

# LA VALUTAZIONE CRITICA DEI TEST DIAGNOSTICI

STEFANO GUAZZETTI  
[stefano.guazzetti@ausl.re.it](mailto:stefano.guazzetti@ausl.re.it)

La diagnosi è l'atto che meglio riassume ed identifica le capacità del clinico. Nonostante ciò la diagnosi è connessa in modo indissolubile all'errore diagnostico e seppure vi sia in genere un ampio accordo sul fatto che la massima abilità del clinico consista nella capacità di sbagliare meno frequentemente, l'errore diagnostico è percepito e vissuto dal veterinario in modo meno che sereno: questo è infatti considerato il più grave degli errori, in quanto dalla diagnosi discendono tutte le azioni del clinico veterinario, sia in senso terapeutico che preventivo. Si deve anzitutto riconoscere che, al di là di casi banali, non è quasi mai possibile conoscere l'esatta condizione di salute o malattia di un animale in vita, sia per la variabilità delle caratteristiche clinicamente misurabili che connotano per l'appunto lo stato fisiologico o patologico, che per l'imperfezione dei mezzi diagnostici utilizzabili nella pratica clinica.

A ben vedere coesistono nel clinico due attitudini opposte: un rigetto formale dell'errore, in quanto componente indesiderata della diagnosi, ed una attitudine a valutare criticamente ed in modo integrato tutti quegli elementi che compongono il quadro diagnostico, dalla anamnesi all'esame più specialistico. Questa ultima attitudine, già presente nei clinici più giovani è affinata con l'esperienza ed - è possibile dimostrarlo - è riconducibile alla comprensione, seppure intuitiva, dei fondamenti della logica induttiva formulati da Thomas Bayes (1702-1761) nel suo teorema.

Senza volere qui entrare nei dettagli matematici, pur semplici, della teoria di Bayes e volendo mantenere anche in questi atti un approccio discorsivo, si cercherà di fornire al clinico una formalizzazione di questa sua naturale attitudine critica, che risulterà sicuramente utile non fosse altro per tentare di quantificare l'incertezza associata alle conclusioni diagnostiche.

## Bayes e il ragionamento diagnostico

Per meglio comprendere l'origine della incertezza associata ad un test diagnostico si deve considerare in primo luogo la variabilità delle caratteristiche misurabili delle popolazioni, in particolare di quelle di rilevanza clinica. Questa variabilità comprende anche quella componente eventualmente introdotta dall'utilizzo di strumenti di misura. Si osservi il grafico *a* della Fig.1: qui è rappresentata la distribuzione di una misura di importanza clinica in una ipotetica popolazione di malati e non malati. Le distribuzioni sono completamente separate ed è possibile identificare un valore che con certezza discrimina i soggetti malati dai sani. Questa situazione ideale - il *gold-standard* - non si realizza quasi mai nella pratica, dove invece si ha piuttosto la situazione schematizzata nel grafico *b*: la sovrapposizione delle distribuzioni dei valori della variabile discriminante impedisce di distinguere con certezza e sulla sola base della misurazione della variabile di interesse, i soggetti malati da quelli sani ma viene introdotta dalla applicazione del test stesso la possibilità di commettere un errore diagnostico, identificando come malato un sano o viceversa.

Il quesito fondamentale del clinico è a questo punto: dato un certo esito del test diagnostico, quale è la probabilità che il soggetto sia malato (o sano?). È quindi nella metrica probabilistica che il quesito deve trovare una soluzione e, si vedrà, questa è data dalla regola di Bayes, che mostra come pesare le evidenze fornite dal test diagnostico in modo da ottenere, a partire dalla probabilità pre-test di malattia, la probabilità post-test.

Per ragioni operative definiamo qui la diagnosi come processo, induttivo, dall'esito incerto, che si realizza iterando il seguente schema:

**informazione pre-test → test → informazione post-test.**

In questo senso **un test diagnostico è qualsiasi informazione che modifichi la nostra convinzione (probabilità) sullo stato di malattia di un animale.**

La conoscenza (probabilità) a priori che un soggetto sia malato interagisce quindi con la probabilità probativa (caratteristica del test) per dar luogo alla probabilità a posteriori (post test).

Il teorema di Bayes <sup>1</sup>: fornisce una risposta al problema: **dato che si è verificato A (test positivo o negativo) qual è la probabilità che la sua causa sia B (la malattia)?** In altre parole qual'è la probabilità di malattia dato un test positivo  $P(M+|T+)$ ?

