppo comprende soggetti che sono stati esposti alla presunta causa (o lo saranno in futuro), e l’altro soggetti che non sono stati esposti (e non lo saranno).

Quindi, i soggetti presi in considerazione vengono seguiti nel tempo e andranno a distribuirsi nelle colonne degli ammalati o dei sani.

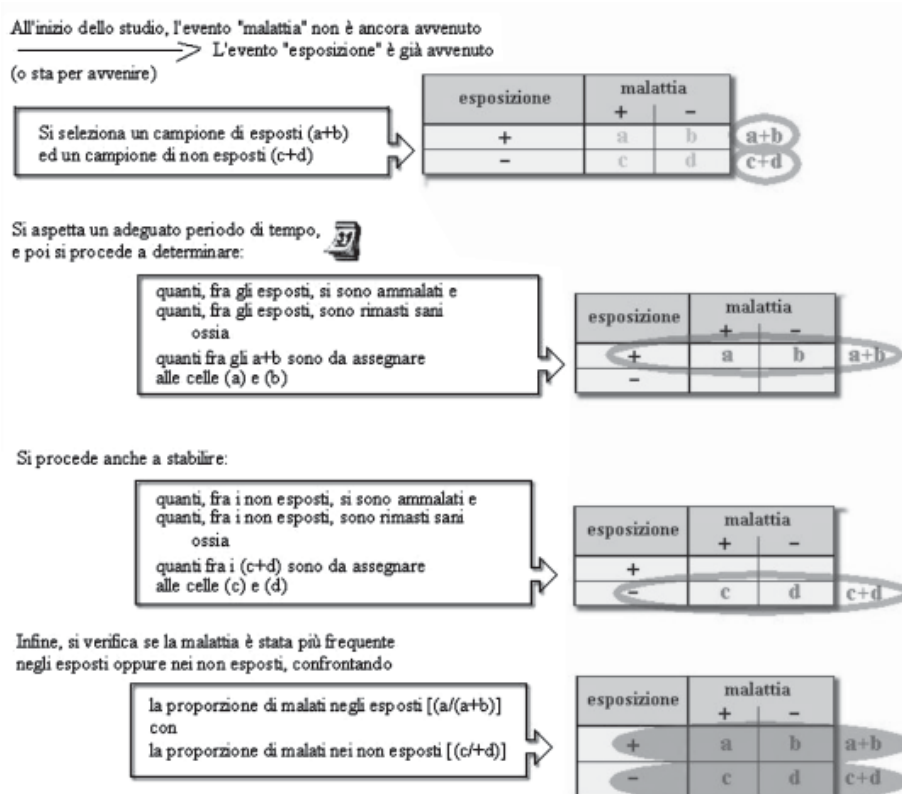
In questo modo, alla fine dell’esperimento, la tabella risulterà completata con i valori a, b, c, d.

Si prosegue effettuando la analisi dei dati, confrontando la proporzione di malati tra gli esposti $[a/(a+b)]$ con la proporzione di malati tra i non esposti $[c/(c+d)]$

Lo studio prospettivo (detto anche “di coorte”) ha lo svantaggio di richiedere più tempo, in quanto si deve seguire nel tempo la comparsa degli eventi. Inoltre, esso non è applicabile a malattie rare per la difficoltà nel reperimento di un numero di casi sufficiente.

Gli studi prospettivi sono superiori a quelli retrospettivi perché meno

Studi prospettivi (di coorte)



soggetti ad “errori sistematici”, in quanto essi non dipendono da dati raccolti in precedenza magari con modalità poco affidabili. Infatti, il ricercatore è in grado di valutare personalmente la qualità dei dati raccolti, soprattutto per quanto riguarda l’esposizione, cosa che invece è sempre un po’ aleatoria negli studi retrospettivi.

Un altro punto a favore degli studi prospettivi è che essi possono fornire una stima della incidenza (ossia del numero di nuovi casi che compaiono in un dato tempo) della malattia e possono essere utilizzati per studiare l’effetto di determinanti rari.

