

L'efficacia diagnostica dei test nelle infezioni da *H.pylori*

G. Giocoli GdL EBM AMCLI

Corso Nazionale sulle Infezioni da *Helicobacter pylori* - Reggio Emilia 11.3.2003

Diagnostic efficacy of tests for *Helicobacter pylori* infection

Parole chiave: sensitivity and specificity, predictive value of tests, likelihood functions, multivariate analysis (tutte presenti nel dizionario dei MeSH terms di Pubmed)

SUMMARY

Different models are necessary in diagnostic research. The efficacy of tests must be studied first in terms of analytical accuracy and diagnostic efficacy, then in terms of clinical impact, where the value of a test is considered in terms of different patient outcomes. The issue of diagnostic efficacy is described here, by taking diagnosis of *Helicobacter pylori* infection as an example.

RELAZIONE

Il termine "test" non riguarda soltanto le prove di laboratorio. Infatti, ai fini della diagnosi, forniscono informazioni sulla salute del paziente qualsiasi prova strumentale, dalla più semplice alle più complesse - dal termometro alla TAC - e lo stesso esame clinico, con l'anamnesi e l'esame obiettivo. Tutti sono definibili "test" (1) e, come altri interventi sanitari, sono da considerare efficaci quando il paziente ne ritrae il massimo vantaggio e il minimo inconveniente (2). L'argomento di questo intervento è la verifica dell'efficacia diagnostica, in particolare dei test per l'infezione da *H.pylori*. Una corretta diagnosi è infatti uno dei contributi che un test può dare ad un esito clinico favorevole.

Prestazioni dei test

Andiamo ad elencare le prestazioni di un test che influiscono sulla sua efficacia (Figura 1).

Prestazioni dei test che influiscono sulla loro efficacia (Knottnerus, 2002)	tecniche	è analiticamente affidabile?
	diagnostiche	è utile per la diagnosi, di per sé o in aggiunta ad altri test?
	impatto clinico	è utile per le decisioni terapeutiche, il monitoraggio, la prognosi?
	impatto organizzativo	è conveniente?

Figura 1

Vi sono prestazioni analitiche, prestazioni diagnostiche, altre prestazioni che contribuiscono alla sua utilità clinica, prestazioni organizzative e di costo beneficio.

Prestazioni analitiche. E' certamente indispensabile che un test sia preciso, che identifichi con accuratezza il suo target (*l'Helicobacter pylori*), abbia un'adeguata sensibilità analitica, un idoneo range di misura, ecc. Tutte queste prestazioni, riassumibili nel termine di "accuratezza analitica", sono adeguatamente assicurate dal controllo di qualità.

Prestazioni diagnostiche. È però altrettanto importante che il test, oltre che il suo target analitico, sia capace di identificare la malattia. In altri termini che sia dotato di **accuratezza diagnostica**: essa dipende dalla sensibilità e dalla specificità.

Questi due attributi conferiscono al test un potere discriminante "nosologico", la capacità di distinguere in una popolazione chi è affetto da una malattia (*l'infezione da H. pylori e/o l'ulcera peptica*) da chi ne è esente. Inoltre gli conferiscono un potere "diagnostico" individuale, la **predittività**, idoneità a fornire nel singolo paziente risultati corrispondenti realmente alla presenza o all'assenza di malattia, in relazione alla sua prevalenza (3).

Il possesso di potere predittivo introduce il test nella sfera dell'utilità clinica. Non basta infatti che un test sia accurato: dev'essere anche utile. È questo il senso di quanto afferma Sackett (4): "L'accuratezza di un test diagnostico non è garanzia di utilità clinica: purtroppo nella letteratura si considera spesso l'accuratezza come un sinonimo di efficacia diagnostica".

A tal riguardo è necessario dire che di solito la diagnosi scaturisce dai risultati di diversi test; più che nel potere predittivo come test singolo, l'utilità consiste nel "valore aggiunto", il contributo del test a modificare in modo significativo la stima predittiva di un modello diagnostico (5).

Impatto clinico. L'efficacia diagnostica, a sua volta, è solo una componente dell'utilità clinica.

Nella malattia da *H.pylori*, un test si rivela efficace se non solo è in grado d'individuare la presenza o l'assenza della malattia, ma se riesce anche a contribuire al suo esito clinico favorevole (es. la scomparsa della dispepsia, la guarigione dall'ulcera, la soddisfazione del paziente in assenza di effetti collaterali).