### La Probabilità

Intuitivamente tutti possediamo il concetto di probabilità: per il nostro uso e senza voler indulgere in eccessivi formalismi, la probabilità di un evento può essere definita come il rapporto tra il numero dei casi favorevoli ad un evento di nostro interesse ed il numero dei casi possibili.

$P(x) = \frac{\text{casi favorevoli}}{\text{casi possibili}}$ , ad esempio, la probabilità che esca 4 (o qualsiasi altro numero)

nel lancio di un dado è  $1/6$  e cioè circa 0.167 (il 16.7%).

Alcune delle proprietà delle probabilità risulteranno utili per comprendere quello che segue e vengono qui illustrate:

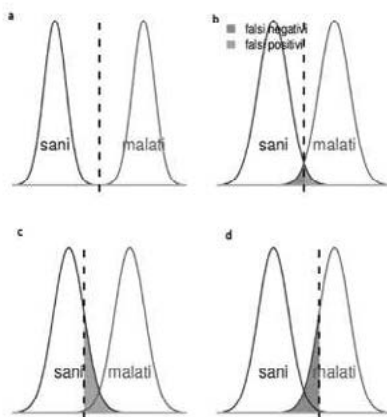
La probabilità è sempre compresa fra 0 ed 1. Se  $P(x)$  è la probabilità che un evento  $x$  si verifichi,  $1-P(x)$  è la probabilità che l'evento  $x$  non si verifichi.

È qualche volta conveniente esprimere la probabilità in termini di *odds*, cioè il rapporto fra i casi favorevoli e i casi sfavorevoli. Ad esempio nel caso del lancio di un dado l'odds di ottenere 4 è

$1/5$ , cioè 0.2. Le relazioni fra probabilità ed odds sono  $\text{odds}(x) = \frac{P(x)}{1-P(x)}$ ;  $P(x) = \frac{\text{odds}(x)}{1+\text{odds}(x)}$

Inoltre, se due eventi sono indipendenti, la probabilità di osservare entrambi è il prodotto delle probabilità di osservare i singoli eventi.  $P(A \cap B) = P(A) \times P(B)$

Fig. 1.



1 Il teorema di Bayes afferma che, date  $H_1, H_2, \dots, H_1, \dots, H_k$  cause tra loro esaustive e incompatibili e un evento  $E$ , la probabilità che avendo osservato l'evento, sia  $H_i$  la causa che lo ha generato è data da:

$$P(H_i | E) = \frac{P(E | H_i) P(H_i)}{\sum_{i=1}^k P(E | H_i) P(H_i)}$$

Il concetto di probabilità permea ogni aspetto del processo diagnostico ed epidemiologico, laboratoristi e clinici hanno la loro particolare visione degli aspetti probabilistici legati alla diagnosi.

**La probabilità pre-test** (a priori) è la probabilità di un soggetto di essere malato prima di essere assoggettato ad un test. La più semplice interpretazione della probabilità pre-test è quella degli epidemiologi, per i quali questa è semplicemente la probabilità che un soggetto estratto casualmente da una popolazione sia malato o abbia una condizione

di interesse  $P(M_{pre}^+) = \frac{\text{n di malati}}{\text{popolazione}}$  e si identifica quindi con la *prevalenza* della

condizione nella popolazione.

Il clinico tende ad attribuire a questa quantità una interpretazione individuale, corrispondente al grado di fiducia che egli ripone al riguardo del particolare stato dell'animale (sano o malato), non esistendo - se non in astratto - una situazione di assoluta ignoranza sullo stato clinico dell'animale, non fosse altro che in genere egli viene chiamato all'osservazione proprio perché un problema clinico sussiste. Anche quando questo non sia, come ad esempio nella pratica della prevenzione, la probabilità pre-test è sempre valutata relativamente all'età, al sesso, alle caratteristiche di allevamento, e cioè a tutti quei fattori che possono modificare la probabilità di malattia in una popolazione. Questi fattori concorrono alla ponderazione che il clinico effettua sull'esito del test diagnostico applicato.

