

(IN)UTILITÀ CLINICA DEI TEST DIAGNOSTICI

V.M. TRANQUILLO¹, L. VENCO², C. GENCHI³

¹Osservatorio Epidemiologico Veterinario Regionale della Lombardia
Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (Brescia)

²Medico Veterinario LP Clinica Citta di Pavia (PV)

³Università di Milano (Milano)

Riassunto

Un test diagnostico può ritenersi utile da un punto di vista clinico quando è in grado di aggiungere una quota rilevante d'informazioni al processo diagnostico. L'utilità clinica non va confusa con l'accuratezza di un test, misurata mediante i parametri di Sensibilità e Specificità, in quanto è funzione della probabilità pre-test di malattia. Da questo punto di vista l'utilità clinica è specifica per ogni singola situazione che il clinico si trova a dover risolvere nella sua pratica quotidiana. Caso per caso potrà selezionare il test diagnostico i cui risultati positivi o negativi determineranno un rilevante cambiamento del giudizio diagnostico. In quest'ottica, indipendentemente dall'accuratezza diagnostica, il clinico potrà giudicare i differenti test come utili o inutili clinicamente. In questo lavoro si illustrano i principi teorici che supportano la definizione di utilità clinica dei test diagnostici e si forniscono alcuni esempi tratti dalla pratica clinica.

Summary

The clinical utility of a diagnostic test depends on the quantity and quality of information it is able to add to the diagnostic process. The clinical utility, which is a function of the pre-test probability of disease, must not be confused with the accuracy, which is measured by the parameters of Sensibility and Specificity. The clinical utility of a given test is specific for each single situation that the clinician has to solve in his practice. Case by case, the clinician should be able to select the diagnostic test whose results (positive or negative) will be able to significantly influence the diagnostic process. In this way, independently from the accuracy, the clinician will be able to judge the different tests as clinically useful or non-useful. This article deals with the theoretical basis of the definition of clinical utility of a diagnostic test. Some practical examples are discussed.

INTRODUZIONE

Il processo diagnostico, sia esso l'esame clinico-fisico o l'esecuzione di un test di laboratorio o strumentale, è un processo probabilistico. Di norma il clinico, sulla scorta della storia del paziente e dell'esame clinico, formula una lista di ipotesi diagnostiche: in tal modo, anche se inconsciamente, egli attribuisce diversi gradi di certezza alle differenti ipotesi generate¹. Questo grado d'incertezza altro non è che il livello di probabilità associato alle diverse ipotesi. L'esclusione o la conferma delle ipotesi diagnostiche in genere viene affidata alla selezione di uno o più test diagnostici di laboratorio o ad esami strumentali. L'utilizzo della diagnostica di laboratorio e strumentale

può essere quindi vista come un supporto per ridurre l'incertezza diagnostica associata alle diverse ipotesi formulate. Un concetto fondamentale che verrà sottolineato in questo lavoro è che un test dovrebbe essere scelto, e i suoi risultati adeguatamente interpretati, in modo che influenzino significativamente la percezione che il clinico ha della probabilità di malattia prima dell'esecuzione del test. In questo senso distingueremo l'utilità diagnostica di un test dalla sua accuratezza, spesso a torto considerata come una misura di utilità clinica. Per fare questo il clinico deve abituarsi a tradurre la sua percezione di probabilità in termini numerici e quindi non esprimere la sua "certezza o incertezza" diagnostica secondo frasi quali: "molto probabile, poco probabile, improbabile", ma in valori da 0 (impossibile) a 1 (certo), oppure in termini percentuali da 0%-a 100%. L'obiettivo di questo lavoro è quello di fornire gli elementi teorici e gli strumenti pratici per una corretta valutazione dell'utilità clinica dei test diagnostici.

¹Articolo ricevuto dal Comitato di Redazione il 2/1/2004 ed accettato per pubblicazione dopo revisione il 26/4/2004".

PRINCIPI TEORICI (IL TEOREMA DI BAYES)

Il problema della diagnosi, intesa come la valutazione della probabilità che un soggetto abbia o meno una determinata malattia considerando la presenza/assenza di segni, sintomi, esiti di esami di laboratorio e strumentali, è una particolare applicazione di un teorema fondamentale della teoria della probabilità: il teorema di Bayes^{2,3}. Si tratta di un teorema a carattere generale che trova applicazione in molti campi della ricerca scientifica. In sostanza questo teorema afferma che la probabilità che si verifichi un determinato evento è funzione della probabilità soggettiva a priori che è attribuita all'evento, e dei dati oggettivi sperimentali. Applicando questo teorema alle problematiche della diagnosi, possiamo affermare che la probabilità che un soggetto abbia o no una determinata condizione è una funzione della probabilità che soggettivamente il clinico attribuisce a priori, prima dell'esecuzione del test, e del potere informativo del test diagnostico (dato sperimentale).

Gli elementi quindi che il clinico deve conoscere per valutare l'utilità di un test diagnostico sono:

- la probabilità di malattia prima dell'esecuzione di qualsiasi test diagnostico (*probabilità pre-test*);
- il *potere informativo* del test diagnostico;
- la *probabilità post-test* della malattia.

