Guida all'uso clinico dei biomarcatori in oncologia: metodi di misura e interpretazione

Massimo Gion[1], Chiara Trevisiol[1,2], Silvia Pregno[3], Aline S.C. Fabricio[1,2] a nome dei Gruppi e Comitati di Progetto della Guida*

[1]Centro Regionale Biomarcatori, U.O. Laboratorio Analisi, Dipartimento di Patologia Clinica, Azienda ULSS 12 Veneziana, [2]Fondazione ABO, Venezia

[3]GRADE Working Group, Cochrane Methodology Review Group, Castelvetro di Modena, MO

**L'elenco completo di tutti gli Autori e Collaboratori della Guida è pubblicato su Biochim Clin 2011;35:91-3.*

ABSTRACT

Guidelines for clinical use of biomarkers in oncology: methods and interpretation. Currently available tumour markers are generally determined by immunoassays. The performance of immunoassays is therefore examined, focusing on flaws that may affect the accuracy of measurements. The use of international reference preparations, IQC and EQA programs are also discussed, focusing on their role in reducing and/or managing imprecision and bias. The diagnostic performance is examined, by defining and discussing sensitivity, specificity, ROC curve analysis, positive and negative predictive values, and likelihood ratio. The difference between sensitivity and positive predictive value is examined in detail with reference to different clinical scenarios characterized by different prevalence of malignancy. The interpretative approaches to be used when serial biomarker measurements are available are also discussed, highlighting their potential usefulness as well as their limitations. Finally, the principles of economic evaluation of diagnostic tests are provided and the phases of the evaluation process are described in detail. The deficiency of economic evaluations on the use of circulating tumour markers ascertained in currently available guidelines is highlighted.

ABBREVIAZIONI E ACRONIMI UTILIZZATI

Parametri ematochimici

5-HIAA	Acido idrossiindolacetico
AbTg	Anticorpi anti-tireoglobulina
ACTH	Ormone adrenocorticotropo
AFP	α-fetoproteina
CA 15.3	Antigene carboidratico 15.3
CA 19.9	Antigene carboidratico 19.9
CA 27.29	Antigene carboidratico 27.29
CA 125	Antigene carboidratico 125
CEA	Antigene carcinoembrionario
CgA	Cromogranina A
Ct	Calcitonina
CTC	Cellule tumorali circolanti
DHEA-S	Deidroepiandrosterone solfato
ECD HER2	Dominio extracellulare di HER2
HAMA	Anticorpi umani anti-immunoglobulina murina
hCG e β-hCG	Gonadotropina corionica umana e subunità β
HER2	Recettore umano per il fattore di crescita epidermico 2
IgM	Immunoglobuline M
LDH	Lattato deidrogenasi
MIA	Melanoma inhibitory activity
MUC-1	Mucin 1 (cell surface associated mucin)
NSE	Enolasi neurone specifica
PAP	Fosfatasi acida prostatica
PSA	Antigene prostatico specifico
PSA L/T	Rapporto PSA libero su PSA totale
SCCA	Squamous cell carcinoma-associated antigen
Tg	Tireoglobulina
TSH	Ormone tireostimolante
VEGF	Fattore di crescita vascolare endoteliale
VES	Velocità di eritrosedimentazione
VIP	Peptide intestinale vasoattivo

Organizzazioni, Società, Gruppi cooperativi

ACCP	American College of Chest Physicians
ACN	Australian Cancer Network
ACOG	American College of Obstetricians and Gynecologists
ACS	American Cancer Society
AGREE	Appraisal of Guidelines for Research and Evaluation
AHRQ	Agency for Healthcare Research and Quality
AIOM	Associazione Italiana di Oncologia Medica
ASCO	American Society of Clinical Oncology
ATA	American Thyroid Association