Nella figura seguente è illustrata schematicamente la differenza fra studi prospettivi e retrospettivi; in particolare, viene evidenziato il diverso momento di inizio dell’osservazione della popolazione in rapporto alla comparsa di malattia.



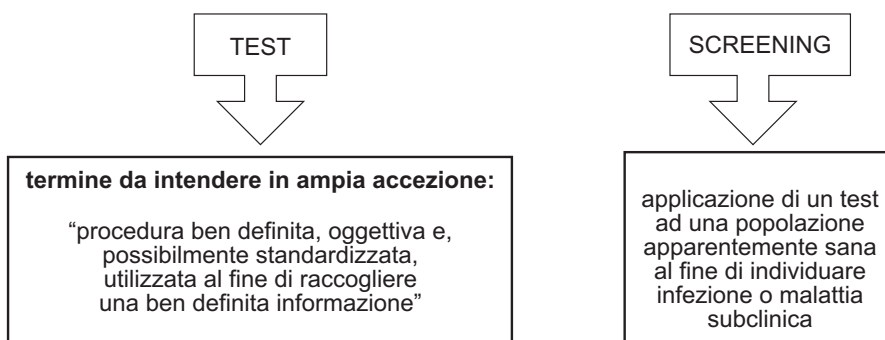
2.4 UTILIZZO DI UN TEST PER LO SCREENING DI POPOLAZIONI.

Molti pensano che un test sia una specifica procedura di laboratorio (es. test della glicemia, colesterolo ecc.) o una procedura che viene valutata attraverso uno strumento e che pertanto è meno dipendente dal giudizio soggettivo dell'esaminatore.

Questa opinione non è certamente sbagliata, ma per i nostri scopi è piuttosto limitativa: infatti per

“test” intendiamo una qualsiasi ben definita procedura, oggettiva e possibilmente standardizzata, che viene utilizzata al fine di raccogliere una ben definita informazione.

Un “test” di screening è un test che viene applicato ad una popolazione apparentemente sana (o a stato sanitario ignoto) soggetta ad una probabilità (“rischio”) più o meno elevata di presentare la malattia considerata. In questo modo, tutti gli individui della popolazione vengono sottoposti al controllo diagnostico. In genere, i test di screening sono procedure poco costose e di rapido e semplice impiego.



Anche se azioni di screening e procedimenti diagnostici possono essere effettuate impiegando lo stesso “test”, tuttavia lo screening differisce dalla diagnosi:

contrariamente allo screening, nel procedimento diagnostico l’eventuale impiego di un “test” viene attuato su soggetti ammalati, cioè che mostrano sintomi clinici che, in una qualche misura, fanno sospettare la presenza di quella malattia; le azioni di screening, anche se implementate in popolazioni nelle quali la malattia è presumibilmente presente, prevedono l’applicazione del test su tutti gli individui della popolazione, indipendentemente dal loro stato di salute; poiché il valore predittivo di un test dipende dalla prevalenza della malattia, ne consegue che la performance del test sarà meno soddisfacente in caso di screening rispetto al caso in cui lo stesso test venga utilizzato a scopo diagnostico. Infatti, è evidente che la prevalenza della malattia fra gli individui che mostrano segni clinici sarà superiore rispetto alla prevalenza considerata nella popolazione nel suo complesso.

Per meglio comprendere la definizione di sensibilità e specificità, è utile riferirsi alla tabella a doppia entrata (Tab. 1) riportata in precedenza, che qui di seguito viene riproposta in maniera più analitica per le successive applicazioni.

esito test ↴	malati (D+)	sani (D-)	TOT.	
+ (T+)	a	b	a+b	totale test-positivi
- (T-)	c	d	c+d	totale test-negativi
TOT.	a+c	b+d	n	
	totale ammalati	totale sani		

Come avevamo già discusso la sensibilità e la specificità sono due misure che vengono impiegate per valutare la capacità di individuare, fra gli individui di una popolazione, quelli provvisti del “carattere” ricercato e quelli che invece ne sono privi. In pratica, per i nostri scopi, il “carattere” è rappresentato quasi sempre dalla malattia o dall’infezione.