LO STUDIO DELLE PRESTAZIONI DIAGNOSTICHE DI UN TEST

Per fare diagnosi d'infezione da *Helicobacter pylori* noi disponiamo di molti test (Figura 2), ma solo alcuni di essi forniscono prestazioni elevate (6, 7). Nella maggioranza dei casi il risultato permette solo delle previsioni parziali sull'esistenza dell'infezione perché l'accuratezza diagnostica è limitata.

INFEZIONE DA H.PYLORI
La capacità di un test di confermare o escludere l'infezione raggiunge raramente il 100%

	METODI INVASIVI Diagn Micr & Inf Dis 2002; 43: 251				METODI NON INVASIVI Fam.Practice 2000;17 (suppl 2):s12				
	NEST PCR	PCR	COL TURA	GIEMSA	C14 UBT	C13 UBT	SIEROL	AG FECI	POCT
Sens %	98	86	80	52	97	96	91	88	86
Spec %	100	100	100	92	95	96	90	92	75
VPP %	100	100	100	92	93	94	86	90	69
VPN %	98	91	87	73	98	97	95	91	85

Figura 2

L'accuratezza di un test diagnostico dicotomico (con due risultati: positivo e negativo) viene valutata con studi detti osservazionali. Essi si basano sull'osservazione dei risultati di un paragone tra il test in studio e un riferimento o gold standard e si effettuano con tabelle 2X2 (o matrici) (Figura 3).

LA MISURA DELL'ACCURATEZZA IN UNO STUDIO OSSERVAZIONALE

L'accuratezza analitica		Gold Stand. il metodo (la coltura)		L'accuratezza diagnostica		Gold Stand. la malattia (l'infezione da H.pylori)	
		Pos	Neg			Pres	Ass
Ag H pylori	Pos	VP	FP	Ag H pylori	Pos	VP	FP
	Neg	FN	VN		Neg	FN	VN

prima fase (ricerca di base) **seconda fase (ricerca applicata)**

Figura 3

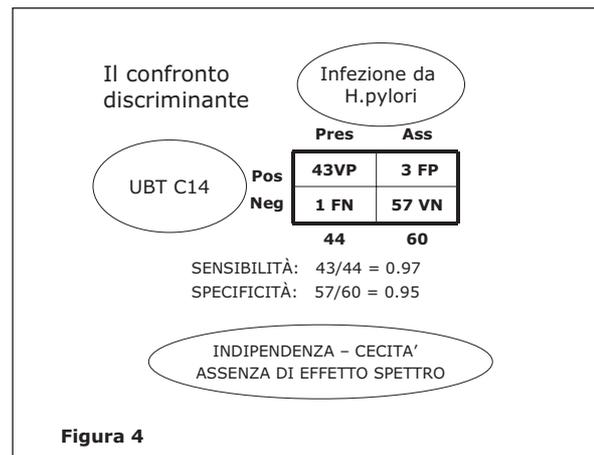
Nelle caselle (con posizioni fisse e standardizzate) si inseriscono Veri Positivi (VP), Falsi Positivi (FP), Falsi Negativi (FN), Veri negativi (VN), in

base alla corrispondenza tra risultati del test in studio e risultati del riferimento.

Nello studio di valutazione si procede per gradi: la prima fase è dedicata alle prestazioni tecnico-analitiche (8): lo studio è focalizzato sul patogeno, non sul paziente che lo ospita; il materiale del gold standard è uno stipe di Helicobacter e il metodo di riferimento è quello più affidabile, ad es. la coltura o un metodo molecolare. Nelle fasi successive si passa all'applicazione clinica. Il test viene studiato sulle persone e il gold standard diviene clinico: l'infezione da *Helicobacter pylori*. Per ottenere una sufficiente sensibilità, sono associati quasi sempre risultati di più metodi (es. coltura + ureasi + breath test) (9).

1. L'accuratezza diagnostica

In un esempio che riguarda l'urea breath test C14 (Figura 4), i parametri discriminanti sono ricavati dai rapporto percentuali tra i VP e il totale dei pazienti infetti (sensibilità) e tra i VN e il totale dei pazienti non infetti (specificità).



Per evitare stime viziate, il confronto deve seguire particolari regole: le principali sono l'indipendenza (entrambe le prove devono essere eseguite su tutti i campioni) e la cecità (chi esegue il test in studio deve ignorare il risultato del gold standard e viceversa) (1).

Inoltre bisogna evitare l'effetto spettro (10), associato alla composizione della popolazione, e ciò si può fare o con il campionamento consecutivo o con la stratificazione.