Come accennato in precedenza, prima di sottoporre ad un test o ad altre procedure diagnostiche un paziente, il clinico formula una o più ipotesi diagnostiche e attribuisce a queste un livello di certezza in termini probabilistici. Questa probabilità è definita come probabilità *pre-test* (a priori o probabilità soggettiva) in quanto è stimata (soggettivamente) sulla base delle conoscenze della diffusione della patologia nella popolazione di riferimento (prevalenza), della storia del paziente e dall'esame clinico-fisico. L'importanza di questa informazione è tale da condizionare largamente l'utilità diagnostica dei test indipendentemente dall'accuratezza del test utilizzato.

Per aiutare il clinico a pensare in termini probabilistici possiamo utilizzare la Figura 1 in cui la probabilità a priori è indicata da un punto (P) su una linea continua che va dalla situazione di assenza della malattia (0 oppure 0%) alla certezza della presenza della malattia (1 o 100%).

Un test negativo determinerà uno spostamento della probabilità verso lo zero, al contrario un test positivo sposterà verso il 100% la probabilità di malattia. L'entità dello

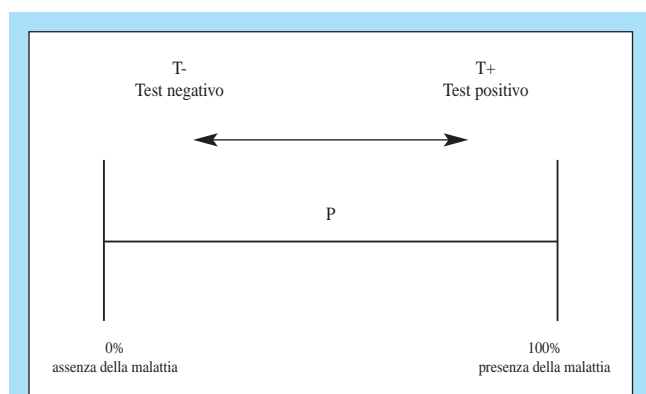


FIGURA 1 - Cambiamento della probabilità di malattia in seguito ai risultati di un test diagnostico.

spostamento del livello di probabilità pre-test dipende dal potere informativo del test che è funzione della sua accuratezza. La probabilità di malattia che si ottiene dopo l'esecuzione del test è detta probabilità post-test ed è comunemente conosciuta come valore predittivo del "risultato" di un test diagnostico.

Consideriamo ora in dettaglio i diversi elementi del processo appena descritto.

LA PROBABILITÀ PRE-TEST DI MALATTIA

Il punto più critico dell'approccio probabilistico alla diagnosi è la stima corretta della probabilità pre-test¹. Quanto più precisa sarà questa stima tanto più efficace sarà l'impiego di un test e il raggiungimento di un elevato grado di certezza diagnostica. Si tratta di una probabilità soggettiva che il clinico stima a partire da diverse fonti:

- conoscenza della diffusione della malattia (prevalenza) nella popolazione di riferimento da cui proviene il paziente;
- dati derivanti dalla raccolta dell'anamnesi (predisposizione genetica, presenza/assenza di fattori di rischio quali età, razza, peso, attitudine, ecc);
- dati clinici evidenziati durante l'esame clinico-fisico.

L'esperienza clinica ha un ruolo fondamentale nel supportare il veterinario nella stima corretta della probabilità pre-test e nella maggioranza dei casi è sufficiente a formulare con un buon grado di certezza la diagnosi senza ricorrere all'impiego di test di laboratorio o strumentali. È fondamentale che il clinico sia in grado di produrre stime corrette di questa probabilità dato che il successo o l'insuccesso diagnostico dipende in larga parte da essa. Trattandosi di una valutazione "soggettiva" sarà affetta da un certo grado di errore, è quindi consigliabile attribuire un possibile *range* di valori della corretta probabilità pre-test in modo da valutare, con i sistemi che saranno illustrati in seguito, quanto questa incertezza sia rilevante nel determinare l'interpretazione dei risultati del test.

IL POTERE INFORMATIVO DI UN TEST DIAGNOSTICO

Come è noto l'accuratezza di un test diagnostico binario è misurata mediante la **Sensibilità** (Se) e la **Specificità** (Sp)^{4,5,6}. La Tabella 1 illustra la classica organizzazione dei dati in formato 2x2, da cui si ricavano i valori di Se e Sp. Va notato che questi valori non sono di nessuna utilità clinica immediata. Infatti i valori Se e Sp di un test non possono essere direttamente impiegati per rispondere alla domanda fondamentale che si pone il clinico: "l'animale positivo al test è veramente ammalato?" oppure, "l'animale negativo al test è davvero sano?". Per rispondere a queste domande bisogna introdurre il concetto di **potere informativo** del test diagnostico.

Il potere informativo o "weight of evidence"^{7,8}, è misurato in termini di rapporti di verosimiglianze o Likelihood Ratios (LR)^{6,1}. Per ogni test esistono due LR: un LR per risultati positivi (LR+) e un LR per risultati negativi (LR-)

che sono calcolati in modo semplice a partire dai dati di Se e Sp, secondo queste formule:

$$LR+ = \frac{\text{Sensibilità}}{1 - \text{Specificità}} = \frac{\text{Proporzione di Veri Positivi}}{\text{Proporzione di Falsi Positivi}}$$

$$LR- = \frac{1 - \text{Sensibilità}}{\text{Specificità}} = \frac{\text{Proporzione di Falsi Negativi}}{\text{Proporzione di Veri Negativi}}$$

Ad esempio, Mueller⁹ et al. (2001) hanno valutato l'accuratezza del riflesso pinna-pedale come strumento per la diagnosi di scabbia in 55 cani con tale patologia e in 533 animali con altre patologie cutanee, ottenendo i seguenti risultati: Se = 82% e la Specificità = 94%.