3. APPLICAZIONI DELLE METODOLOGIE UTILIZZATE

3.1 TEOREMA DI BAYES E VALORI PREDITTIVI POSITIVI E NEGATIVI

Mentre la sensibilità e la specificità di un test diagnostico sono parametri determinanti per i clinici per valutare l'efficacia dei test nel discriminare i soggetti sani da quelli malati. Tale valutazione viene effettuata "ex- ante" e cioè prima che il test venga eseguito. Secondo l'ottica del paziente, la questione non è "Ogni quante volte il test rivela realmente la malattia?" ma piuttosto, "Adesso che i risultati dei test sono noti, qual è la possibilità che io abbia la malattia?". Il paziente vuole conoscere la probabilità condizionale che è l'inverso della sensibilità. Se noi usiamo Dx per intendere un risultato positivo e D per intendere la presenza attuale di malattia, il paziente non è preoccupato per la sensibilità, $\text{Prob}(Dx/D)$ ma per VPP, $\text{Prob}(D/Dx)$, vale a dire la possibilità che sia presente la malattia poiché il test era positivo.

Entrambe sensibilità e specificità sono considerate caratteristiche dei test invariate e innate. Al contrario, il VPP e il VPN dipendono non soltanto dalla sensibilità e dalla specificità ma anche dalla prevalenza della patologia, $\text{Prob}(D)$. Essi possono essere combinati usando la formula di Bayes che correla il VPP, $\text{Prob}(D/Dx)$, alla sensibilità, $\text{Prob}(Dx/D)$, alla specificità, $1 - \text{Prob}(Dx/D\emptyset)$, e alla prevalenza, $\text{Prob}(D)$.

$$\text{VPP} = \frac{\text{Sensibilità} \times \text{Prevalenza}}{(\text{Sensibilità} \times \text{Prevalenza}) + (1 - \text{Specificità}) \times (1 - \text{Prevalenza})}$$

In questa equazione, il denominatore è il numero totale dei risultati positivi attesi nella popolazione, mentre il numeratore è il numero dei risultati positivi che si accompagnano all'attuale presenza di patologia.

Supponiamo di considerare un test diagnostico per screening che ha entrambi il 95% sia di sensibilità (positivo nel 95% di tutti i casi dove è presente la patologia) che di specificità (negativi nel 95% di tutti i casi dove la malattia è assente). Al posto di usare l'esercizio laborioso di costruire tabelle, la formula di Bayes ci dà direttamente il VPP. Per la prevalenza del 10%, $\text{Prob}(D)=0.10$, noi abbiamo $\text{VPP} = (0.95 \times 0.10) / (0.95 \times 0.10) + (0.05 \times 0.90) = 0.095 / (0.095 + 0.045) = 0.679$.

3.1.2. ESEMPI DI TEST DI SCREENING

I principali test di screening, di prevenzione soprattutto oncologica sono i seguenti:

- a) *Screening della mammella*, utile per diagnosticare il tumore della mammella in fase iniziale, che rappresenta il 28% dei casi diagnosticati nella popolazione femminile ed è la neoplasia più frequente tra le donne. La mammografia è una radiografia del seno ed è un esame semplice e indolore. Le donne tra i 45 e i 49 anni vengono chiamate annualmente, mentre quelle tra i 50 e i 69 anni ogni 2. La mammografia è il mezzo più efficace e sicuro per scoprire i tumori al seno in fase iniziale, con ottime possibilità di recupero completo. Aiuta ad individuare piccole modifiche del seno prima che appaiano altri segni; per questo è bene eseguirla ogni 2 anni anche in assenza di disturbi.

La mammografia è un test affidabile e solo in rari casi non è in grado di trovare il tumore.

