

Gli strumenti della Evidence Based Medicine nella Medicina di Laboratorio

Romolo M Dorizzi, Marco Caputo

*Laboratorio Analisi Chimico-Cliniche ed Ematologiche,
Azienda Ospedaliera di Verona, Piazzale Stefani 1, 37126 Verona
E-mail: dorizzi@easynet.it*

Il concetto di Evidence Based Laboratory Medicine (EBLM) deve essere affrontato con grande attenzione. E' da più parti sostenuto che la medicina di laboratorio, usata male, risulti inutile ed addirittura controproducente per quanto riguarda la salute dei cittadini. L'introduzione della EBLM ci consentirebbe tre risultati: 1) il miglioramento della qualità degli esami di laboratorio; 2) il "raffreddamento" dell'aumento del numero delle procedure diagnostiche di efficacia non sufficientemente dimostrata; 3) l'adeguamento, alla richiesta, non solo delle Società di Assicurazione ma anche dei Servizi Sanitari Nazionali, di documentare che il costo degli esami sia proporzionato alla loro utilità.

APPROPRIATEZZA DEGLI ESAMI DI LABORATORIO

L'aumento degli esami diagnostici induce un aumento delle procedure terapeutiche anche se la loro utilità non è stata dimostrata. La letteratura sull'argomento è vasta. Per esempio Wenneberg et al hanno studiato la relazione tra numero di coronarografie eseguite e numero di procedure di rivascolarizzazione trovando una correlazione indipendentemente dalla prevalenza della patologia e dalla disponibilità di servizi chirurgici nella regione (1). In un altro studio è stato analizzato l'archivio del Medicare National Claims History relativo a circa 30 milioni di cittadini statunitensi con più di 65 anni. E' stato dimostrato un aumento dei test diagnostici (cateterismo cardiaco, imaging della colonna, mammografia, biopsia prostatica) compreso tra il 50% ed il 300% in 7 anni a cui corrispondeva un aumento delle procedure terapeutiche come rivascolarizzazione, interventi chirurgici alla colonna, gastrotomia percutanea, biopsia della mammella, prostatectomia (2). Anche se si ritiene che i benefici ottenuti dall'aumento delle spese diminuisce con l'aumentare del volume degli esami diagnostici non vi è dubbio che l'aumento degli esami rappresenta (quindi più indirettamente che direttamente) uno dei fattori principali dell'aumento degli interventi terapeutici

e quindi dei costi.

Recentemente sono stati condotti in Canada alcune valutazioni di cosa si debba intendere appropriato impiego del laboratorio sia per quanto riguarda la scelta, la frequenza e la temporizzazione della raccolta del campione sia per come si possa intervenire per modificare le modalità di richiesta degli esami di laboratorio (3). L'appropriatezza dell'impiego degli esami di laboratorio (di chimica clinica, di ematologia e di microbiologia) è stata valutata in situazioni diverse (al momento del ricovero ospedaliero, sia di elezione che in urgenza, in terapia intensiva, nella pratica ambulatoriale). La percentuale degli esami richiesti appropriatamente andava dal 5 al 95% e gli esami eseguiti più frequentemente in modo non appropriato erano test comuni come tempo di protrombina, calcemia, VDRL nel liquido cerebrospinale, determinazione della concentrazione nel siero dei farmaci antiepilettici. Lo stesso gruppo ha studiato come interventi diversi quali la produzione di linee guida, la modifica dei moduli di richiesta degli esami ed il cambiamento delle modalità di rimborso delle prestazioni hanno modificato la richiesta della velocità di eritrosedimentazione, l'esame microscopico delle urine, gli esami di funzionalità renale, quelli relativi allo studio dei depositi del ferro e di funzionalità tiroidea tra il 1991 ed il 1997 (4). Le richieste della VES ed urea sono diminuite rispettivamente del 58% e del 57% e la richiesta della determinazione del ferro è diminuita dell'80% e quella di ferritina del 34% (dopo la rimozione dal modulo di richiesta e la diffusione di linee guida sul corretto impiego di questi esami), la richiesta di analisi delle urine non accompagnate da esame microscopico è aumentato del 1700%, quella di analisi microscopica delle urine è diminuito del 14% (dopo la diffusione di linee guida), la determinazione della tiroxina totale è diminuita del 96% e quella del TSH del 12% (dopo rispettivamente l'interruzione del rimborso della prestazione e la rimozione dal modulo di richiesta degli esami). Questo studio era corredato da un editoriale di George Lundberg che rilevava come è confermata la possibilità di modificare l'uso degli esami se il direttore del laboratorio ha ben

chiaro come intervenire, se ha il coraggio di farlo ed opera in una struttura che lo sostiene in questa azione (5).

Un precedente editoriale dello stesso autore aveva ricordato come “la frequenza con cui un esame di laboratorio può essere richiesto va da zero al monitoraggio continuo di tutti i parametri. Qual è la frequenza ottimale: richiedere ognuno degli esami disponibili una volta nel corso della vita, una volta l’anno, ogni ora o mai? L’unica risposta è quella che non devono essere richiesti esami senza un piano di come reagire di fronte alla risposta. Si deve avere chiaro prima di chiedere un esame come reagire sia nel caso che il risultato sia all’interno dell’intervallo di riferimento, sia aumentato o sia diminuito” (6).