Da cui si ricava:

$$LR+ = \frac{0.82}{(1-0.94)} = 13.6 \quad LR- = \frac{(1-0.82)}{0.94} = 0.20$$

Vediamo il significato di questi valori. In caso il test fornisca un risultato positivo, la probabilità che l'animale sia davvero ammalato (vero positivo) è circa 14 volte più grande della probabilità che sia invece sano (falso positivo). Nel caso di test negativo, la probabilità che un animale sia ammalato (falso negativo) è 0.20 volte più piccola della probabilità che sia sano (vero negativo). Se un test non ha nessuna capacità discriminante (Se e Sp = 50%) sia LR+, sia LR- avranno un valore pari a 1. Al contrario un test perfetto (gold standard) avrà un LR+ uguale a ∞ (100% / 0%) e un LR- uguale a 0 (0%/100%).

Il rapporto di verosimiglianza, fornisce al clinico un valore direttamente spendibile nel processo decisionale diagnostico. Il clinico potrà utilizzare i valori di LR, ad esempio, per selezionare i test che forniscono il maggior valore di LR+ in caso di test positivo o il minor valore di LR- in caso di test negativo. Supponiamo di essere nella situazione di dover scegliere tra due test diagnostici che presentano i seguenti parametri di accuratezza: Test A Se = 82%, Sp = 94% Test B Se = 94%, Sp = 82%. Si noterà che i due test presentano valori di Se e Sp opposti. Quale è il più accurato? Non si può rispondere a questa domanda semplicemente considerando separatamente i valori di Se e Sp. Se confrontiamo i due test rispetto ai valori di LR+ e LR- avremo i risultati riportati in Tabella 2.

Il test A (ad elevata specificità e LR+ più grande) fornisce risultati positivi più credibili del test B (ad elevata sensibilità e LR- più piccolo), che invece a sua volta è più credibile quando fornisce risultati negativi. Da questi risultati possiamo dedurre una semplice regola per l'interpretazione dei risultati quando siamo nella condizione di impiego di test noti per l'elevata sensibilità o per l'elevata specificità. La regola si traduce in due acronimi facili da memorizzare¹:

1. **SnNout** è un acronimo inglese che deriva da "When a test has a very high Sensitivity, a Negative result rules out the diagnosis", ovvero: quando un esame ha una sensibilità molto elevata (>95%) un risultato negativo virtualmente esclude la diagnosi;
2. **SpPin** deriva invece da "When a test has a very high Specificity, a Positive result rules in the diagnosis" ovvero: quando un esame ha una specificità molto elevata un risultato positivo virtualmente conferma la diagnosi.

I rapporti di verosimiglianza di un test sono proprietà insite nel test e teoricamente stabili (cioè costanti) e non dipendono dalla situazione clinica contingente a cui il test è applicato.

Il potere discriminante o, nella concezione bayesiana, il potere informativo di un test espresso in termini di LR non è però ancora sufficiente a indicare l'utilità clinica. Se escludiamo i casi in cui possiamo applicare la regola SpPin e SnNOut, nella maggioranza dei casi è necessario acquisire un'altra informazione per poter esprimere un giudizio sull'utilità clinica di un test: **la probabilità post-test.**

CALCOLO DELLA PROBABILITÀ POST-TEST

La probabilità post-test è una misura "oggettiva" che indica qual è la probabilità di malattia in base al risultato del test ed è comunemente indicata come Valore Predittivo Positivo (quando si riferisce a un risultato positivo del test) e Valore Predittivo Negativo (quando si riferisce a risultati negativi).^{10,6,7} La probabilità post-test si ottiene integrando le informazioni considerate precedentemente: probabilità pre-test e rapporto di verosimiglianza (Likelihood Ratio). Per poter relazionare queste due grandezze è necessario utilizzare le stesse dimensioni. Abbiamo visto che il rapporto di verosimiglianza è un rapporto tra due probabilità (probabilità di essere un vero positivo/1-probabilità di essere un vero negativo). In termini statistici questo rapporto è noto con il termine inglese di Odds¹¹. Analogamente possiamo riferirci alla probabilità pre-test in termini di Odds rapportando la probabilità pre-test di malattia alla probabilità pre-test di non-malattia. Ad esempio, se ad un cane viene attribuita la probabilità

Tabella 1
Tabella 2x2 per la stima dei parametri di accuratezza di un test diagnostico (Sensibilità e Specificità)

Risultato del test	Malati	Non malati	
T+	Veri positivi a	Falsi Positivi b	Positivi al test a+b
T-	Falsi Negativi c	Veri Negativi d	Negativi al test c+d
	Malati a+c	Sani b+d	N a+b+c+d

Se = Veri positivi/Malati = a/(a+c): indicati come "i positivi tra i malati"
Sp = Veri negativi/Sani = d/(b+d): indicati come "i negativi tra i sani"

Tabella 2
Confronto tra due test sulla base dei loro LR's

TEST	LR+	LR-
A (Se=82%, Sp=94%)	13.6	0.20
B (Se=94%, Sp=82%)	5.2	0.07

pre-test di avere una infestazione da filarie, pari al 30%, il suo Odds di malattia sarà pari a $30\% / (100-30\%) = 0.43$. Questo risultato indica che il cane ha una probabilità di essere infestato 0.43 volte più bassa della probabilità di non esserlo.