- b) *Tumore della cervice uterina*, utile per diagnosticare il tumore della cervice uterina in fase precoce. Il carcinoma del collo dell'utero è la seconda forma tumorale più diffusa tra le donne al di sotto dei 50 anni ed è al primo posto, in molti paesi in via di sviluppo, nella fascia di età compresa tra i 35 e i 45 anni. In Italia ci sono ogni anno, circa 3500 nuovi casi ed il cervicocarcinoma risulta al quinto posto, per incidenza, tra la popolazione femminile. Lo screening della cervice uterina, da eseguire ogni 3 anni, ha ridotto la mortalità del 50% negli ultimi venti anni ed è rivolto alle donne di età compresa tra i 25 e i 64 anni.

Il test di screening per la prevenzione di questo tipo di tumore è il pap-test e serve a scoprire eventuali alterazioni che possono trasformarsi in tumore. L'ostetrica o il Ginecologo prelevano dal collo dell'utero con una spatola un piccolo campione di cellule che viene analizzato al microscopio, per un'eventuale presenza ad esempio del L'HPV è un virus (il papilloma virus) che colpisce la maggior parte delle donne almeno una volta nella vita e dal 2007 in Italia, ed è stata autorizzata la vaccinazione contro tale virus.

- c) *Tumore del colon retto*, utile per diagnosticare il tumore del colon retto in fase precoce. Il tumore del colon retto, ultimo tratto dell'intestino, è la seconda neoplasia più frequente nelle donne e la terza nei maschi. E' possibile scoprire in tempo l'insorgenza di un tumore grazie ad un test

semplice ed affidabile, che svela nelle feci la presenza di sangue occulto, cioè non visibile ad occhio nudo. Il test è associato a una riduzione della mortalità per tumore del colon retto di almeno il 20%. La presenza di sangue nelle feci spesso è dovuta a cause banali (emorroidi) ma può anche essere il primo segnale di un tumore o di un adenoma (polipo tumorale benigno) del colon retto e per diagnosticarlo sono necessari ulteriori accertamenti (colonscopia). Gli adenomi diagnosticati in fase iniziale, possono essere curati efficacemente, evitando la trasformazione in neoplasia. Su 100 persone che eseguono il test 5 possono risultare positive per la presenza di sangue nelle feci. In questo caso, è necessario eseguire la colonscopia. Tale esame serve a controllare le pareti interne del colon retto e si esegue con il colonscopio, uno strumento munito di una telecamera che, illuminando le pareti interne del colon, permette di individuare eventuali alterazioni.

3.1.3 IMPIEGO DEL TEOREMA DI BAYES NEI TEST DIAGNOSTICI

Il teorema di Bayes viene impiegato per calcolare la probabilità di una causa che ha scatenato l'evento verificato, come una determinata patologia.

Con questo teorema si può calcolare la probabilità che una certa persona soffra della malattia per cui ha eseguito il test diagnostico (nel caso in cui questo sia risultato negativo) o viceversa non sia affetta da tale malattia (nel caso in cui il test sia risultato positivo), conoscendo la frequenza con cui si presenta la malattia e la percentuale di efficacia del test diagnostico.

Proviamo adesso a pensare agli eventi H come le cause che determinano l'evento O. Allora, se si è verificato O, con quale probabilità la causa è H? In altre parole si vuole conoscere la probabilità $P(H|O)$:

$$P(H|O) = \frac{P(O|H) \times P(H)}{P(O|H) \times P(H) + P(O|-H) \times P(-H)}$$

con la seguente simbologia:

$P(H)$ = probabilità che sia vera l'ipotesi da testare;

$P(-H)$ = probabilità che l'ipotesi H sia falsa;

$P(O|H)$ = probabilità che l'osservazione O sia riscontrata se l'ipotesi H è vera

$P(O|-H)$ = probabilità che l'osservazione O sia riscontrata se l'ipotesi H è falsa

$P(H|O)$ = probabilità a posteriori che l'ipotesi H sia vera se l'osservazione O è positiva;