### **La probabilità probativa: sensibilità e specificità dei test e il rapporto di verosimiglianza.**

Osserviamo di nuovo la figura 1: i pannelli c e d mostrano come variando la posizione del valore soglia che distingue i malati dai non malati, possiamo aumentare o diminuire la proporzione dei falsi negativi e dei falsi positivi. Sensibilità e specificità sono caratteristiche proprie dei test diagnostici e, almeno idealmente, non dipendono dalle caratteristiche del soggetto sul quale il test viene applicato.

La *sensibilità* può essere definita come misura della capacità di un test di identificare correttamente i malati ed è quindi la proporzione di malati che risulta positiva al test. In termini probabilistici la sensibilità rappresenta la probabilità di avere un test positivo, data

la malattia:  $Se = P(T^+ | M^+)$ ,  $Se = \frac{\text{veri positivi}}{\text{veri positivi} + \text{falsi negativi}}$  e il suo

complemento,  $1-Se$ , è la proporzione (probabilità) di falsi negativi.

Il pannello c della Fig. 1 mostra, per un test imperfetto, una situazione di massima sensibilità. Si noti come questa massima sensibilità comporti inevitabilmente una certa quota di falsi positivi.

In modo analogo la *specificità* può essere definita come la misura della capacità di un test di identificare correttamente i non malati ed è cioè la proporzione di non malati che risulta negativa al test. In termini probabilistici la Specificità è la probabilità di avere un test negativo, data la

condizione di non malattia:  $Sp = P(T^- | M^-)$ ,  $Sp = \frac{\text{veri negativi}}{\text{veri negativi} + \text{falsi positivi}}$

e il suo complemento,  $1-Sp$ , è la proporzione di falsi positivi.

Una situazione di massima specificità di un test imperfetto è mostrata nel pannello d della

figura 1, e ben si vede come ciò comporti una certa quota di falsi negativi.

**In ragione di quanto detto può essere utile ricordare, nella pratica clinica, che per escludere la malattia serve un test di grande sensibilità; in quanto un risultato negativo sarà poco probabile in un soggetto malato. Se viceversa l'interesse è rivolto alla conferma di una condizione patologica servirà un: test di grande specificità, di modo che un risultato positivo sia poco verosimile in assenza di malattia.**

Una quantità utile al clinico per poter valutare correttamente l'esito di un certo test diagnostico e che riassume le caratteristiche del test diagnostico è il *rapporto di verosimiglianza* (LR, likelihood ratio): questo ci dice quante volte è più probabile un risultato, positivo o negativo, in un soggetto malato rispetto ad un non malato.

Il LR è una funzione di sensibilità e specificità, è espresso in termini di odds e può assumere valori compresi fra 0 ed infinito. In particolare il *rapporto di verosimiglianza per un risultato positivo* LR<sup>+</sup> dice quante volte è più probabile un risultato positivo (T<sup>+</sup>) in un soggetto malato rispetto ad un non malato ed è calcolato come la probabilità di avere un test positivo essendo malato diviso la probabilità di avere un test positivo essendo non malato:

$$LR^- = \frac{P(T^- | M^+)}{P(T^- | M^-)} = \frac{\text{falsi negativi}}{\text{veri negativi}} = \frac{1 - \text{Sensibilità}}{\text{Specificità}}$$

Idealmente un test perfetto ha LR<sup>+</sup> infinita.

Allo stesso modo, il *rapporto di verosimiglianza per un risultato negativo* LR<sup>-</sup> dice quante volte è più probabile un risultato negativo (T) in un soggetto malato rispetto ad un non malato ed è la probabilità di avere un test negativo essendo malato diviso la probabilità di avere un test negativo essendo non malato:

$$LR^+ = \frac{P(T^+ | M^+)}{P(T^+ | M^-)} = \frac{\text{veri positivi}}{\text{falsi positivi}} = \frac{\text{Sensibilità}}{1 - \text{Specificità}}$$

Un test perfetto ha idealmente LR<sup>-</sup> uguale a 0.

Per comprendere cosa comporti l'applicazione di un test imperfetto in un popolazione si considerino i seguenti esempi: Si immagini di utilizzare un test con Se e Sp di 0.9 (90%), in una condizione prevalenza 0.01 (1%), n=1000, utilizzando lo schema di tabella 1.