1.1 L'effetto spettro

L'effetto spettro si genera quando la stima dell'accuratezza è fornita da un confronto su una popolazione di pazienti troppo selezionati (Figura 5).

E' in causa uno studio sull'accuratezza di un test per svelare l'ulcera peptica in pazienti dispeptici: lo si può eseguire con l'adozione di un disegno "caso-controllo" oppure di un disegno "crociato trasversale" (cross-sectional).

Lo studio caso-controllo riguarda due gruppi di

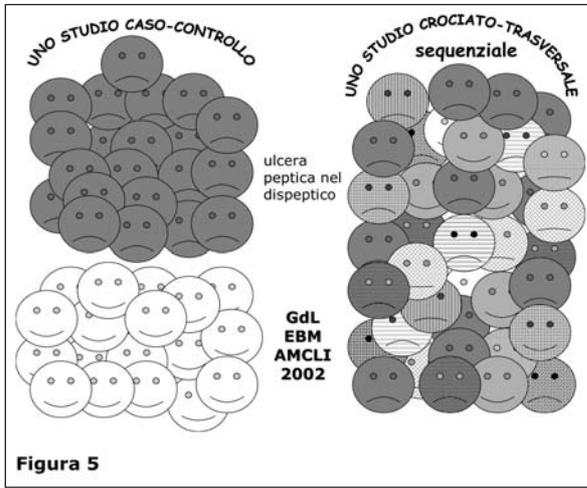


Figura 5

persone: uno affetto da ulcera allo stato conclamato e l'altro (sano) che fa da controllo. La selezione è massima: malattia o assenza di malattia; così l'effetto spettro. Lo studio *cross-sectional* include invece pazienti nei quali l'ulcera è sospetta, non accertata. L'arruolamento è consecutivo, cioè senza una selezione. E' il modo migliore per simulare le circostanze in cui il test sarà applicato clinicamente, su casi lievi o gravi ed anche su persone che hanno disturbi simili, ma di altra natura (es. disfagici, disfunzionali, epatopatici, enteropatici). Nello studio *cross-sectional* l'effetto spettro si riduce e la sua gerarchia nella scala delle evidenze è nettamente superiore, perché la casistica è molto più ricca e il test viene messo alla prova nella sua capacità di discriminare le persone con la patologia che interessa (11). L'accuratezza tende invece ad essere sopravvalutata negli studi caso-controllo, che sono utili, ma solo per una valutazione preliminare (12).

Uno studio crociato trasversale può essere stratificato, quando si cerca di includere nel campione una vasta gamma di dispeptici. La stratificazione è un surrogato della sequenzialità: sia l'una che l'altra assicurano che l'indagine sia praticata su una popolazione di pazienti il più possibile simile a quella che si presenta al medico.

2. Il potere predittivo

Con la determinazione del potere predittivo si entra nel capitolo dell'utilità clinica di un test. I valori predittivi si ricavano con un calcolo di cui si comprende facilmente la logica: il valore predittivo positivo (VPP) è il rapporto tra i veri positivi e il totale dei risultati positivi (veri e falsi), e il valore predittivo negativo (VPN) è il rapporto tra i veri negativi e il totale dei risultati negativi, (veri e falsi). Un grave inconveniente dei valori predittivi riportati negli articoli scientifici è che essi non sono validi per una popolazione con prevalenza di malattia diversa da quella del campione studiato. I confronti nella Figura 6 riguardano l'Urea

Breath Test (UBT) C14. Nelle due popolazioni, la prevalenza dell'infezione da Helicobacter è diversa (Paesi Bassi: 15%; Italia: 60%) (13). Supponendo che il potere discriminante dell'UBT sia invariato (sens 97% spec 95%) constatiamo

I valori predittivi e gli effetti della prevalenza

		Infezione da H.pylori			Infezione da H.pylori		
		Pres	Ass		Pres	Ass	
UBT C14	Pos	15 VP	4 FP	19	60 VP	2 FP	62
	Neg	1 FN	84 VN	85	2 FN	40 VN	42
		16	88	104	62	42	104

PREVALENZA: 16/104 = 0.15

VPP: 15/19 = 0.79

VPN: 84/85 = 0.99

PREVALENZA: 62/104 = 0.60

VPP: 60/62 = 0.97

VPN: 40/42 = 0.95

Figura 6

che i valori predittivi cambiano al cambiare della prevalenza.

Con l'aumentare di essa il VPP aumenta (da 79 a 97%), ma il VPN diminuisce (da 99 a 95%).

Per evitare equivoci e porre rimedio a questi inconvenienti è opportuno servirsi delle *Likelihood ratio* (LR), o rapporti di verosimiglianza.