Il calcolo della probabilità post-test passa quindi per i seguenti passaggi (Sackett¹ e Smith⁶ Gambino⁷):

1. calcolo dell'odds pre-test: $p/(1-p)$
2. calcolo dell'odds post-test: $LR \times \text{odds pre-test}$
3. conversione dell'odds post-test in probabilità post-test: $\text{odds post-test} / (1+\text{odds post-test})$

Facciamo un esempio:

Supponiamo che il cane di cui sopra, venga sottoposto ad esame diagnostico con un ipotetico test dotato di un $LR+ = 15$: dopo un test positivo la sua probabilità post-test sarà:

1. odds pre-test = $30\%/70\% = 0.43$
2. odds post-test = $15 \times 0.43 = 6.45$
3. probabilità post-test = $6.45 / (1+6.45) = 0.86$ (86%).

In caso di test positivo la probabilità che il cane sia infestato da filarie è dell'86%.

Consideriamo ora l'eventualità che il test dia un risultato negativo, ci si chiede qual è la probabilità che il cane sia comunque positivo (e che quindi il test abbia fallito) e per contro qual è la probabilità che il test sia corretto e quindi l'animale sia davvero negativo.

I passi sono identici a quelli già visti, con la differenza che in questo caso utilizzeremo il valore del rapporto di verosimiglianza del test riferito ad un risultato negativo: $LR-$, che supponiamo sia 0.10.

Riprendiamo la procedura di calcolo:

1. odds pre-test = $30\%/70\% = 0.43$
2. odds post-test = $0.1 \times 0.43 = 0.043$
3. probabilità post-test = $0.043 / (1+0.043) = 0.04$ (4%).

Quindi in caso di esito negativo del test la probabilità che sia un falso negativo e che quindi l'animale sia comunque positivo è solo del 4% e la probabilità che l'animale sia davvero sano è pari a $100\% - 4\% = 96\%$.

Fortunatamente, nella pratica si può ricorrere ad un sistema semplificato per il calcolo della probabilità post-test, grazie all'ausilio di uno strumento noto come: nomogramma (o nomografo) di Fagan¹², riportato in Figura 2. Il suo utilizzo è molto semplice: si ricerca nella linea dei valori di probabilità pre-test quella che è attribuita al caso specifico sotto indagine (ad esempio 30% o 0.30), poi si cerca nella linea centrale il valore del $LR+$ o $LR-$ a seconda che il test risulti positivo o negativo e con un righello si traccia una linea retta tra i due punti fino ad individuare il valore di probabilità post-test corrispondente. Nel caso di utilizzo del $LR-$ ci si deve ricordare di sottrarre a 100 il risultato per avere la probabilità che l'animale sia davvero negativo e di confrontarlo con la probabilità di non malattia pre-test che si ottiene. In Figura 2 è riportato un esempio dell'utilizzo del nomogramma di Fagan utilizzando i dati dell'ipotetico test per la diagnosi di filariosi.

UTILITÀ CLINICA DEI TEST

Ora abbiamo tutti gli elementi per valutare l'utilità clinica del test: Probabilità pre-test, Likelihood Ratio e Probabilità post-test.

In accordo a Sackett¹ e Knotteraus¹³ definiamo utilità clinica di un test la quantità d'informazione che il test aggiunge alla probabilità soggettiva del clinico. La misura dell'utilità clinica di un test è quindi la differenza (assoluta) tra la probabilità post-test e la probabilità pre-test.

$$\text{Utilità Clinica (UC)} = \text{Probabilità Post-test} - \text{Probabilità Pre-test}$$

Consideriamo l'esempio del cane testato per filaria con un test con $LR+ = 15$ e $LR- = 0.10$. Da quanto detto sopra abbiamo che la probabilità pre-test è pari al 30% mentre quella post-test in caso di test positivo è uguale all'86%. Da questi dati ricaviamo che:

$$UC = 86 - 30 = 56.$$

In uno scenario clinico come quello da noi descritto il test dà un importante contributo informativo al clinico nel definire la probabilità che il cane abbia la filaria, fornendo un incremento di 56 punti percentuali alla stima soggettiva indicata prima di effettuare il test. In caso di un risultato negativo il test avrà una $UC = 70 - 96 = 26$ punti percentuali. È interessante notare come dai risultati dell'UC il test ha maggior utilità clinica quando fornisce un test positivo. Utilizzando il nomogramma di Fagan, il clinico potrà quindi prima di effettuare il test verificare per ogni singolo caso clinico, quale test è dotato di maggior utilità clinica e qual è il peso probabilistico associato al risultato. Facciamo notare che questa decisione deve essere presa in base