Test	Malattia presente	Malattia assente
Positivo (T <sup>+</sup> )	VP	FP
Negativo (T <sup>-</sup> )	FN	VN
	M <sup>+</sup>	M <sup>-</sup>

**Tabella 1** VP=Veri Positivi, FP=Falsi Positivi, FN=Falsi Negativi, VN=Veri Negativi. La tabella 2 permette di computare le frequenze attese degli esiti del test diagnostico:

Test	Malattia presente	Malattia assente	totali
T <sup>+</sup>	9	99	108
T <sup>-</sup>	1	891	892
	10	990	1000

**Tabella 2**

La prevalenza apparente: sarà in questo caso  $108/100=0.108$  (10.8%), a cui corrisponde un valore predittivo di un risultato positivo (VPP) (rapporto fra veri positivi e tutti i positivi) di  $9/108=0.0833$  (8.3%) ed un valore predittivo di un risultato negativo (VPN) (rapporto fra veri negativi e tutti i negativi): di  $891/892=0.9988$  (99%).

Sempre a titolo esemplificativo si osservi quanto atteso nella stessa condizione di prevalenza (1%) ma applicando in una popolazione di 1000 soggetti un test con Se e Sp di 0.99 (99%) (Tab. 3):

Test	Malattia presente	Malattia assente	totali
T <sup>+</sup>	10	10	20
T <sup>-</sup>	0	980	980
	10	990	1000

**Tab. 3.**

La prevalenza apparente sarà qui circa  $20/1000=0.02$  (2%) ed il VPP  $10/20=0.5$  (50%), mentre il VPN sarà circa  $980/980=1$  (100%)

Per ultimo si consideri un test con Se e Sp di 0.9 (90%) ed una prevalenza del 10% (Tab. 4)

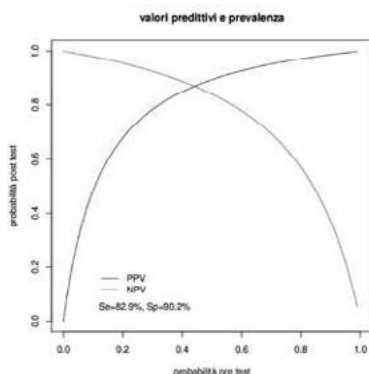
Test	Malattia presente	Malattia assente	totali
T <sup>+</sup>	90	90	180
T <sup>-</sup>	10	810	810
	100	900	1000

**Tab 4.**

Prevalenza apparente:  $180/1000=0.18$  (18%) VPP= $90/180=0.5$  (50%); VPN= $810/820=0.987$  (98.7%)

La Fig. 2 mostra, per un test con sensibilità 82.9% e specificità 90.2%, come i valori predittivi per un risultato negativo o positivo varino in funzione della probabilità pre-test

**Valori predittivi e prevalenza:** gli esempi qui sopra servono a far comprendere quale sia l'importanza della probabilità pre-test nel determinare il valore predittivo di un test. Sensibilità e specificità del test, che pur rivestono grande importanza ( in particolare per i laboratoristi), non sono da sole di grande utilità al clinico, che necessita di integrare la conoscenza di sensibilità e specificità con (la sua stima del) la probabilità a priori per ottenere il valore predittivo del test (cioè la probabilità post-test) che, come abbiamo visto, è la quantità di suo interesse, perché risponde al quesito fondamentale: dato questo risultato, qual'è la probabilità che il soggetto sia ammalato?



**Fig. 2:** rapporto fra valori predittivi e prevalenza.

### La probabilità post-test

Il T. di Bayes può essere riscritta nei termini delle quantità sopra descritte: sensibilità, specificità e probabilità pre-test o, oppure in termini di odds e rapporto di verosimiglianza. I calcoli, sulla scala di odds, sono molto semplici ed il T. di Bayes può essere scritto in forma semplificata:  $\text{odds post-test} = \text{odds pre test} \times \text{LR}$ , valendo sempre le relazioni

e  $P(x) = \frac{\text{odds}(x)}{1 + \text{odds}(x)}$  e potendo quindi riportare sempre il risultato in termini di probabilità.