2.1 Le Likelihood ratio

Le LR (Figura 7) sono indici del potere discriminante indipendenti dalla prevalenza e utili per calcolare la predittività. Il concetto è semplice: nei test dicotomici le LR sono due [LR(+) e LR(-)] e si calcolano facendo rispettivamente il rapporto percentuale VP/FP e FN/VN. Più alto è il rapporto % tra VP e FP e più aumenta la predittività positiva; più basso il rapporto % tra FN e VN, più aumenta la predittività negativa (2).

Le likelihood ratio sono indici indipendenti dalla prevalenza e utili per la stima della predittività

		Malattia	
		Pres	Ass
UBT C14	Pos	97% VP	5% FP
	Neg	3% FN	95% VN

Likelihood ratio (+) = 0.97/0.05 = 19

Likelihood ratio (-) = 0.03/0.95 = 0.03

Figura 7

Nella (Figura 8) sono elencate le LR positive e negative di vari test per l'infezione da Helicobacter. Sono considerati test di buona capacità discriminante quelli con LR(+) > 10 (%VP

maggiore più di 10 volte della %FP) e $LR(-) < 0.1$ (%FN inferiore di 10 volte o più alla %VN).

LR di test diagnostici per l'infezione da *H.pylori*

	METODI INVASIVI				METODI NON INVASIVI				
	NEST PCR	PCR	COL TURA	GIEM SA	C14 UBT	C13 UBT	SIER OL	AG FECI	POCT
LR (+)	inf	inf	inf	6	19	24	9	11	3
LR (-)	0.02	0.14	0.2	0.5	0.03	0.04	0.1	0.13	0.19

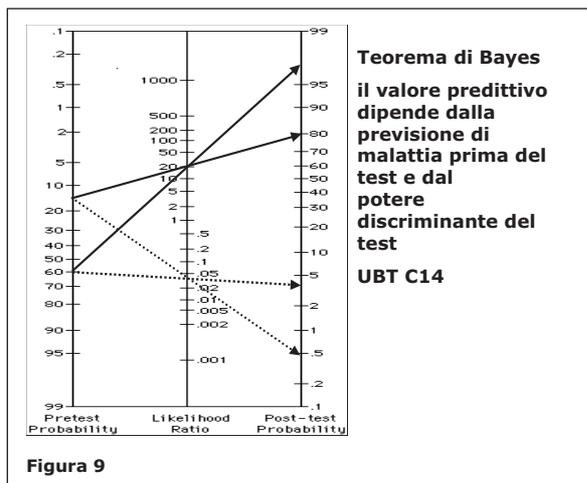
Figura 8

Una LR(+) ottimale è pari a infinito. Ciò accade quando la percentuale dei FP è uguale a zero (specificità = 100%). Una LR(-) ottimale è pari a zero. Ciò accade quando la percentuale dei FN è uguale a 0 (sensibilità = 100%). Un valore di LR uguale a 1 è indifferente.

Le LR servono ad applicare una regola intuitiva: ottenuto il risultato di un test, le probabilità di malattia dipendono dalla previsione di malattia prima del test e dal potere discriminante del test. In parole semplici, questo è il teorema di Bayes (14).

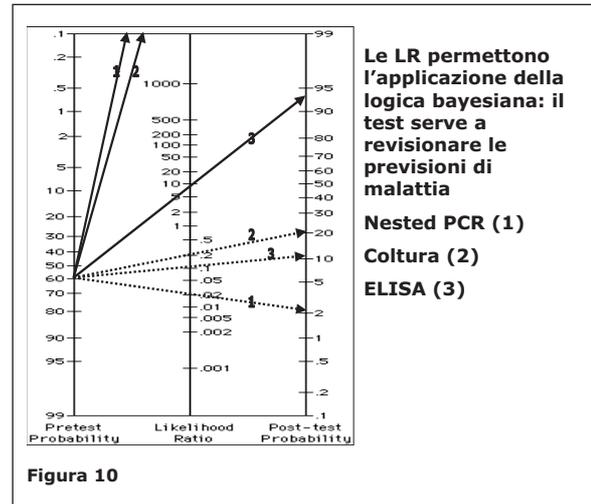
La previsione di malattia si basa sulla sua prevalenza. Questa ha due componenti, una epidemiologica (la sua frequenza in seno alla popolazione cui appartiene il paziente) e una fornita dalle probabilità a livello individuale (anamnesi, es. fisico, altri test ...).