Corrispondenza a: Massimo Gion, Centro Regionale Biomarcatori, U.O. Laboratorio Analisi, Dipartimento di Patologia Clinica, Azienda ULSS 12 Veneziana, Campo Santi Giovanni e Paolo, 30122 Venezia. Tel. 0415294260, Fax 0415294910, E-mail massimo.gion@ulss12.ve.it

EMEA	European Medicines Agency
ESMO	European Society for Medical Oncology
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
NACB	National Academy of Clinical Biochemistry
NCCN	National Comprehensive Cancer Network
NGC	National Guideline Clearinghouse
NICE	National Institute for Health and Clinical Excellence
SIGN	Scottish Intercollegiate Guidelines Network
SNLG	Sistema Nazionale Linea Guida

Miscellanea

CAS	Chemical abstract service
CC	Consensus conference
CRM 457	Human thyroglobulin reference material 457
DT	Tempo di raddoppio
EBG	Linea guida evidence-based
EBM	Evidence-based medicine
ER	Esplorazione rettale
FMTC	Carcinoma midollare tiroideo familiare
FNAC	Fine needle aspiration cytology
HCC	Epatocarcinoma
HTA	Health technology assessment
IRP	International reference preparation
LG	Linea guida
LoE	Level of evidence
MTC	Carcinoma midollare della tiroide
NEBG	Linea guida non evidence-based
NSCLC	Tumore non a piccole cellule del polmone
NSGCT	Tumori germinali non seminomatosi del testicolo
PCR	Reazione a catena della polimerasi
PET	Tomografia a emissione di positroni
Qaly	Quality adjusted life years
RMI	Risk of malignancy index
RT-PCR	Reazione a catena della polimerasi in tempo reale
SCLC	Tumore a piccole cellule del polmone
SI	Standard internazionale
SoR	Strength of recommendation
TAC	Tomografia assiale computerizzata

METODI DI MISURA: IMPORTANZA NELLE DECISIONI CLINICHE

I marcatori vengono misurati principalmente con metodi immunometrici i quali, sfruttando le caratteristiche di alta affinità di un anticorpo specifico contro l'antigene cercato, garantiscono una sensibilità analitica adeguata al livello di concentrazione dei marcatori nei liquidi biologici (10^{-9}-10^{-10} g). Tali metodi presentano alcune caratteristiche che devono essere conosciute per dare un adeguato valore clinico al dato di laboratorio.

Esattezza del metodo: i metodi immunometrici misurano il valore "vero" del marcatore?

Prima di tutto è necessario precisare un aspetto relativo alla terminologia. Nell'uso comune per definire la capacità di misurare il valore "vero" si utilizza il termine "accuratezza". In realtà il termine corretto è quello di "esattezza" ("trueness"). L'accuratezza rappresenta lo scostamento di un'unica determinazione rispetto al valore "vero" e comprende quindi non solo l'esattezza, ma anche l'imprecisione del metodo stesso. L'esattezza può essere determinata confrontando il valore "vero" con la media di una serie di misure, eliminando quindi l'imprecisione intrinseca. Non si tratta solo di un aspetto formale. Infatti, spesso le valutazioni analitiche sono effettuate, o i risultati sono riportati, in maniera tale che può non essere chiaro a cosa ci si riferisce con la misura trovata.

Si deve quindi ricordare che metodi immunometrici sono metodi "comparativi" in quanto l'analita non viene identificato nel campione biologico dal punto di vista chimico, ma viene quantificato per confronto con una curva di calibrazione ottenuta con un antigene standard. A ciò si aggiunga che la stessa identificazione di numerosi marcatori è avvenuta su base immunologica. L'identità tra diversi calibratori e tra calibratore e campione è spesso relativa e questo determina sia una bassa (spesso non determinabile) esattezza intrinseca dei metodi che, conseguentemente, una scarsa confrontabilità tra diversi metodi. La complessità biochimica di molti marcatori fa sì che a tutt'oggi esistano materiali internazionali di riferimento solo per CEA, AFP, hCG, Ct, Tg e PSA. Negli altri casi, le concentrazioni dei marcatori vengono calcolate usando calibratori preparati dall'industria e vengono espresse in unità arbitrarie.