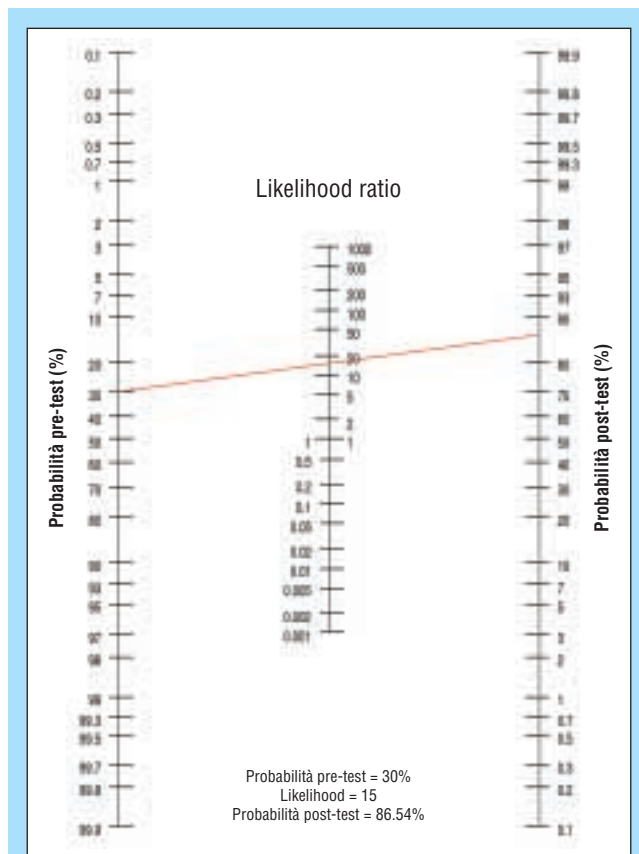


FIGURA 2 - Nomogramma di Fagan. Data una probabilità di malattia pre-test del 30% un rapporto di verosimiglianza per un test positivo di 15, in caso di positività al test si otterrà una probabilità post-test dell'86.54%.

al livello di probabilità pre-test del caso clinico e non in base alle caratteristiche del test, che verranno prese in esame solo nella fase successiva: un test, indipendentemente dalle sue caratteristiche di accuratezza, può rilevarsi utile o inutile a seconda della probabilità pre-test del caso clinico a cui sarà applicato.

Chiariamo questo concetto con un ipotetico, ma realistico, caso clinico: considereremo l'applicazione del test di soppressione a basse dosi del desametasone (LDD) per la diagnosi di ipercorticosurrenalismo applicato a 3 differenti scenari clinici:

- 1) Barboncino M 10 anni. Polidipsia, poliuria, polifagia. Aumento di peso. Addome rigonfio. Alopecia regione dei fianchi bilaterale simmetrica con pelle sottile e moderatamente pigmentata. Emocitometria nella norma. Ematochimica: AST 110 UI ALT 85 UI, Fosfatasi alcalina 1410 UI, P.S. urine 1010. Sulla scorta delle evidenze cliniche e di laboratorio raccolte il clinico ipotizza che il paziente abbia un'elevata probabilità di essere affetto da ipercorticosurrenalismo. Decide quindi di sottoporre il soggetto al test diagnostico di soppressione a basse dosi del desametasone (LDD). In letteratura sono riportati diversi valori relativi all'accuratezza di questo test, con Sensibilità e Specificità molto variabili¹⁴. Consideriamo, ai fini dell'esempio, che si utilizzi un ipotetico test dotato di una sensibilità e una specificità rispettivamente del 94% e 80%, da cui ricaviamo che il rapporto di verosimiglianza per un test positivo è pari a $LR+ = 4.7$ e per un test negativo $LR- = 0.075$. Quale è l'utilità diagnostica di questo test nel modificare significativamente la probabilità di malattia che il clinico ha attribuito al paziente prima dell'esecuzione dell'esame? Sappiamo che per rispondere a questa domanda è necessario che il clinico quantifichi la sua "ipotesi" diagnostica. Un quadro come quello descritto è compatibile con una elevata probabilità di ipercorticosurrenalismo che quantificheremo pari al 90%. In caso di un esame positivo, utilizzando il nomogramma di Fagan, la probabilità di malattia risulterà pari al 98%, modificando la probabilità iniziale di 8 punti percentuali. Consideriamo ora l'eventualità che l'esame risulti negativo, avremo che la probabilità di malattia sarà scesa al 60%, modificando considerevolmente la probabilità iniziale di 30 punti percentuali. A causa dell'elevata sensibilità il test di soppressione fornisce infatti informazioni più utili in caso di esiti negativi.
- 2) Cane Pastore tedesco M 7 anni. Dimagrimento da 1 mese. Ora addome rigonfio. Sintomi: disoressia, polidipsia, vomito saltuario, pallore delle mucose. Emocitometria nella norma. Ematochimica: AST 145 UI ALT 100 UI Fosfatasi alcalina 510 UI P.S. urine 1013. In un questo paziente il clinico è più propenso a considerare tra le diagnosi più probabili, malattie quali l'emangiosarcoma splenico e/o epatico o l'insufficienza cardiaca, relegando in fondo alla lista delle possibili diagnosi quella di ipercorticosurrenalismo. Supponiamo quindi che per questo soggetto la probabilità di ipercorticosurrenalismo sia solo del 5% e che per scrupolo il clinico decida di applicare il test di soppressione a ragione della sua elevata accuratezza. In caso di esito positivo al test la probabilità di ipercorticosurrenalismo risulterà pari al 20%, modificando quindi la probabilità

iniziale di 15 punti percentuali, al contrario in caso di esito negativo la probabilità di malattia sarà dello 0.40%, modificando la probabilità iniziale di 4 punti percentuali.