Per convertire comodamente le previsioni di malattia in valori predittivi del test si può utilizzare il nomogramma di Fagan (Figura 9). Si traccia una retta che parte dalle "previsioni" (probabilità pre-test o prevalenza), passa attraverso le LR positiva o negativa e va a raggiungere la colonna di destra con i valori predittivi (probabilità post-test di malattia).



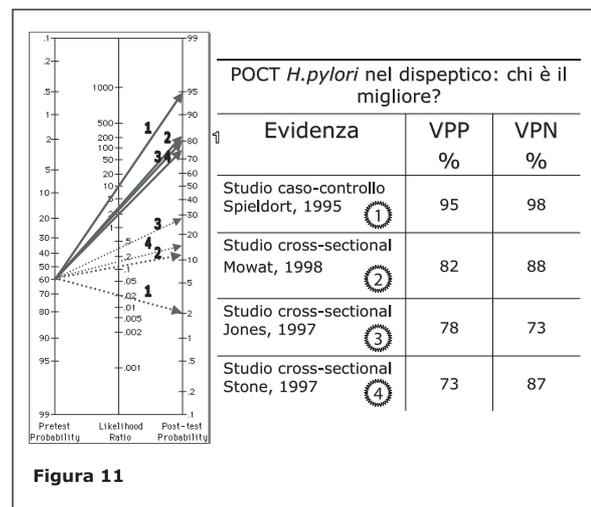
La figura mostra i valori predittivi dell'UBT quando la prevalenza è di ca. il 60% (Italia) e 15% (Paesi Bassi). Le frecce continue sono tracciate dalle prob. pre-test a quelle post-test attraverso l'LR(+), quelle punteggiate attraverso l'LR(-).

La (Figura 10) mostra i valori predittivi di altri test quando la prevalenza è di ca. il 60% (Italia). Le frecce dirette in alto e che non raggiungono la colonna di destra si riferiscono ai test le cui LR(+) sono pari a infinito.



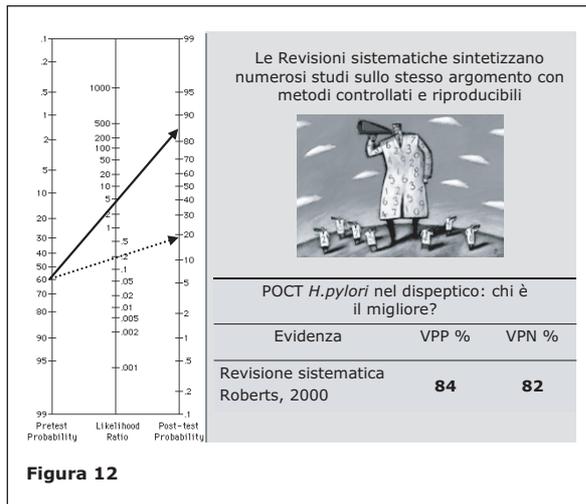
3. Le revisioni sistematiche

Sappiamo tutti che i parametri discriminanti e predittivi di un test sono quasi sempre diversi in studi diversi. Consideriamo le prestazioni di un test rapido EIA nel rivelare l'infezione da Helicobacter (Figura 11). I risultati migliori sono quelli di uno studio (numero 1) che ostenta un potere predittivo positivo del 95% e negativo del 98%. Ma perché quello studio dà risultati migliori degli altri? Si tratta di uno studio caso controllo, mentre gli altri sono studi crociato-trasversali (cross-sectional).



Le revisioni sistematiche sono sintesi critiche

della letteratura che – al contrario delle rassegne tradizionali o narrative – sono basate su metodi standardizzati e riproducibili, non su criteri personali. I revisori escludono i protocolli contenenti errori di vario tipo o cercano di correggerli, quando è possibile. In ogni caso non tengono in alcun conto gli studi caso-controllo. A volte la selezione può ridurre del 90% il numero degli studi inizialmente presi in considerazione. I risultati della revisione sistematica del test rapido per l'*Helicobacter* (Figura 12) (7 4) ci forniscono la valutazione depurata e ponderata del potere predittivo: discreto potere medio di conferma e di esclusione.



3.1 L'iniziativa STARD

Per rendere più facile ed efficace il lavoro dei revisori, nel gennaio del 2003 una diecina di riviste mediche hanno richiamato l'attenzione degli studiosi sul documento STARD, una sigla che sta per *Standards for Reporting of Diagnostic Accuracy* (15, 16). L'iniziativa ha lo scopo di aiutare gli autori degli studi di accuratezza diagnostica dei test ad esporre in modo completo e diligente i loro rapporti, con l'ausilio di una checklist e di un diagramma di flusso.