Applicare alla Medicina di Laboratorio la classica definizione della EBLM “l’esplicito e prudente impiego delle migliori prove di efficacia nel prendere decisioni circa la cura di un singolo paziente” richiede siano affrontate problematiche particolari legate alle intrinseche peculiarità di questa disciplina. L’EBM può insegnarci a tenere conto dei bias che possono influenzare i risultati ma nella diagnostica di laboratorio aspetti come la frequente mancanza di un “gold standard” assoluto con cui confrontare i nuovi esami e la grande diversità tra i valori numerici forniti da metodi diversi per la misurazione di uno stesso analita (basta pensare a troponina I e PSA) ostacolano l’applicabilità dei principi dell’EBLM.

L’obiettivo è quello di ricercare tipi diversi di prove di efficacia; non solo dati circa le prestazioni analitiche, dati relativi alla verifica interna ed esterna di qualità; dati circa sensibilità e specificità in particolari circostanze cliniche ma anche dati circa la differenza circa degli esami di laboratorio nella diagnosi e nel monitoraggio.

COME VALUTARE LA LETTERATURA RELATIVA AGLI ESAMI DIAGNOSTICI DI LABORATORIO

Vi sono almeno 7 standard metodologici negli esami diagnostici che si devono considerare quando si valutano articoli ad essi dedicati (7):

- **Spettro della composizione della popolazione studiata:** sensibilità e specificità di un esame dipendono dalle caratteristiche della popolazione studiata. Questo aspetto è rilevante; infatti la maggior parte degli esami diagnostici è valutata inizialmente in popolazioni con le forme più gravi di malattia e quindi i valori di sensibilità e specificità ottenuti non si applicano alle forme meno gravi di malattia in cui saranno applicati in genere nella pratica clinica.
- **Gruppi pertinenti:** Sensibilità e specificità possono rappresentare un valore medio per una popolazione. A meno che la condizione per la quale un

esame è utilizzato sia definita in modo molto ristretto, sensibilità e specificità possono variare in sottogruppi clinici diversi e vanno quindi personalizzati per i diversi sottogruppi. Un articolo, per essere adeguato, deve contenere degli indici di accuratezza personalizzati per ogni sottogruppo demografico o clinico pertinente (per esempio pazienti sintomatici vs pazienti asintomatici).

- **Bias nella conferma:** Si deve evitare che i pazienti in cui l’esame diagnostico ha dato un risultato positivo e quelli in cui ha dato un risultato negativo siano poi sottoposti a conferma con l’esame di riferimento in percentuale diversa. Tutti i soggetti devono essere sottoposti sia all’esame di riferimento (gold standard) che a quello in fase di verifica.
- **Bias nella revisione:** Deve essere evitato che esame gold standard e quello in fase di verifica non siano esaminati in modo obiettivo ed i risultati dei due esami vanno interpretati separatamente senza conoscere il risultato dell’altro.
- **Validità dei risultati relativi alla accuratezza dell’esame:** L’affidabilità della sensibilità e della specificità dipende dal numero dei pazienti che sono stati esaminati. La misura deve avere quindi una stima dell’intervallo di confidenza e questa deve essere indicata qualunque sia la sua entità.
- **Presentazione dei risultati indeterminati:** La presenza di risultati indeterminati è evento comune nella valutazione di un esame e la loro frequenza può limitare l’applicabilità di questo ovvero può aumentarne il costo in quanto richiede l’esecuzione di esami di conferma.
- **Riproducibilità dell’esame:** Deve essere sempre compiuta una stima della variabilità di un risultato e si deve sempre stimarne la causa.

Read et al (7) esaminano quanto 112 articoli relativi ad esami radiologici ed immunometrici pubblicati tra il 1978 ed il 1993 in giornali di grande prestigio come Lancet, British Medical Journal, New England Journal of Medicine e Journal of American Medical Association rispettino questi standard e concludono che solo il 24% di essi ne rispettavano 4 ed addirittura solo il 6% ne rispettava 6. La Figura 1 mostra comunque che la percentuale degli standard è aumentata tra il 1978 ed il 1993. L’utilità di una valutazione sistematica degli esami prima della loro introduzione assicura vantaggi importanti: 1) elimina esami di qualità scadente o nulla prima della loro introduzione; 2) migliora la qualità della informazione dell’esame diagnostico; 3) diminuisce i costi sanitari; 4) migliora la cura del paziente.

Il capitolo dedicato a diagnosi e screening del classico volume di Sackett et al (8) esordisce con l’affermazione che l’EBLM deve aiutare a rispondere a 3 domande:

- Le prove di accuratezza di un esame diagnostico sono valide?
- Queste prove (valide) dimostrano che questo

esame è in grado di distinguere in modo accurato i pazienti con una determinata malattia da quelli che non la presentano?

- Posso usare questo esame in un particolare paziente?

A seconda dei casi può essere preferibile trovare prima la risposta alla seconda domanda poiché se si conclude che l'impatto di un test è poco importante a chi interessa se è valido? D'altra parte può essere conveniente considerare prima la validità di un test; se un esame non è valido, cosa importa se è importante? E' invece ineludibile l'assunto che non può essere considerata l'applicabilità di un esame ad uno specifico paziente prima che i primi due punti non siano stati considerati.