Nel caso che ci interessa, l'utilizzo del Teorema di Bayes è abbastanza semplice: basta applicare la formula appropriata, ossia quella che consente di ottenere, conoscendo Se e Sp del test e la Prevalenza, il VPP (valore predittivo positivo) cioè la probabilità (espressa in percentuale) che un soggetto positivo al test sia effettivamente malato. Dalle definizioni di prevalenza, sensibilità e specificità, ne deriva che:

$$P(H) = \text{prevalenza};$$

$$P(O|H) = \text{sensibilità};$$

$$P(O|\bar{H}) = 1 - \text{specificità};$$

poiché la specificità indica la probabilità che il test sia negativo per i soggetti sani, per ottenere i casi in cui il test è negativo nei soggetti malati, occorre calcolare il complemento a 1.

3.1.4 ESEMPI APPLICATIVI

Un test diagnostico, effettuato su 250 soggetti affetti da una data patologia, ha esito negativo in 15 casi. Lo stesso test provato su un campione costituito da 150 soggetti non affetti dalla patologia è risultato positivo in 18.

Riassumiamo i dati in una tabella a due entrate, come quella descritta precedentemente:

Test	Malati	Non malati	TOT
+	VP = 235	FP = 18	253
-	FN = 15	VN = 132	147
TOT	250	150	400

$$\text{Sensibilità} = \Rightarrow = 0.94 \rightarrow 94\%$$

$$\text{Specificità} = \frac{VN}{FP + VN} \Rightarrow \frac{132}{150} = 0.88 \rightarrow 88\%$$

$$\text{Valore predittivo positivo} = \frac{VP}{VP + FP} \Rightarrow \frac{235}{235 + 18} = 0.9288 \rightarrow 93\%$$

$$\text{Valore predittivo negativo} = \frac{VN}{FN + VN} \Rightarrow \frac{132}{15 + 132} = 0.8979 \rightarrow 90\%$$

$$\text{Efficienza} = \Rightarrow \frac{VP + VN}{TOT} = \frac{235 + 132}{400} = 0.9175 \rightarrow 92\%$$

Il test è in grado di identificare il 94 % delle situazioni patologiche e di discriminare i non malati nell'88 % dei casi; un risultato positivo corrisponde a un malato con una probabilità del 93 % e i risultati complessivi del test mostrano un' affidabilità del 92%.

Consideriamo un esempio con applicazione del teorema di Bayes:

La probabilità che una persona con più di cinquant'anni senza sintomi abbia un cancro colon-rettale è dello 0,3%. Se una persona ha un cancro colon-rettale, c'è una probabilità del 50% che il test del sangue occulto nelle feci risulti positivo (sensibilità = 50%). La specificità del test è del 97%; questo significa che c'è una probabilità del 3% che il test del sangue occulto nelle feci risulti comunque positivo anche in assenza di un cancro colon-rettale.

Considerando una persona ultracinquantenne, senza sintomi, che risulti positiva al test del sangue occulto nelle feci, qual è la probabilità che abbia veramente un cancro colon-rettale?

Dai dati risulta che:

$P(H) = 0,3\% = 0,003 = 3$ persone malate ogni 1000 = Prevalenza

$P(-H) = 1 - P(H) = 1 - 0,003 = 0,997 = 997$ persone sane ogni 1000;

$P(O|H) = 50\%$ = probabilità di riscontrare O se l'ipotesi è vera = Sensibilità

$P(O|-H) = 1 - \text{specificità} = 1 - 0,97 = 0,03$ probabilità di riscontrare O se l'ipotesi è falsa = Specificità

Applicando la formula di Bayes, risulta:

$$P(H|O) = \frac{P(O|H) \times P(H)}{P(O|H) \times P(H) + P(O|-H) \times P(-H)}$$

$$= 0.048 = 4.8\% \quad \frac{0.5 \times 0.003}{0.5 \times 0.003 + 0.03 \times 0.997}$$

$$\frac{0.0015}{0.0314}$$

Questo significa che su un campione di 100 persone, la probabilità di un vero positivo è circa 5 casi, mentre i falsi positivi sono 95.