Il CEA ha rappresentato un esempio dell'importanza del materiale di riferimento per la confrontabilità dei risultati in centri diversi. La variabilità dei risultati ottenuti in vari laboratori usando gli standard convenzionali preparati dalle ditte era intorno al 60%, ma si ridusse al 20% quando tutti i risultati vennero normalizzati al materiale internazionale di riferimento mediante un apposito fattore di conversione. Un miglioramento confrontabile si è verificato nel caso della Tg quando si rese disponibile un materiale internazionale e, più recentemente, per il PSA in seguito alla disponibilità del materiale WHO 96/670.

La disponibilità di un materiale internazionale, tuttavia, può non risolvere completamente il problema, come dimostra in modo paradigmatico il caso della hCG. Come è noto, questa molecola è costituita da due subunità, α e β. Nei liquidi biologici, accanto alla molecola nativa e biologicamente attiva sono presenti in percentuali variabili, su base individuale e in relazione alla condizione morbosa, anche altre forme (molecola "nicked" inattiva per clivaggio interno, subunità β, subunità β "nicked", frammento "core" della subunità β, subunità α e subunità α "pesante" per aggiunta di catene carboidratiche). Gli anticorpi utilizzati nei diversi kit commerciali riconoscono in modo differente le varie forme di hCG. È inoltre necessario ricordare che i kit commerciali utilizzano preparazioni standard secondarie ottenute da pool di urine, calibrati in riferimento al materiale internazionale. Tali preparazioni standard secondarie non sono tuttavia interscambiabili, in quanto hanno una variabilità in composizione legata al materiale utilizzato e alle modalità di preparazione e conservazione. Sono state descritte variazioni consistenti (fino a 300%) del valore nominale

dello standard secondario di un dato kit quando misurato con un altro. In queste condizioni il materiale internazionale, pur migliorando di molto la confrontabilità fra i risultati di metodi diversi, non li rende ancora interscambiabili.

Un'analoga situazione è stata recentemente documentata anche nel caso del PSA. Confrontando diversi metodi tutti calibrati con lo standard internazionale WHO si riscontra una significativa riduzione della variabilità intermetodo, ma non si ottengono risultati sovrapponibili in termini di classificazione come positivo/negativo nei campioni clinici. Quindi, nemmeno l'adozione di un unico materiale internazionale di riferimento rende i metodi del tutto interscambiabili.

Per queste ragioni, vanno considerati con cautela i confronti quantitativi fra referti ottenuti in diversi laboratori o nello stesso laboratorio qualora venga cambiato il prodotto commerciale per il dosaggio di un dato marcatore.

Precisione del metodo: il dosaggio di un marcatore è ripetibile?

Il risultato di un marcatore, come di un qualsiasi altro analita, espresso abitualmente come unico valore numerico, corrisponde in realtà ad un intervallo di concentrazione i cui estremi sono determinati dalla variabilità del metodo. Il basso livello di concentrazione dei marcatori circolanti e le caratteristiche intrinseche dei metodi immunometrici fanno sì che la ripetibilità della determinazione analitica sia in linea di massima peggiore di quella riscontrabile in chimica clinica. Orientativamente, uno stesso campione di siero, se misurato più volte, può presentare variazioni di CV intorno al 3-5% nello stesso dosaggio e al 5-10% in dosaggi diversi. L'entità dell'imprecisione è comunque molto variabile a seconda del tipo di marcatore e del metodo usato.

Alla variabilità analitica si sovrappone poi la variabilità biologica, risultante dalle fluttuazioni spontanee, indipendenti cioè da situazioni patologiche, dei livelli circolanti di una data molecola.