Sackett accetta i punti considerati nell'articolo di Read circa la validità della valutazione di un esame diagnostico. La validità di un esame è provata se:

- E' stato confrontato in cieco con il metodo di riferimento (gold standard) per fare diagnosi. Si deve comunque essere cauti nel considerare il metodo di riferimento e tenere presente che perfino nella interpretazione di biopsie di mammella, cute e fegato la concordanza tra esperti non raggiunge il 50% (9). Un esame diagnostico classifica in un modo o nell'altro un risultato come normale o come anomalo. La normalità di un risultato secondo l'EBLM va inteso in senso diagnostico: non indica cioè un soggetto sano rispetto a quello malato, ma indica quell'ambito di risultati oltre il quale la patologia in esame diventa altamente probabile. La definizione gaussiana (media \pm 2 Deviazioni Standard) o in percentili (che possono andare dal percentile 2.5 al percentile 97.5 o dal percentile 5 al percentile 97.5) della normalità non è accettabile perché si concentra esclusivamente sul risultato del test diagnostico. Queste definizioni sono basate su proprietà statistiche che non solo implicano che tutte le anomalie si verificano con la stessa frequenza ma comportano che se eseguiamo un numero elevato di esami in un paziente aumentiamo progressivamente la probabilità di trovare un risultato anomalo e di portare quindi alla esecuzione di ulteriori esami (non appropriati). Il 64% dei soggetti in cui sono eseguiti 20 esami di laboratorio indipendenti ne presenta almeno uno al di fuori dell'intervallo di riferimento [$100 * (1 - 0.95^{20}) = 1 - 0.358 = 100 * 0.642 = 64.2\%$ (10).
- La modalità di riferimento è stata applicata a tutti i pazienti (sia in quelli in cui l'esame in valutazione è risultato positivo che in quelli in cui è risultato positivo).
- L'esame è stato valutato in uno spettro appropriato di pazienti.
- L'esame è stato validato in un secondo (ed indipendente) gruppo di pazienti. Gli esami diagnostici servono a predire non a spiegare una diagnosi. La prima valutazione di un esame non consente di

distinguere tra reale accuratezza diagnostica per una determinata malattia e associazione causale relativa solo al particolare gruppo iniziale di pazienti. Il modo migliore per valutare l'accuratezza di un esame è quello di dimostrare livelli simili di accuratezza in un secondo gruppo di pazienti. Nel caso questo non sia stato fatto è opportuno riservarsi il giudizio sull'esame.

COME VALUTARE LA CAPACITA' DI UN ESAME DI DISTINGUERE IN MODO ACCURATO I PAZIENTI AFFETTI O MENO DA UNA PARTICOLARE MALATTIA

Sensibilità', Specificità' e Quoziente di probabilità'

La EBLM si concentra sui classici concetti di sensibilità e specificità solo come mezzo per calcolare il Quoziente di Probabilità (Likelihood Ratio, LR), lo strumento cruciale che consente di calcolare come la probabilità di una diagnosi sia modificata dal risultato di un esame (ovvero come si calcola la probabilità post-test a partire dalla probabilità pre-test). Questo è importante dal punto di vista pratico sia perché questi strumenti possono consentire di valutare se un esame diagnostico può essere utile ovvero quale esame è da preferire quando ve ne siano numerosi disponibili per risolvere un quesito diagnostico. Nelle Tabelle I e II è sintetizzato un classico esempio su come la determinazione della concentrazione della ferritina nel siero possa essere impiegato per diagnosticare l'anemia sideropenica. Impiegando i dati provenienti da un articolo di Guyatt et al viene riassunto come sono calcolati Sensibilità, Specificità, Valore Predittivo Positivo e Negativo, Quoziente di Probabilità Positivo e Negativo, Odds Pre- e Post-test e Probabilità Post-test (11). La Tabella I ci ricorda che la sensibilità esprime la percentuale di pazienti con anemia sideropenica con una concentrazione di ferritina inferiore a 30 mg/L e che il 15% dei pazienti con anemia da altra causa presentano una concentrazione di ferritina inferiore a 30 mg/L. Questo equivale a dire che la concentrazione della ferritina è inferiore a 30 mg/L 6 volte più frequentemente (90%/15%) nei pazienti con anemia sideropenica rispetto ai soggetti che non presentano questa patologia; questo concetto è definito Quoziente di probabilità Positivo (LR+) e può essere calcolato dividendo la sensibilità per il reciproco della specificità.

Quello che interessa il clinico è però quale è la probabilità che un paziente con una concentrazione, per esempio, di 26 mg/L di ferritina sia affetto da anemia sideropenica. In questo caso deve essere tenuto conto della prevalenza della malattia [nel caso dell'esempio si può calcolare come il rapporto tra malati con anemia sideropenica (809) e tutta la popolazione studiata (2579) che porta ad una

prevalenza del 31%; gli Odds pretest sono dati dal rapporto tra prevalenza e reciproco della prevalenza $= 31/69\% = 0.45$]. Se moltiplichiamo gli Odds pretest per il Quoziente di Probabilità ($0.45 * 6$) otteniamo gli Odds post-test (2.7); vale a dire gli Odds (la probabilità) post-test è a favore della anemia sideropenica nel rapporto di 2.7:1. Questa può essere convertita in probabilità con il rapporto $2.7/(2.7+1) = 2.7/3.7 = 73\%$ (Tabella II).