Illustriamo un ulteriore esempio è relativo alla campagna di screening sulla tubercolina. Si vuole iniziare una campagna di screening per la individualizzazione dei soggetti positivi alla tubercolina.

La popolazione su cui fare lo screening è costituita da 10000 individui.

Come in precedenza riassumiamo i dati in nostro possesso con la seguente tabella a doppia entrata:

Test	Malati	Non malati	TOT
Tubercolina +	VP = 96	FP = 594	690
Tubercolina -	FN = 4	VN = 9306	9310
TOT	100	150	10000

Sensibilità = $VP/VP+FN = 96/100 = 0.96 = 96\%$

Specificità = $VN/FP+VN = 9306/9900 = 0.94 = 94\%$

Prevalenza = $VP+ FN/ TOT = 96+4/10000 = 100/10000 = 0.01 = 1\%$

Inoltre, osserviamo che:

$P(H) = \text{Prevalenza} = 0.01$

$P(O|H) = \text{Sensibilità} = 0.96$

$P(O|-H) = 1 - \text{Specificità} = 1 - 0.94 = 0.06$

$P(-H) = 1 - \text{Prevalenza} = 1 - 0.01 = 0.99$

Applicando la formula del teorema di Bayes avremo:

$$\begin{aligned}
 P(H | O) &= \frac{P(O | H) \times P(H)}{P(O | H) \times P(H) + P(O | -H) \times P(-H)} \\
 &= \frac{\text{Se} \times \text{preval}}{(\text{Se} \times \text{preval.}) + [(1 - \text{Sp}) \times (1 - \text{Preval.})]} \\
 &= \frac{0.96 \times 0.01}{0.96 \times 0.01 + 0.6 \times 0.99} = \\
 &= \frac{0.0096}{0.069}
 \end{aligned}$$

Se calcoliamo il VPP, con la formula tradizionale:

$$VPP = VP / VP + FP = 96/96+594 = 96/690 = 0.139 = 13.9\%$$

Notiamo che confrontando il risultato del VPP con quello del teorema di Bayes affermiamo che mediante il teorema ci consente di determinare il

valore predittivo positivo (VPP) di un test di screening, cioè ci permette di calcolare la probabilità che un soggetto positivo al test è veramente malato.

TEOREMA DI BAYES



VPP

(PROBABILITÀ CHE UN SOGGETTO POSITIVO AL TEST È VERAMENTE MALATO)

4 CONCLUSIONI

Nel lavoro, dedicato agli strumenti statistici utilizzabili in campo medico, abbiamo evidenziato i concetti e le problematiche fondamentali collegate alla diagnostica e alle altre principali metodologie utilizzate. Abbiamo, in particolare visto come un test diagnostico sia caratterizzato da caratteristiche diverse, spesso in contrasto tra loro, come sensibilità e specificità. Abbiamo, anche, visto, come la prevalenza della malattia sia un fattore non trascurabile sui risultati di un test.

Abbiamo anche evidenziato il ruolo importante del teorema di

Bayes sulla determinazione della probabilità che il risultato di un test diagnostico sia collegato alla realtà.

In definitiva, il messaggio complessivo che abbiamo voluto trasferire è quello di porre la massima attenzione, prima di giungere a conclusioni definitive in un campo tanto delicato per il coinvolgimento di persone, quanto stimolante per i benefici che si possono apportare, studiando e considerando tutte le implicazioni, caso per caso. La metodologia statistica è di valido ausilio perché è in grado di quantificare i rischi collegati ad ogni conclusione diagnostica.

THE DIAGNOSTIC TESTS IN EPIDEMIOLOGY: OUTLINE OF THEORETICAL & METHODOLOGICAL DEVELOPMENTS IN APPLICATION EXAMPLES

Summary

This paper sets out in detail the relation between aspects of epidemiological related to the distribution of diseases and statistical techniques more suitable, such as diagnostic tests, retrospective studies, the ROC curve, allowing clinicians and social workers to make decisions scientifically more consistent for the treatment of patients suffering from diseases of different possible extent and severity. In the work are also illustrated several examples explaining the statistical techniques described.