- 3) Fox terrier F 13 anni moderatamente sovrappeso. Ciclo estrale irregolare. Appetito conservato. Polidipsia, poliuria. Addome rigonfio. Cataratta bilaterale. Seborrea. Emocitometria: modesta leucocitosi, G.B. 21.000 mm^3 . Ematochimica: AST 100 UI, ALT 79 UI, Fosfatasi alcalina 710 UI, P.S. urine 1008. In questo soggetto la formulazione delle ipotesi diagnostiche risulta complessa essendo il quadro clinico compatibile sia con l'ipercorticosurrenalismo, sia con piometra o con il diabete mellito. Il clinico può quantificare la probabilità di ipercorticosurrenalismo pari al 50%. In questo caso ci troviamo di fronte alla situazione di maggiore incertezza diagnostica. Il clinico necessita quindi di incrementare le informazioni in suo possesso per poter esprimere un giudizio diagnostico. Supponiamo che anche in questo caso esplori la possibilità che il paziente sia affetto da ipercorticosurrenalismo praticando il test di soppressione del desametasone. In caso di un risultato positivo avremo che la probabilità di malattia post-test del paziente sarà pari all'83%, incrementando quindi di ben 33 punti percentuali la probabilità iniziale. In caso di esito negativo il paziente avrà una probabilità di ipercorticosurrenalismo pari al 7%, riducendo la probabilità iniziale di 43 punti percentuali.

Il "peso" del risultato di un test diagnostico (positivo/negativo), o la sua utilità, nel modificare in maniera rilevante il giudizio clinico-diagnostico misurabile attraverso la stima della probabilità post-test, è differente per ogni caso essendo correlato al valore "soggettivo" che il clinico esprime prima dell'esecuzione del test. Nei casi sopra esaminati possiamo dire che nell'eventualità che il paziente abbia una elevata probabilità di malattia (90%) (Fig. 3A), l'esecuzione del test di soppressione si mostra particolarmente utile nel caso fornisca risultati negativi in quanto modifica sostanzialmente la probabilità di malattia. Al contrario un risultato positivo modifica di poco questa probabilità. In questo caso quindi il clinico può verificare a priori il comportamento del test diagnostico e decidere se in caso di esito negativo la probabilità di malattia post-test del 60% sia sufficientemente "bassa" da escludere la malattia dalla lista delle possibili diagnosi oppure ancora sufficientemente alta da confermare la diagnosi e procedere quindi con un test differenziale come LDD per l'accertamento delle cause di iperadrenocorticismo. Oppure, considerando che l'eventualità di un test negativo non esclude la diagnosi di iperadrenocorticismo, può decidere che il test di stimolazione positivo non incrementi significativamente il suo giudizio diagnostico pre-test già molto alto e decidere di passare direttamente all'esecuzione di un test differenziale. Nel secondo caso (Fig. 3B), quando le manifestazioni cliniche siano compatibili con altre malattie relegando l'ipotesi di ipercorticosurrenalismo ad una bassa probabilità, si osserva che il test fornisce utili informazioni quando dà un risultato positivo, anche se il "peso" probabilistico di questa ipotesi è ancora relativamente basso (20%) rispetto alle altre patologie. Al contrario, in caso di un risultato negativo il test modifica poco il giudizio clinico. Facciamo notare che affermare che il paziente abbia

solo una probabilità del 5% di avere il morbo di Cushing corrisponde ad affermare che ha 95 probabilità su 100 di non avere la malattia. In questo caso un test negativo corrisponde ad una probabilità post-test che l'animale non abbia il Cushing pari al 99.6%. Nonostante l'elevata percentuale, è evidente che il clinico era già giunto ad una conclusione analoga, prima di eseguire il test. In questo senso eseguire il test per escludere la malattia è di fatto irrilevante. Al contrario il clinico deve essere cosciente che se esegue su quel tipo di paziente il test di Cushing, dovrà considerare con una certa attenzione gli eventuali risultati positivi. Sarà il clinico che dovrà giudicare se una probabilità di malattia del 20% è sufficiente a giustificare un ulteriore approfondimento diagnostico o, al contrario, considerare il risultato come un falso positivo. Questi "calcoli" eseguiti prima di qualsiasi test diagnostico permettono al clinico di decidere in modo razionale se effettuare o meno il test. Nel terzo caso presentato (Fig. 3C), le evidenze cliniche e di laboratorio lasciano il veterinario in uno stato d'incertezza. In questo caso dai risultati sopra riportati possiamo osservare che sia risultati positivi, sia negativi, modificano in modo rilevante la probabilità a priori espressa dal clinico. Questo comportamento è costantemente presente per ogni test diagnostico qualunque sia la sua accuratezza. Possiamo affermare quindi che in caso di probabilità pre-test del 50%, i test forniscono la loro massima utilità clinica¹. Quanto fin qui trattato permette di evidenziare come l'interpretazione corretta del risultato di un test diagnostico sia strettamente correlata alla probabilità "soggettiva" pre-test di malattia. Da questo punto di vista diventa rilevante per il clinico produrre una stima di questa probabilità che sia la più attendibile possibile. A causa della natura soggettiva di questa stima non è possibile fornire nessun sistema di "calcolo" valido. Maggiori sono le informazioni che si hanno a disposizione sulla malattia (prevalenza nella popolazione, prevalenza in sottogruppi specifici di popolazione quali razza, età e sesso, fattori di rischio associati alla malattia) maggiore sarà l'accuratezza con cui il clinico stimerà la probabilità pre-test. A questo insieme d'informazioni che possono essere reperite in letteratura, va aggiunta la componente dell'esperienza. Clinici con maggior "esperienza" riconosceranno con maggior precisione profili clinici "caratteristici" per le diverse