L'LR consente anche di tenere conto di come la concentrazione più o meno alta del risultato influenza la probabilità di un risultato. Mentre impiegando i concetti di sensibilità (90% nell'esempio) e specificità (85% nell'esempio) il risultato ha due soli livelli (positivo e negativo), l'LR ci consente di attribuire un diverso peso a seconda del risultato numerico ottenuto; l'esempio in Tabella III mostra come un risultato molto positivo porta ad un LR molto elevato (52) che modifica in modo importante i risultati dell'esempio prima citato. Gli Odds post-test diventano infatti 23.4:1 con una probabilità post-test del 96%. Quando la prevalenza della malattia è dell'ordine di grandezza dell'esempio soprariportato un test con una LR superiore a 10 raggiunge una probabilità post-test superiore all'80% e quindi utile dal punto di vista diagnostico. LR inferiori a 0.1 quando la probabilità pre-test è quella dell'esempio, la probabilità post-test è molto bassa, inferiore al 5% (Odds post-test = $0.45 * 0.1 = 0.045$; probabilità post-test $0.045/1.045 = 4.5\%$). Jaeschke et al (12) classificano grossolanamente il significato nella pratica diagnostica del LR nel modo seguente:

1. $LR > 10$ o < 0.1 modificano in modo spesso conclusivo la probabilità di malattia
2. $LR > 5$ e < 10 o > 0.1 e < 0.2 modificano in modo discreto la probabilità di malattia
3. $LR > 2$ e < 5 o > 0.2 e < 0.5 modificano modestamente (talvolta importante) la probabilità di malattia
4. $LR > 1$ e < 2 o > 0.5 e < 1 modificano scarsamente la probabilità di malattia e solo di rado sono importanti.

Va d'altra parte ricordato che in molti casi, e quello della ferritina è uno di questi, una sola tabella non può essere applicata a pazienti di ogni età, maschi e femmine ma vanno personalizzate di conseguenza. Questa limitazione ostacola l'impiego di questo tipo di approccio in clinica a meno che molti centri, sia diagnostici che clinici, non collaborino nella definizione di queste tabelle..

Un modo ulteriore per valutare un esame diagnostico è l'NND (Number Needed to Diagnose = numero di test necessari per ottenere una diagnosi positiva) che si calcola conoscendo la sensibilità e specificità di un esame. L'NND si ottiene infatti calcolando il reciproco della differenza tra sensibilità e reciproco della specificità (13). Questo concetto deriva, infatti, da quello del NNT (Number Needed to Treat) che è dato dal reciproco della percentuale dei soggetti

migliorati dopo trattamento con il farmaco investigato meno la percentuale dei soggetti migliorati senza trattamento. Un calcolo analogo per l'esame diagnostico è dato dal reciproco della percentuale di risultati positivi nei malati (SENSIBILITA') meno la percentuale di risultati positivi nei soggetti non malati (RECIPROCO DELLA SPECIFICITA')(14)[$NND = 1/ \text{SENSIBILITA}' - (1 - \text{SPECIFICITA}')$]. Nell'esempio indicato nella Tabella IV l'aumento della concentrazione di ferritina porta ad un aumento dell'NND, anche se al di sopra di 29 mg/L, l'LR si abbassa al di sotto di 1 e si ottiene un valore di NDD negativo. Nella pratica l'NND può essere utile nel confrontare il rapporto costo beneficio degli esami di laboratorio, soprattutto nei casi in cui più esami possono essere impiegati in un determinato contesto clinico. In effetti un problema può essere determinato dalla negativizzazione del valore nel caso di un esame con modeste prestazioni diagnostiche.

Nell'appendice sono riportati esempi di LR positivi e negativi (LR-); l'LR- può essere facilmente calcolato con il rapporto tra il reciproco della sensibilità e la specificità (nel caso dell'esempio prima considerato $10/85 = 0.12$).

Uno degli aspetti più rilevanti della EBLM è l'approccio pratico che introduce nell'impiego dei risultati di laboratorio nella gestione del singolo paziente. Due esempi di questo approccio sono gli acronimi SnNout e SpPin (15) ed il nomogramma di Fagan (16).

I due acronimi rendono più facile ricordare, soprattutto per i non addetti ai lavori, quale è l'impiego ottimale di un esame diagnostico molto sensibile o molto specifico.

SnNout è un acronimo inglese che deriva da When a test has a very high **S**ensitivity, a **N**egative result rules **out** the diagnosis (*Quando un esame ha una sensibilità molto elevata un risultato negativo virtualmente esclude la diagnosi*); **SpPin** deriva invece da When a test has a very high **S**pecificity, a **P**ositive result rules **in** the diagnosis (*Quando un esame ha una specificità molto elevata un risultato positivo virtualmente conferma la diagnosi*). Formule di questo tipo consentono di avere sempre presente quale è la interpretazione di uno dei numerosi test diagnostici che il clinico ha necessità di gestire quotidianamente fornendogli un modo "automatico" per dare a test diversi un "peso" diverso ma, anche se grossolanamente, efficace.

Il nomogramma di Fagan consente di semplificare molto la conversione della probabilità pre-test in quella post-test. Il nomogramma si usa ancorando una linea retta in corrispondenza della probabilità pre-test indicata sull'asse di sinistra e ruotandola finché raggiunge l'LR relativo all'esame che si sta considerando indicato nell'asse al centro (Figura 1). E' sufficiente proseguire la retta e si troverà sull'asse di destra la probabilità post-test (17). Il nomo-

gramma può essere tenuto nella tasca del clinico e del laboratorista per un impiego immediato.

I vantaggi dell'impiego dei LR e del nomogramma di Fagan sono numerosi; per esempio:

1. sono più comprensibili e più facili da usare rispetto a sensibilità e specificità;
2. possono essere calcolati a livelli diversi di risultato;
3. possono essere usati in modo sequenziale in modo che la probabilità post-test che si ottiene da un esame può diventare la probabilità pre-test per il successivo;
4. consentono di embricare medicina di laboratorio e clinica in modo molto efficace;
5. forniscono gli strumenti per introdurre forme di audit clinici nella pratica. Consente, infatti di valutare l'efficacia diagnostica degli esami richiesti dal clinico ed eseguiti dal laboratorio

Uno schema che risulta assai utile nella pratica clinica è quello indicato in Figura 2 che rappresenta la soglia di probabilità (A) superata la quale il clinico richiede un esame e quella (B) superata la quale inizia il trattamento.