BIBLIOGRAFIA

- ALONZO T.A., PEPE M.S. (1999),
“Using a combination of reference tests to assess the accuracy of a new diagnostic test, *Statistics in Medicine*”; 18: 2987-3003.
- BAMBER D. (1975), “The area above the ordinal dominance graph and the area below the receiver operating graph, *Journal of Mathematical Psychology*”; 12: 387-415.
- BECKER M.P. (1994), “Analysis of cross-classifications of counts using models for marginal distributions: an application to trends in attitudes on legalized abortion, *Sociological Methodology*”; 24: 229-65.
- BEGG C.B., METZ C.E. (1990), “Consensus diagnosis and gold standard, *Medical Decision Making*”; 10: 29-30.
- BOTTARELLI E., PARODI S (2003), “Un approccio per la valutazione della validità dei test diagnostici: le curve ROC (Receiver Operating Characteristic)”, *Ann. Fac. Medic. Vet. Di Parma (Voi. XXIII)*: 49-68.
- BRANSCUM A.J., JOHNSON W.O., HANSON T.E., GARDNER I.A. (2008), “Bayesian semiparametric ROC curve estimation and disease diagnosis, *Statistics in Medicine*”; 27: 2474-2496.
- CAVALLI SFORZA L. (1965) *Analisi statistica per medici e biologi*. Boringhieri, Torino.
- CUTTINI M., DI CIOMMO V., RAVA' L., TOZZIA E. *Glossario di epidemiologia e biostatistica* (<http://www.ospedalebambinogesu.it/Portale2008/>)

- FASOLO A., Dizionario di Biologia, UTET, Torino, 2004
- FRIEDMAN G.D. (1995) Epidemiologia per discipline bio-mediche. Edizione italiana a cura di L. Sebastiano Annichiarico Petruzzelli, McGraw-Hill Libri Italia, Milano, IV ed.
- GALLIA., GREPPI G.F. Bioinformatica - Statistica. <http://www.bioinfovet.unimi.it/binfovet/bioin/biostat/indibiom.htm>.
- GALLI A., GREPPI G.F., P. Roncada á Epidemiologia. <http://www.bioinfovet.unimi.it/binfovet/bioin/biostat/epidem.htm>
- GLANZ S.A. (1994) Statistica per discipline bio-mediche. Ed. italiana a cura di A. Marinoni, V. Balestra, S. Favilli. McGraw-Hill Libri Italia, Milano, III ed.
- JEKEL J.F., ELMORE J.G., KATZ D.L. (1996) Epidemiologia, Biostatistica e Medicina preventiva. EDISes, Napoli
- KIRKWOOD B.R., STERNE J.A.C. (2005) Essential Medical Statistics. 2nd ed., Blackwell Publishing.
- LANG T.A., SECIC M. (2006) How To Report Statistics In Medicine. 2nd ed., American College of Physiscians, Philadelphia, USA,
- NALDI L. (1998), Metodi di ricerca clinica, ISED, Brescia
- NORMAN G.R., STREINER D.L. (2000) Biostatistica. Casa Editrice Ambrosiana, Milano.
- PICCOLO D. (2000) Statistica. Seconda edizione. Il Mulino, Bologna.
- PROVENZANO F, D'ARRIGO G, ZOCCALIC, TRIPEPI G. (2011) "La regressione logistica nella ricerca clinica". G Ital Nefrol; 28: 210-3.
- SELVIN S. (1996) Statistical analysis of epidemiological data. Oxford University Press, New York, 2nd ed.
- SOLIANI L. (2004) . Statistica applicata alla ricerca e alle professioni scientifiche.
- SWETS JA. (1998)
"Measuring the accuracy of diagnostic systems.
Science"; 240: 1285-93.
- YOU DEN WJ. (1950) "Index for rating diagnostic tests". *Cancer*; 3: 32-5.