Nella (Figura 13) un particolare della checklist, che comprende 25 voci: sono richieste l'accurata enunciazione nel titolo dell'oggetto dell'indagine, le parole chiave da usare; il tipo di campionamento effettuato ... Bisogna specificare nelle varie sezioni del lavoro tutti quei particolari idonei a migliorare la comprensione del suo disegno e della sua logica, al fine di considerarne l'inclusione nelle revisioni sistematiche e nelle metanalisi.

4. Il valore aggiunto

È ben raro che un medico basi la sua diagnosi su un solo test, clinico o di laboratorio. Questo gli può essere utile solo se è in grado di aggiungere peso predittivo all'insieme dei dati di cui già dispone. Come già detto, questa capacità è quantiz-

LA CHECKLIST DELLO STARD	
Sezione e argomento	Descrizione
TITOLO/RIASSUNTO/PAROLE CHIAVE	Articolo sullo studio dell'accuratezza diagnostica (raccomandato il MeSH heading 'sensitivity and specificity')
INTRODUZIONE	Quesiti che sono oggetto di ricerca, p.es.: stima dell'accuratezza diagnostica o confronto dell'accuratezza tra test o gruppi di partecipanti
METODI	
Partecipanti	Campionamento dei partecipanti: è stato eseguito su una serie consecutiva di pazienti definiti con i criteri di selezione (3) e (4)? Se no, specificare come i pazienti sono stati selezionati.....

Figura 13

zabile come "valore aggiunto" e contribuisce alla sua utilità clinica (17).

Un esempio è visibile nella (Figura 14): alcuni gastroenterologi olandesi hanno determinato il valore aggiunto di un test rapido non invasivo (determinazione degli anticorpi anti-*Helicobacter* su sangue intero) nei confronti di un gruppo di dati anamnestici, per diagnosticare l'ulcera peptica. Il fine era la ricerca di criteri per attivare la strategia *test & treat* in medicina di base senza ricorrere all'endoscopia (18).

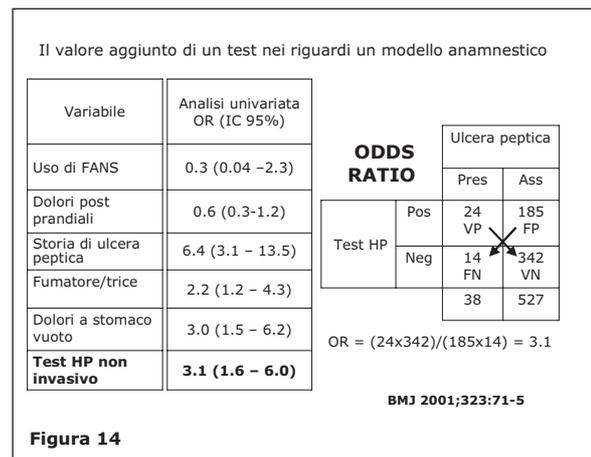


Figura 14

S'inizia con il calcolo delle *odds ratio* diagnostiche (OR) e dei loro intervalli di confidenza (IC) (a destra nella figura). Le OR esprimono il potere discriminante delle singole variabili (o determinanti) nei riguardi dell'ulcera peptica.

Le OR diagnostiche sono affini alla LR, ma hanno un valore meno selettivo e sono più adatte per questo tipo di confronti (analisi univariata e multivariata): per un test dicotomico equivalgono al rapporto tra il prodotto delle previsioni vere e il prodotto di quelle false – il che equivale al rapporto LR(+)/LR(-) (19).

Per quattro determinanti (a sinistra nella figura) esiste una significativa associazione con l'ulcera,

in quanto l'OR (compreso il suo intervallo di confidenza) è superiore a 1 (nel caso del test rapido l'OR è 3.1 con IC 1.5-6.2). Come per le LR, un valore uguale a 1 indica "nessuna associazione, discriminazione zero".

4.1 L'analisi multivariata

Bisogna capire però che in queste stime esistono interferenze da eliminare, in quanto le variabili s'influenzano a vicenda e bisogna controllare se sono veramente indipendenti (Figura 15). Nella parte sinistra della figura sono visibili le quattro variabili significative ottenute con l'analisi univariata. Per verificarne l'indipendenza si ricorre ad un'analisi multivariata: la determinazione delle OR non più per le singole variabili isolate ma per ciascuna di esse in quanto parte di un modello multiplo (parte destra della figura).