Dalla Figura emerge che se l'LR di un esame è molto basso, la probabilità di malattia diventa così bassa che tale ipotesi diagnostica può essere abbandonata e si devono considerare altre ipotesi diagnostiche; un risultato negativo può consentire quindi di non oltrepassare la soglia dell'esecuzione di esami (A). Un risultato positivo che produce un elevato LR aumenta talmente la probabilità post-test da consentire di non eseguire ulteriori esami e di passare alla terapia; è stata quindi oltrepassata la soglia del trattamento. Se il risultato del nostro test ci lascia nell'area compresa tra A e B, sono necessari altri esami per cercare di arrivare ad oltrepassare la soglia del trattamento. L'abilità del clinico sta nel calcolare intuitivamente delle soglie per l'esecuzione dell'esame e per l'inizio del trattamento che siano corrette.

Un aspetto che Sackett sottolinea (8) è che nella pratica clinica vengono eseguiti numerosi esami diagnostici ed in questo caso possiamo calcolare la probabilità post-test complessiva semplicemente iterando il procedimento prima descritto un numero di volte pari al numero di esami eseguito. Gli odds post test ottenuti dopo l'esecuzione del primo esame rappresentano gli odds pre-test del secondo esame eseguito e così via. Non dobbiamo fare altro che continuare a moltiplicare il prodotto ottenuto per l'LR generato dall'esame successivo. Se un paziente che si presenta dal medico ha una probabilità pre-test del 6% di coronaropatia in cui l'anamnesi positiva produce un LR di 13 e l'ECG dopo sforzo produce un LR di 11, la probabilità post-test di stenosi coronarica è data dalla probabilità pre-test moltiplicata per le due LR

$$0.06/0.94 = 0.06 * 13 = 0.83 * 11 = 9.13; 9.13/10.13 = 90\%.$$

Va tenuto conto che questi calcoli sono del tutto accurati solo se gli esami diagnostici sono "indipendenti" e questa evenienza nella pratica clinica è ottenuta solo di rado. Comunque questo esempio ci dimostra come il ragionamento diagnostico, che si solito è un procedimento implicito, può essere esplicitato ed oggettivizzato.

CONCLUSIONI

Il clinico usa dei metodi impliciti per usare i risultati degli esami nella propria pratica. Se un esame è chiaramente positivo e negativo, egli può decidere di usarlo o meno per orientare il suo intervento diagnostico e terapeutico; se il risultato è indeterminato può eseguire ulteriori esami per chiarire il quadro. La EBLM può consentire da una parte di dare maggiore evidenza a queste decisioni e dall'altra di aiutare il clinico meno esperto nell'interpretare i risultati di laboratorio in modo efficiente.

Di fronte ad un esame il clinico deve chiedersi se questo sarà in grado di aiutare a risolvere un problema diagnostico, clinico o terapeutico; per fare questo dovrà considerare sia l'esame stesso sia il contesto clinico in cui esso sarà usato. Il clinico dovrà compiere una rassegna sistematica sull'argomento e dovrà verificare (17):

1. se l'esame è disponibile, accurato, riproducibile ed accessibile nel contesto in cui opera;
2. qual'è la probabilità pre-test;
3. se la probabilità post-test che si ottiene è in grado di modificare la gestione del paziente;
4. se le conseguenze mediche di un particolare esame sono accettabili per il paziente.

Il laboratorio di qualunque tipologia e dimensione deve attrezzarsi per partecipare al processo che porta alla risposta di questi quesiti. E' auspicabile che l'EBLM entri nella attività quotidiana e che alla conoscenza di concetti come sensibilità e specificità si aggiungano concetti come probabilità pre- e post-test e quoziente di probabilità. Questi possono aiutare a valutare l'efficacia degli interventi diagnostici in modo più obiettivo. Introdurre gli strumenti dell'EBLM nell'attività quotidiana del laboratorio consente infatti di valutare gli esami diagnostici (sia tra quelli già in uso che quelli in via di introduzione) dotati della maggiore efficienza diagnostica ed affidabilità, meglio tollerati dal paziente e più vantaggiosi dal punto di vista economico. Su questo versante la diagnostica ha sicuramente fatto meno progressi rispetto alla terapia come è stato dimostrato dal basso numero di articoli dedicati ad esami che rispettano standard rigorosi di qualità pubblicati anche da giornali di grande prestigio (7).

D'altra parte non è pensabile che oggi il singolo laboratorio sia attrezzato con sofisticate tecnologie informatiche come recentemente descritto dal gruppo che opera presso i National Institutes of Health gui-

dati da Zweig che hanno recentemente descritto i Prevalence-Value-Accuracy-Plots. Questi grafici considerano nella valutazione degli esami diagnostici, oltre all'accuratezza del metodo, l'effetto della prevalenza ed il costo della misclassificazione (18). Rimane da considerare il punto che gli strumenti dell'EBLM sono utili in ambito diagnostico che non rappresenta la quota prevalente dell'attività del laboratorio, soprattutto rivolta al monitoraggio di una determinata condizione o di una terapia e nei quali altri concetti, quali intervallo di riferimento, livelli decisionali, variabilità biologica o differenza critica sono spesso determinanti.

NOTA ALL'ESERCIZIO

L'esercizio è ricavato da uno studio pubblicato nel 1978 da Casscells et al (19).