Riprendiamo il terzo esempio (Tab. 4) per mostrare come i calcoli siano piuttosto banali: per un test con Se e Sp di 90% ed una prevalenza del 10%, ponendo un test con esito positivo avremo:

$$\text{odds( pre-test)} = \frac{0.1}{0.9} = 0.1\bar{1} \text{ e } \text{LR}^+ = \frac{0.9}{0.1} = 9, \text{ odds (post-test)} = 0.1\bar{1} \times 9 = 1$$

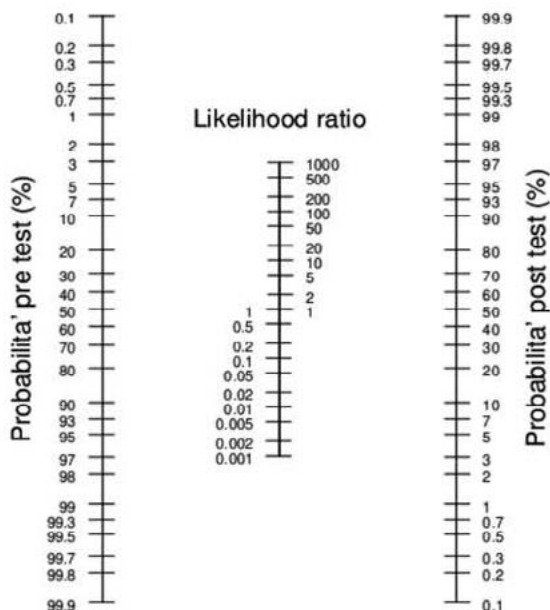
passando dagli odds alle probabilità  $P(\text{post - test}) = \frac{1}{1+1} = 0.5$ : in caso di esito positivo cioè vi è una probabilità del 50% che questo rifletta una condizione di malattia .

Per un risultato negativo, invece, il  $\text{LR}^- = \frac{0.1}{0.9} = 0.1\bar{1}$  e gli odds post test saranno ,

$0.1\bar{1} \times 0.1\bar{1} = 0.01235$  corrispondenti ad una probabilità post-test di  $0.01235/1.01235$  e cioè del 98.8% che è per l'appunto il valore predittivo per un risultato negativo.

La computazione di queste quantità, seppure non presenti alcuna difficoltà e possa essere effettuata anche senza l'uso del calcolatore, può essere ulteriormente semplificata con il ricorso ad uno strumento grafico, come il nomogramma di Fagan. Il suo uso è semplice: con un righello si prolunga un segmento che dall'asse della probabilità pre-test intercetta il valore di LR, per trovare il corrispondente valore di probabilità post-test.

Fig. 3: il nomogramma di Fagan.



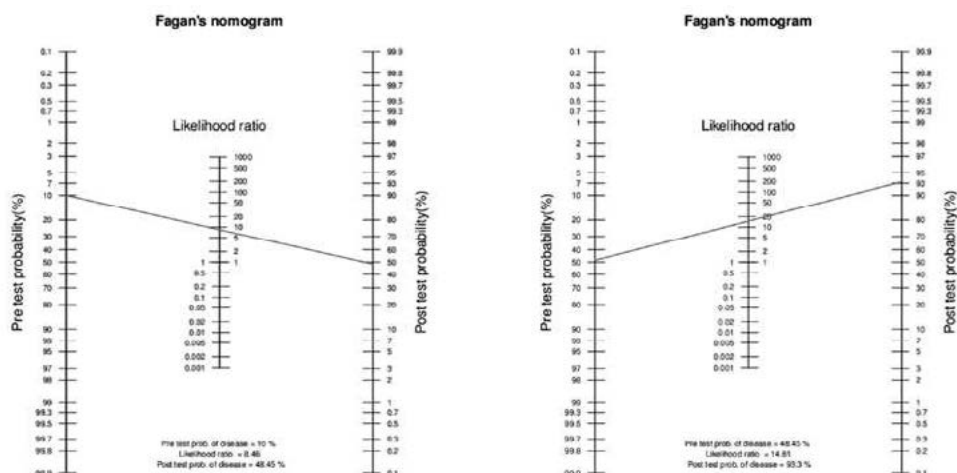
## L'utilizzo di più di un test

Quanto visto fino ad ora si riferisce all'utilizzo di un solo test. Nella realtà il clinico ricorre sempre a più di un test per porre una diagnosi. Seppure nella pratica la distinzione possa risultare meno netta, si riconoscono due strategie di somministrazione di test: in serie o in parallelo.

Se i test vengono somministrati in serie il soggetto è considerato ammalato solo quando è positivo a due (o più) test applicati in modo (con)sequenziale. In questo caso il primo test deve essere il più sensibile (ed economico) ed il secondo test deve essere il più specifico. La probabilità post-test del primo test diventa la probabilità pre-test del secondo test. Ad esempio (Straw et al. Diseases of Swine, IX ed., capitolo 10.) il Modified Agglutination Test (MAT) per *Toxoplasma gondii* ha  $Se=82.9\%$  e  $Sp=90.2\%$  ed il Latex Agglutination Test (LAT) ha  $Se=45.9\%$  e  $Sp=96.9\%$ , la prevalenza attesa, sulla base delle conoscenze del clinico è del 10%.