patologie. Restano comunque le difficoltà dovute all'eventuale "rarietà" della patologia che si sta considerando e il fatto che si tende a memorizzare con maggior precisione i casi più recenti e in genere i casi più "eclatanti" risolti con successo o i casi non risolti, con esito infausto del paziente. Mantenere un dettagliato archivio dei casi clinici può rivelarsi utile nel fornire informazioni spendibili nella formulazione della probabilità pre-test. Per ridurre gli errori nella stima della probabilità pre-test è consigliabile individuare un "intervallo" di valori di probabilità pre-test in cui far ricadere il giudizio clinico relativo al paziente in esame e, di conseguenza, effettuare i calcoli della probabilità post-test utilizzando i limiti inferiori e superiori degli intervalli di probabilità considerati. È consigliabile quindi definire inizialmente in termini soggettivi la probabilità pre-test come ad esempio: molto bassa, bassa, media, medio-alta, alta, e quindi stabilire gli intervalli numerici associati. Ad esempio nel caso clinico 2, si potrebbe considerare che il cane abbia una bassa probabilità di essere affetto dal morbo di Cushing, stimabile fino ad un massimo del 10%. Di conseguenza il clinico potrà verificare con l'ausilio del nomogramma di Fagan quanto la sua "incertezza" si rifletta sulla probabilità post-test in caso di test positivo o negativo. Ad esempio nel caso la probabilità pre-test sia pari al 10%, potrà verificare che in caso di un test positivo avrà una probabilità post-test pari al 35%, mentre in caso di test negativo la probabilità di malattia sarà solo dello 0.83%. Osserviamo che in questo caso l'incertezza iniziale tende a riflettersi principalmente in caso di un risultato positivo, in quanto la probabilità post-test inizialmente stimata (20%) è salita al 35%. Spetterà al clinico stabilire se l'incertezza intorno alla stima della probabilità pre-test sia sufficientemente stretta da non modificare sostanzialmente il suo giudizio clinico alla luce di risultati positivi o negativi del test. In caso contrario tenderà ad utilizzare stime di probabilità pre-test più conservative che a suo giudizio permettano di esplorare le diverse ipotesi diagnostiche.

CONCLUSIONI

L'utilità clinica di un test diagnostico è una funzione di tre grandezze: la probabilità a priori di malattia, il peso dell'evidenza oggettiva fornita dal test diagnostico (Likelihood Ratio) e infine dalla probabilità post test di malattia che è facilmente ricavabile attraverso l'impiego del nomogramma di Fagan. L'utilità clinica è quindi specifica per ogni singolo scenario clinico e non deve essere confusa con l'accuratezza del test diagnostico. In questo lavoro si è cercato di sintetizzare quali sono i presupposti teorici e gli strumenti che permettono di stimare quale sarà il reale vantaggio in termini di miglioramento della certezza diagnostica nell'utilizzo di un test, prima della sua esecuzione. Un pratico nomogramma di Fagan è reso disponibile presso il sito del Children's Mercy Hospital di Kansas City (USA) (<http://www.childrens-mercy.org/stats/sliderule.asp>).

Per poter utilizzare al meglio le informazioni fornite da diversi strumenti diagnostici in uso nella pratica clinica veterinaria è importante che si disponga di accurate e precise stime dei parametri di Sensibilità e Specificità da cui poter ricavare secondo le formule sopra indicate i valori di LR+ e LR- che permettono l'uso del nomogram-

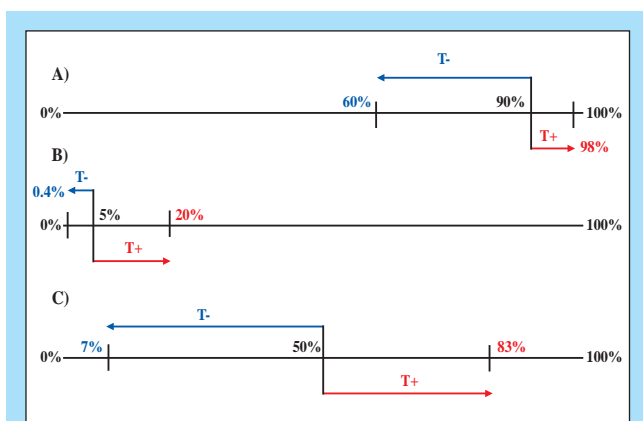


FIGURA 3 - Modifiche della probabilità pre-test per differenti scenari clinici di sospetto diagnostico del morbo di Cushing: **A)** caso clinico 1 (probabilità pre-test = 90%); **B)** caso clinico 2 (probabilità pre-test = 5%); **C)** caso clinico 3 (probabilità pre-test = 50%).