A 20 studenti di Medicina del 4 anno, 20 strutturati che operavano in reparti medici e 20 medici non strutturati incontrati casualmente in Ospedale era posto questo quesito: Se un esame ha una sensibilità ed una specificità del 95% per una patologia, quale è la probabilità di un paziente di soffrire di quella malattia nel caso l'esame risulti positivo e la prevalenza sia dell'1/1000.

Esistono molti modi rapidi per la soluzione:

- 1. empirico; il senso comune ci dice che solo 1 su 1000 soggetti sarà, in media, malato; il 5% degli altri ($0.05 * 999 = \approx 50$) darà un risultato falso positivo poiché la specificità è del 95%. Solo 1 su 51 soggetti ($\approx 2\%$) sarà malato.*
- 2. Si calcolano gli odds pre-test ($0.001/0.999 = 0.001001$); si calcola l'LR + ($95/5 = 19$); si calcolano poi gli odds post-test ($0.001001 * 19 = 0.019$) e la probabilità post-test ($0.019/1.019 = 0.0186 * 100 = 2\%$)*
- 3. Si calcola l'LR + ($95/5 = 19$) e poi si calcola la probabilità post-test utilizzando il nomogramma di Fagan.*

Nello studio citato 11 su 60 (18%) partecipanti al sondaggio indicavano la risposta corretta; 4 su 20 studenti di medicina; 3 su 20 degli strutturati e 4 su 20 degli altri medici. La risposta più comune, data da 27, è stata 95% e le risposte andavano da 0.095% a 99%. La media di tutte le risposte è stata 55.9%, una sovrastima di 30 volte della reale probabilità

L'articolo concludeva che una della cause del cattivo uso del laboratorio sembra risiedere nella cattiva interpretazione dei dati di laboratorio ed auspicava che piuttosto che uno studio formale dell'analisi decisionale fosse promossa la conoscenza di nozioni di tipo pratico per l'interpretazione delle analisi di laboratorio.

Questo auspicio sembra poco realizzato come dimostra una recente indagine compiuta negli Stati Uniti che valutava l'uso di metodi formali come tenere conto di sensibilità, specificità, valore predittivo

positivo, LR, e curve ROC (Receiving Operating Characteristics) da parte di medici di 6 specialità diverse che dedicavano almeno il 40% del tempo in attività assistenziale (20). Attraverso domande del tipo "Quando richiedete ed interpretate esami di laboratorio considerate i valori di sensibilità e specificità?" gli autori concludevano che i metodi bayesiani erano usati dal 3% dei medici e curve ROC ed LR dall'1% (nessuno dei ginecologi e dei medici di famiglia usavano questi strumenti). Secondo gli autori i risultati degli esami di laboratorio non sono presentati in modo intuitivo ed utile per il clinico. Le informazioni sugli esami devono essere "immediatamente disponibili nel momento in cui sono richiesti; deve essere aumentato l'addestramento formale in questo ambito; le informazioni della letteratura sono utili solo se riflettono la popolazione in cui debbono essere applicate dal clinico.

L'EBLM ha anche lo scopo di diffondere questi strumenti di lavoro sia tra i laboratoristi sia tra i clinici in modo che il peso delle informazioni che il laboratorio fornisce sia compreso e, soprattutto, impiegato nel modo corretto. L'articolo si conclude con una citazione da Peabody degli anni 20: "La buona medicina non consiste nell'indiscriminata esecuzione degli esami di laboratorio, ma piuttosto in una chiara comprensione delle probabilità di una patologia e di quale esame può essere utile.... Ogni ospedale ha il dovere di impedire che nessun medico riceva il suo diploma a meno che abbia dimostrato di conoscere come usare i risultati degli esami nella cura dei suoi pazienti".

BIBLIOGRAFIA

1. Wenneberg DE, Kellet, Dickens JD, Malenka DJ, Keilson LM, Keller RB. The association between local diagnostic testing and invasive cardiac procedures. *JAMA* 1996; 275: 1161-4.
2. Verrilli D, Welch G. The impact of diagnostic testing on therapeutic interventions. *JAMA* 1996; 275: 1189-91.
3. Van Walraven C, Naylor D. Do we know what inappropriate laboratory utilization is? *JAMA* 1998; 280: 550-8.
4. Van Walraven C, Goel V, Chan B. Effect of population-based interventions of laboratory utilization. *JAMA* 1998; 280: 2028-33.
5. Lundberg G. Changing physician behavior in ordering diagnostic tests. *JAMA* 1998; 280: 2036.
6. Lundberg G. The need for an outcomes research agenda for clinical laboratory testing. 1998; 280: 565-6.
7. Read MC, Lachs Ms, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA* 1995; 274: 645-51.
8. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Hayes RB. Evidence-Based Medicine. 2nd Ed. Edinburgh: Churchill Livingstone; 2000.
9. Anonimo. Melanoma diagnosis. *Bandolier* 1998; 5 (1): 8.
10. Solberg HE. Establishment and use of reference values. In: Burtis CA, Ashwood ER (eds): Tietz