In caso di MAT negativo la probabilità di malattia (VPN) è del 2% circa (e concludiamo che è poco verosimile la presenza di malattia), concludendo qui gli accertamenti.

In caso di MAT test positivo la probabilità di malattia (VPP) è del 48.5% circa: esito che non possiamo ritenere conclusivo (Fig. 4, pannello sinistro).



**Fig. 4:** Utilizzo sequenziale del test MAT e LAT per *Toxoplasma gondii*.

se anche il LAT risulta positivo la probabilità post-test di malattia sarà del 93.3% circa (Fig. 4, pannello destro).

Nel caso dei test in parallelo due o più test sono somministrati contemporaneamente e il risultato viene ritenuto positivo se almeno uno dei test fornisce esito positivo. Questa strategia è adatta quando entrambi i test hanno scarsa sensibilità, in quanto aumenta la sensibilità ed il valore predittivo di un risultato negativo, mentre diminuiscono specificità e valore predittivo di un risultato positivo.

Si vuole qui accennare al caso, comune nella clinica di allevamento, della interpretazione

di test a livello aggregato, cioè applicati non ad un singolo animale ma ad un campione di una popolazione oggetto di indagine clinica. Si consideri che all'aumentare del numero dei soggetti testati (posto il numero critico di positivi per il quale classificare come ammalato il gruppo di animali ad 1) la specificità cala con il crescere del numero di soggetti esaminati: per  $n=2$  la probabilità di identificare correttamente i soggetti come sani sarà  $Sp^2$ , per  $n=3$  sarà  $Sp^3$  ed in generale la specificità di gruppo (Herd-Specificity) sarà  $HSp = Sp^n$ .

La sensibilità viceversa cresce all'aumentare di  $n$ :  $HSe = 1 - (1 - \text{Prevalenza apparente})^n$ . La prevalenza apparente: è data da  $PA = (Se \times \text{Prevalenza Vera}) + (1 - Sp) \times (1 - \text{Prevalenza Vera})$  dove i due prodotti a destra rappresentano la quota di veri positivi e di falsi positivi.

La prevalenza vera sarà invece 
$$\frac{PA + Sp - 1}{Se + Sp - 1}$$

## CONCLUSIONI

Quanto qui trattato vuole solamente essere uno stimolo per il clinico a riconoscere e a fare propri due importanti concetti della epidemiologia clinica, che dovrebbe tenere a mente nella sua pratica.

Il primo e più importante concetto del quale i clinici dovrebbero appropriarsi è che il processo diagnostico è sempre associato alla probabilità di errore diagnostico. I clinici dovrebbero inoltre essere costantemente coscienti del fatto che utilizzano il ragionamento bayesiano nel rivedere in modo iterativo le loro convinzioni a proposito della condizione clinica in esame alla luce degli esiti dei test.

Ciò comporta che i test possano essere interpretati solo alla luce della conoscenza (incertezza) a priori e che non sia in alcun modo possibile delegare ai test diagnostici stessi (in particolare alla diagnostica di laboratorio) la diagnosi, come talora si potrebbe sperare di fare né ad algoritmi che, seppure si debba riconoscere abbiano una qualche valenza didattica, non possono tenere in considerazione la componente probabilistica (cioè la componente di errore) insita nei procedimenti diagnostici.

## BIBLIOGRAFIA

- C J Gill, L Sabin, C H Schmid *Why clinicians are natural bayesians* BMJ Vol 330, 7 may 2005
- Straw et al. *Diseases of Swine*, 9<sup>th</sup> ed., capitolo 10.
- Noordhuizen et al. *Application of Quantitative Methods in Veterinary Epidemiology*, capitolo IV,4.
- Greiner, M., Gardner, I.A. *Application of diagnostic tests in veterinary epidemiologic studies*, Preventive Veterinary Medicine 45 (2000), 43-59
- Christensen, J., Gardner, I.A. *Herd-level interpretation of test results for epidemiologic studies of animal diseases*, Preventive Veterinary Medicine 45 (2000), 83-106