Indipendenza dell'associazione tra variabili e ulcera peptica			
Variabile	Analisi univariata OR (IC 95%)	Modelling multivariabile (regressione logistica) OR (IC 95%)	
		Modello anamnestico	Modello anamnestico potenziato
Storia di ulcera peptica	6.4 (3.1 - 13.5)	5.5 (2.6 - 11.8)	4.6 (2.1 - 10.1)
Fumatore/trice	2.2 (1.2 - 4.3)	2.0 (1.0 - 4.0)	1.9 (0.9 - 3.8)
Dolori a stomaco vuoto	3.0 (1.5 - 6.2)	2.8 (1.4 - 6.0)	2.8 (1.3 - 5.9)
Test HP non invasivo	3.1 (1.6 - 6.0)		2.7 (1.4 - 5.5)

Figura 15

Qui i modelli sono due, uno puramente anamnestico e uno potenziato: il precedente cui è aggiunto il test di cui si vuol calcolare il valore aggiunto. I calcoli si basano sulla regressione logistica multipla (che fa uso dei logaritmi naturali sfruttandone le proprietà additive) (19).

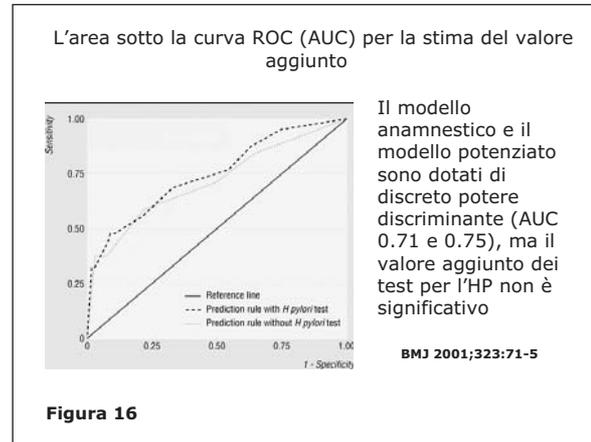
Per il primo modello predittivo "3 variabili anamnestiche" osserviamo come il potere discriminante dei determinanti s'indebolisce quando è eliminato l'influsso delle covariabili, presente nell'analisi univariata. Nel "modello potenziato", ottenuto aggiungendo la variabile "test non invasivo", le OR diminuiscono ancora, ma restano significative.

Si tratta dunque di variabili indipendenti e di per sé il test rapido è dotato di potere discriminante. Ma vediamo ora qual'è il valore aggiunto.

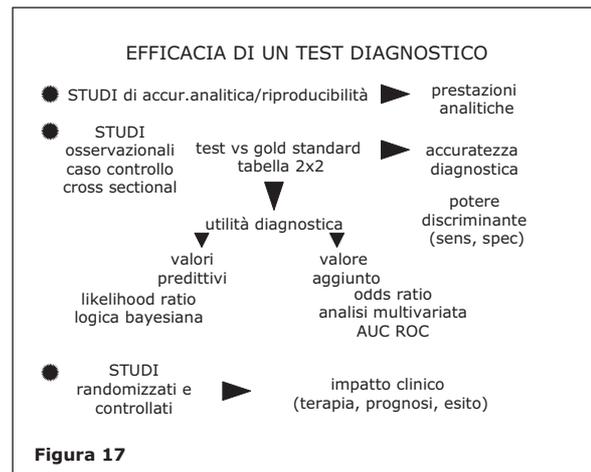
4.2 L'area sotto la curva ROC

Bisogna misurare il potere discriminante complessivo dei due modelli: per far questo si disegnano le curve ROC e si calcolano le "aree sotto la curva" (*area under the curve* = AUC) (Figura 16). Le AUC sono rispettivamente 0.71 e 0.75. Un valore di 0.71 significa che vi sono 71 probabilità

su 100 che il modello riesce a discriminare tra presenza e assenza di ulcera peptica. L'incremento da 0.71 a 0.75 (il 4%), tenendo conto degli intervalli di confidenza, non è significativo: il test rapido per l'*Helicobacter* non possiede sufficiente valore aggiunto (18).



5. Per ricapitolare (Figura 17): la stima dell'efficacia di un test diagnostico comprende diversi capitoli, ai quali corrispondono diversi tipi di studi.