- Textbook of Clinical Chemistry. Philadelphia: Saunders; 1999.
11. Guyatt GH, Oxman AD, Ali M, William A, Mallory W, Patterson C. Laboratory diagnosis of iron-deficiency anemia: an overview. *J Gen Intern Med* 1992; 7: 145-53.
 12. Jaeschke R, Guyatt GR, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. *JAMA* 1994; 271: 703-7.
 13. Galloway MJ, Reid MM. Is the practice of hematology evidence based? III. Evidence based diagnostic testing. *J Clin Pathol* 1998; 51: 489-92.
 14. Anonimo. How good is a test II. *Bandolier* 1996; 27: 64-6.
 15. Fleming K. Evidence based pathology. *Evidence-Based Medicine* 1997; 2: 132
 16. Fagan TJ. Nomogram for Bayes's theorem. *N Engl J Med* 1975; 293:257.
 17. Sackett DL, Straus S. On some clinically useful measures of the accuracy of diagnostic tests. *Evidence-Based Medicine* 1998; 3: 68-70.
 18. Remaley AT, Sampson ML, De Leo JM, Remaley NA, Farsi BD, Zweig MH. Prevalence-value-accuracy plots: a new method for comparing diagnostic tests based on misclassification costs. *Clin Chem* 1999; 45: 934-41.
 19. Casscells W, Schoenberger A, Graboys TB. Interpretation by physician of clinical laboratory results. *N Engl J Med* 1978; 299: 999-1001.
 20. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgements: practising physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; 104: 374-80.
 21. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical decision making. Boston: Butterworths; 1988.
 22. Black ER, Bordley DR, Tape TG, Panzer RJ. Diagnostic strategies. Philadelphia: American College of Physicians; 1999.
 23. Dorizzi RM, Senna GE, Caputo M. Likelihood ratios and Fagan's nomogram: valuable but underrated tools for in vitro latex sensitisation assessment. *Chim Clin Acta* 1999; 282: 175-83.
 24. Dorizzi RM, Caputo M, Morando G. Algorithms or nomograms for the busy physician? *Am Heart J* 2000

Appendice. Esempi di Quoziente di Probabilità positivo (LR+) e negativo (LR-) (21-24)

ESAME/CONDIZIONE CLINICA	LR+	LR-
ADDOME IN BIANCO Appendicite	4	0.67
Occlusione	10	0
Colelitiasi	2.5	0.83
Urolitiasi	4	0.67
ACE Sarcoidosi I	12	0.41
Sarcoidosi II	15	0.24
Sarcoidosi III	19	0.07
Arteriografia addominale Aneurisma aorta addominale	14	0.32
Cancro pancreas	6.5	0.39
Sanguinamento gastrointestinale cronico	> 45	< 0.56
Stenosi arteria renale	> 100	0
Arteriografia cerebrale Aneurisma con emorragia subaracnoidea	> 95	< 0.05
Massa/Tumore	> 96	< 0.04
Tubo digerente Cancro colon	19	0.05
Polipi, < 1 cm	5.4	0.77
Polipi, 1-2 cm	16.6	0.18
Polipi, > 2 cm	18	0.11
Osso radiografia, metastasi Cancro mammella	> 66	0.34
Cancro polmone	> 86	0.65
Cancro melanoma	90	0.4
Cancro mieloma	> 97	0.09
Cancro ovaio	> 100	0
Cancro prostata	> 44	0.57
Cancro rene	> 7.1	0.32
Osso scintigrafia, metastasi Cancro mammella	6	0.05
Cancro polmone	32.3	0.03
Cancro prostata	> 99	<0.01
Cancro mammella Esame del medico	3.8	0.69
Mammografia	7.5-8.7	0.14-0.28
Termografia	1.67	0.71
CEA Cancro colon Duke A	1.6	0.87
Duke B	2.6	0.66
Duke C	4.4	0.3
Duke D	4.9	0.19
Radiografia torace Cancro polmone	15	0.42
Adenopatia ilare	5.2	0.14
Adenopatia mediastinica	2.4	0.57
Tubercolosi attiva	9	0.49
Colonscopia Polipi	> 90	< 0.1
Cancro	> 90	< 0.1

Digossina Intossicazione (> 2 mg/L)	5.7	0.16
Gravidanza ectopica HCG	100	0
VES (> 20 mm/h) Endocardite infettiva	23	0.07
Infiammazione pelvica	21.5	0.15
Tubercolosi attiva	20-23	0.7-2.1
Cancro	7.5-22.5	0.1-0.74
Arterite temporale	24.8	0.01
Sangue Occulto Feci Cancro colon	8	0.1
Polipi colon	8	0.46
Pletismografia ad impedenza Trombosi venosa profonda	9	0.11
Pielografia intravenosa Stenosi arteria renale	5.4	0.29
Cancro vescica	4.9	0.32
Cancro renale	30	0.1
Ematuria	4	0.1
Pancreatite acuta Amilasi (2 limite superiore intervallo riferimento)	47.5	0.05
Tripsinogeno	5.7	0.04
Lipasi	87	0.13
Embolia polmonare Difetto scintigrafia	1.04	0.08
Difetto scintigrafia segmentata	3.8	0.3
Esame fuzione respiratoria FEV < 80% in giovane asintomatico	1.4	0.98
FEV < 80% in adulto paucisintomatico	2.7	0.81
Liquido cerebrospinale Glucosio < 0.3 g/L	4.7	0.35
Bande oligoclonali nella sclerosi multipla	5.7	0.16
Teofillina Intossicazione (> 20 mg/L)	6	0.44
Ecografia addominale Pancreatite acuta	8.8	0.13
Pancreatite vs cancro	3.1	0.54
Ascesso pancreas	3.8	0.69
Pancreatite cronica	11.9	0.05
Ascesso pelvico	8.8	0.13
Colecistite cronica	5.5	0.5
Massa	23.8	0.05
Ostruzione ureterale	9.8	0.02
Ecografia tiroide Nodulo solido vs cistico	> 95	< 0.05
Analisi urine Esterasi per piuria	22	0.13
IMA ECG Nuova onda Q o soprasslivellamento ST	11	0.34
Qualunque anomalia	1.3	0.04
IMA CK-MB Determinazioni seriate	50	0.01
IMA Troponina I 9 ore dopo insorgenza dei sintomi	47	0.03
IMA Troponina I 24 ore dopo insorgenza dei sintomi	47	0.03
IMA CK-MB 24 ore dopo insorgenza dei sintomi	18	0.46
IMA troponina I 0.4 µg/L	7.5	
IMA tropinina I 1.5 µg/L	21.8	
IgE al lattice 7.06 KU/L	16	
Ig E al lattice 0.81 KU/L	2	
Colon cancro Sigmoidoscopia flessibile	>65	0.35
Tubo digerente	10	0.52
Colonscopia	>94	< 0.06
Metastasi epatiche ALP elevata	1.9	0.54
ALP molto elevata	5.4	0.51
Gamma GT	1.4	0.56
CEA > 5 mg/L	2.9	0.28
CEA > 10 mg/L	5.4	0.34
Scintigrafia	8	0.22
Ecografia	8	0.22
TAC	9	0.11
Epatite virale AST anomala ma < 200 U/L	0.4	
201-400 U/L	1	
401-600 U/L	7	
601-1000 U/L	20	
>1000 U/L	µ	
LES ANA	5	0.013
Anti-DNA	37	0.28
Arterite temporale VES > 30 mm/h	2.5	0.02
Biopsia arteria temporale	16	0.21
Iperparatiroidismo PTH intatto (> 6.3 pmol/L)	14	0.07
Albumina (> 40 g/L)	5.1	0.23
Cloro (> 102 mmol/L)	2.5	0.4
ALP (< 96 U/L)	2	0.51