ma di Fagan. È compito quindi della ricerca clinica veterinaria produrre e riportare in letteratura le stime di questi parametri, secondo rigorose metodologie ben descritte dal gruppo STARD (The Standards for Reporting of Diagnostic Accuracy "STARD" steering group)¹⁵ e da Irwing et al. (2002)¹⁶, Sackett et al. (2002)¹⁷ e Knottnerus J.A. et al. (2002)¹³.

L'impiego di test diagnostici di cui non si conosca l'accuratezza o di cui si hanno fondati motivi di dubitare della validità di essa a causa di errori metodologici in corso di valutazione, dovrebbe essere evitato in quanto porterebbe ad una valutazione dei risultati del test che ignora l'importante componente "probabilistica" che caratterizza il processo diagnostico. Quanto detto in questo lavoro dovrebbe chiarire il significato del titolo. Infatti, in assenza di una modifica sostanziale della probabilità di malattia dopo l'esecuzione del test (differenza tra probabilità pre-test e probabilità post-test), il clinico può, sempre rispetto allo specifico caso clinico che ha di fronte, ritenere un test diagnostico, inutile e quindi evitare di eseguirlo. Riteniamo sia importante quindi che la scelta del test da utilizzare durante il work up diagnostico non si basi esclusivamente sulle performance del test, ma che si proceda sempre ad una valutazione a priori della sua reale utilità clinica, al fine di ottimizzare il processo diagnostico.

Parole chiave

Test diagnostico, sensibilità, specificità, rapporti di verosimiglianza, accuratezza, diagnosi.

Key words

Diagnostic test, sensibility, specificity, likelihood ratio, accuracy, diagnosis.

Bibliografia

1. Sackett D.L., Haynes R.B., Guyatt G.H., Tugwell P. (1991). Clinical epidemiology: a basic science for clinical medicine. Boston:Little, Brown. Second Edition.
2. Fleiss J.L. (1981). Statistical methods for rates and proportions. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc. USA. Second Edition.
3. Scandellari C. (1991). The Bayesian approach to evaluation of diagnostic data. Ann. Ist. Super. Sanità Vol. 27, N. 3 (1991), pp. 385-394.
4. Henken A.M., Graat E.A.M., Casal J. (1997). Measurement of disease frequency. In: Application of quantitative methods in veterinary epidemiology. Editors: Noordhuizen J.P.T.M., Frankena K., van der Hoofd C.M., Graat E.A.M. Wageningen Pers. Wageningen (The Netherlands).
5. Dawson B., Trapp R.G. (2001). Basic & Clinical Biostatistics. Lange Medical Books/McGraw-Hill. Medical Publishing Division. Third Edition. New York.
6. Smith R. Veterinary Clinical Epidemiology: A Problem-Oriented Approach, Second Edition. 1995. CRC Press.
7. Gambino R. (1991). The misuse of predictive value-or why you must consider the odds. Ann. Ist. Super. Sanità. Vol. 27, N. 3 pp. 395-400.
8. Good I.J. (1960) Weight of evidence, corroboration, explanatory power, information and the utility of experiments. J. R. Statis. Soc. 22(B) 319-331.
9. Mueller R.S., Bettenay S.V., Shipstone M. (2001). Value of the pinnal-pedal reflex in the diagnosis of canine scabies. Vet Rec 2001 May 19; 148(20):621-3.
10. Martin S.W., Meek A.H., Willenberg P. Veterinary Epidemiology: Principles and Methods. Iowa State University Press, 1987.
11. Grassi M. (1994). Statistica in medicina: un approccio basato sulla verosimiglianza. McGraw-Hill.
12. Fagan T.J. (1975). Nomogram for Bayes's theorem. New Engl. J. Med. July. 31; 293(5):257.
13. Knottnerus J.A., Weel C., Muris J.W.M. Evidence base of clinical diagnosis: Evaluation of diagnostic procedures BMJ 2002; 324:477-480.
14. Zerbe C.A. (2001). Test di screening per la diagnosi dell'iperadrenocorticismo del gatto e nel cane. Supplemento a Veterinaria, Anno 15, n. 3, Ottobre 2001.
15. Patrick M. Bossuyt, Johannes B. Reitsma, David E. Bruns, Constantine A. Gatsonis, Paul P. Glasziou, Les M. Irwing, Jeroen G. Lijmer, David Moher, Drummond Rennie, Henrica C.W. de Vet. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ VOLUME 326 4 JANUARY 2003 bmj.com.
16. Irwing L., Bossuyt P., Glasziou P., Gatsonis C., Lijmer J. (2002) Evidence base of clinical diagnosis: Designing studies to ensure that estimates of test accuracy are transferable BMJ 2002; 324:669-671.
17. Sackett D.L. and Haynes R.B. Evidence base of clinical diagnosis: The architecture of diagnostic research BMJ 2002; 324:539-541.