Per valutare l'accuratezza analitica sono necessari studi di base per la sensibilità analitica (valore minimo determinabile), il range di misura, la linearità, la riproducibilità, ecc., prestazioni da assicurare in seguito con il controllo di qualità. L'efficacia diagnostica viene valutata con studi osservazionali che analizzano i risultati del confronto test - gold standard; essi comprendono l'accuratezza diagnostica (potere discriminante) e l'utilità diagnostica (potere predittivo). L'approfondimento dell'utilità clinica di un test si basa sugli studi d'impatto clinico, che valutano i risultati finali dell'uso di un test: gli effetti che una specifica diagnosi induce sulla terapia, sul-

l'impiego delle risorse mediche, sul gradimento del paziente, e infine sull'esito della malattia. Per la valutazione di questi *outcomes* (esiti clinici) sono necessari studi sperimentali (trial randomizzati e controllati) e bisogna dire che in campo diagnostico sono ancora pochi, o quanto meno non tanti quanti ne esistono nel campo della valutazione dell'efficacia dei farmaci.

La Figura 18 mostra un esempio di studio randomizzato e controllato sulle procedure diagnostiche nel paziente dispeptico (20). Qui si va molto oltre l'efficacia diagnostica del test. Gli esiti qui presi in considerazione, ai fini della valutazione dell'impatto clinico, sono: primo l'attenuazione della dispepsia, poi l'uso di altre risorse mediche (farmaci, ulteriori accertamenti), gradimento del paziente e innocuità.

I risultati sono gli stessi e l'efficacia globale delle due procedure è la stessa.

Studio d'impatto clinico

Uno studio randomizzato e controllato che confronta l'endoscopia associata al breath-test per *H. pylori* con il breath test da solo nel trattamento della dispepsia

K E L McColl, et al. BMJ 2002;324:999

I risultati sull'attenuazione della dispepsia, l'uso di risorse mediche, gradimento del paziente, innocuità, sono gli stessi

Figura 18

10. Mulherin SA. et al. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann.Intern.Med.* 2002; 137:598-602
11. Black ER et al. Diagnostic strategies for common medical problems 2nd. Ed. American College of Physicians, 1999.
12. Ljimer JG et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-6
13. De Wit NJ et al. Guidelines on the management of *H. pylori* in primary care: development of an implementation strategy. *Fam. Practice* 2000;17:S27-S32
14. Bossuyt PM. No burial for Bayes' rule. *Epidemiology* 1997. 8: 4-5
15. Bossuyt PM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41-4
16. The STARD Group. Towards complete and accurate reporting of studies on diagnostic accuracy: the STARD initiative. <http://www.consort-statement.org/stardstatement.htm> versione italiana: <http://www.gimbe.org/Link/STARD.zip>
17. Moons KGM. et al. Diagnostic studies as multivariable, prediction research. *J Epidemiol Community Health* 2002;56:337-8
18. Weijnen CF et al. Testing for *H. pylori* in dyspeptic patients suspected of peptic ulcer disease in primary care: cross sectional study. *BMJ* 2001; 323: 71-5
19. Habbema JDF et al. Analysis of data on the accuracy of diagnostic tests. In: JA Knottnerus. The evidence base of clinical diagnosis. 2002 BMJ Bookshop
20. McColl KEL, et al. Randomised trial of endoscopy with testing for *Helicobacter pylori* compared with non-invasive *H. pylori* testing alone in the management of dyspepsia. *BMJ* 2002;324:999

BIBLIOGRAFIA

1. Jaescke R et al. for the Evidence-Based Medicine Working Group. Users' Guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA* 1994; 271:389-91 B. What are the results and will they help me in caring my patients? *JAMA* 1994;271:703-7
2. Sackett DJ et al. Evidence-based Medicine (2nd Ed) 2000. Churchill Livingstone
3. Knottnerus JA. Challenges in dia-prognostic research - The relevance of design issues and new methods. *J.Epidem.Comm.Health* 2002;56:340-1
4. Sackett DL et al. Clinical Epidemiology. A basic science for clinical medicine (2nd Ed) 1991. Little Brown & Co.
5. Moons KGM et al. Diagnostic studies as multivariable, prediction research. *J.Epidemiol Community Health* 2002;56:337-8
6. Kisa O et al. The evaluation of diagnostic methods for the detection of *H. pylori* in gastric biopsy specimens. *Diagn Micr & Inf Dis* 2002; 43: 251
7. Roberts AP et al. Tests for *Helicobacter pylori* infection: a critical appraisal from primary care. *Fam. Practice* 2000;17:S12-S20
8. Price CP. Evidence-based laboratory medicine: supporting decision-making. *Clin Chem.* 2000;46:1041-50
9. Laheij RJF, et al. Evaluation of commercially availa-