Tabella I. Come si calcolano gli strumenti dell'EBLM. I.

CONCENTRAZIONE FERRITINA	ANEMIA SIDEROPENICA		
	Presente	Assente	TOTALE
< 30 µg/L	731 a	270 b	1001 a+b
≥30 µg/L	78 c	1500 d	1578 c+d
TOTALI	809 a+c	1770 b+d	2579 a+b+c+d

- $a/(a+c) = \text{Sen} = 731/809 = 90\%$
- $d/(b+d) = \text{Spe} = 1500/1770 = 85\%$
- $\text{LR+} = \text{Sens}/(1-\text{Spec}) = 90\%/15\% = 6$
- $\text{LR-} = (1-\text{Sens})/\text{Spec} = 10\%/85\% = 0.12$

Tabella II. Come si calcolano gli strumenti dell'EBLM. II.

CONCENTRAZIONE FERRITINA	ANEMIA SIDEROPENICA		
	Presente	Assente	TOTALE
< 30 µg/L	731 a	270 b	1001 a+b
≥30 µg/L	71 c	1500 d	1578 c+d
TOTALI	809 a+c	1770 b+d	2579 a+b+c+d

- $\text{NPV} = d/(c+d) = 1500/1578 = 95\%$
- $\text{PRE} = a+c/a+b+c+d = 809/2579 = 31\%$
- $\text{PREODDS} = \text{PRE}/(1-\text{PRE}) = 31\%/69\% = 0.45$
- $\text{POSTODDS} = \text{PRE} * \text{LR} = 0.45 * 6 = 2.7$
- $\text{POST PROB} = \text{POSTODDS}/(\text{POSTODDS}+1) = 2.7/3.7 = 73\%$

Tabella III. Utilita' dell'impiego di 5 livelli di cut-off per un esame diagnostico.

	FERRITINA (µg/L)	LR
● >>POSITIVO	<6.8	52
● POSITIVO	6.8-15.5	4.8
● NEUTRO	15.5-29	1
● <NEGATIVO	29-43	0.39
● << NEGATIVO	> 43	0.08

Tabella IV. NND relativi a 5 concentrazioni sieriche di ferritina.

FERRITINA (µg/L)	SE	1-SP	LR	NND
● <6.8	.58	.011	52	1.7
● 6.8-15.5	.22	.04	4.8	5.8
● 15.5-29	.10	.10	1.	208
● 29-43	.04	.09	.4	-17
● > 43	.06	.75	.08	-1.4

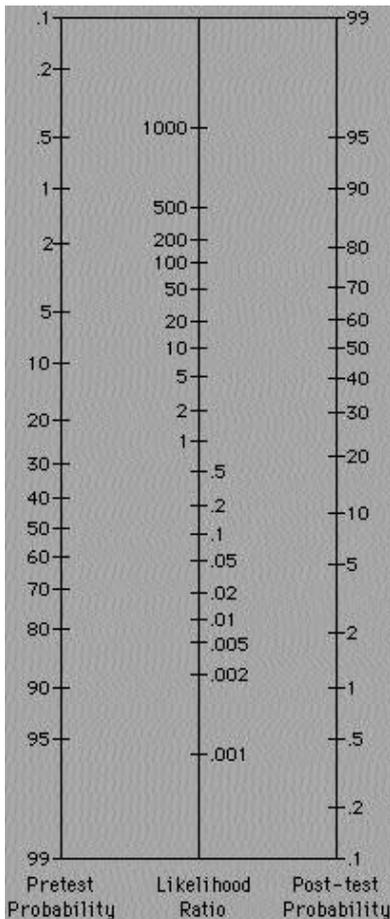


Figura 1. Nomogramma di Fagan.

Figura 2. Soglia dell'esecuzione dell'esame (A) e soglia del trattamento (B).