Ogni laboratorio dovrebbe stabilire a priori in maniera oggettiva l'accettabilità dell'imprecisione di un metodo. Purtroppo, nel caso dei marcatori, la modalità più comunemente utilizzata a questo scopo (i limiti ricavati tramite i dati di variabilità biologica) sono solo parzialmente applicabili, dato che lo stesso concetto di variabilità biologica è difficilmente applicabile ad analiti di questo tipo. In alternativa possono essere utilizzate le informazioni relative allo stato dell'arte delle prestazioni analitiche, ricavabili dai programmi di VEQ. Dovrebbero in ogni modo essere disponibili notizie idonee sia ad identificare le caratteristiche del metodo utilizzato, sia a valutare la variabilità della misura ad un dato livello di concentrazione.

Le problematiche appena discusse giustificano il fatto che i metodi immunometrici presentino un rischio intrinseco di errore analitico più elevato nei confronti di altri metodi di chimica clinica, per ciò che riguarda sia l'errore sistematico (inesattezza) che l'errore casuale (controllato dalla stima della imprecisione). Questi errori analitici sono monitorati da procedure di controllo della qualità e sono nella maggior parte dei casi identificabili.

Fase preanalitica

Come tutte le indagini di laboratorio, anche il dosaggio dei marcatori è soggetto a errori preanalitici. Gli errori preanalitici, prodotti nelle fasi di raccolta, preparazione e trasporto del campione, possono essere sospettati e a volte riconosciuti valutando la plausibilità clinica del risultato. È pertanto opportuno che il laboratorista sia adeguatamente informato sulle condizioni del paziente al momento del prelievo (stato della malattia, terapia in corso o precedente, ittero, iperlipemia, patologia benigna concomitante, ecc.). Il ruolo del laboratorio è critico nella ottimizzazione della fase preanalitica. È oggi riconosciuto che fra le competenze del laboratorio devono essere anche incluse:

1. la messa a punto di procedure ottimizzate per la preparazione del paziente, per la raccolta, il trasporto e la conservazione del campione;
2. un'attività continua di formazione rivolta a chi ha la responsabilità operativa della fase preanalitica;
3. il monitoraggio permanente della corretta esecuzione di tutte le procedure relative.

Quindi, non differentemente da altre procedure diagnostiche, anche nel caso dei marcatori il miglioramento continuo della qualità deve spostare il suo obiettivo dal risultato (fase analitica) al processo (fasi pre-analitica, analitica e post-analitica).

INTERPRETAZIONE DEL DATO DI LABORATORIO

Nell'utilizzo clinico dei marcatori il clinico può trovarsi di fronte a due possibilità:

a) disponibilità di un unico referto;
b) disponibilità di più referti successivi per lo stesso paziente.

Nel primo caso, la valutazione del marcatore viene fatta secondo un criterio dicotomico; il marcatore viene cioè classificato come positivo o negativo nei confronti di un dato valore soglia. Nel secondo caso, la valutazione è di tipo dinamico; infatti vengono considerate le variazioni del marcatore nel tempo nei confronti del o dei valori precedenti. Questo tipo di valutazione viene oggi considerato come il modo biologicamente più corretto di valutare i marcatori in tutte le situazioni in cui è possibile disporre di dosaggi seriati (variazione post-operatoria, "follow-up", monitoraggio della terapia). Il criterio dicotomico (positivo o negativo nei confronti di un valore soglia) trova un razionale impiego nel bilancio iniziale prima del trattamento del tumore primitivo e, quando indicato, in fase diagnostica, in quanto in tali situazioni cliniche non sono generalmente disponibili più determinazioni nel tempo.

Valutazione di un unico referto: il problema del valore soglia

Quando si disponga di un unico risultato, per interpretare l'informazione fornita dal marcatore è necessario usare un valore soglia positivo/negativo. A tutt'oggi non esiste tuttavia un marcatore assolutamente specifico per tumore, cioè rilevabile se c'è il tumore e non rilevabile negli altri casi. La concentrazione di un marcatore nel sangue è dovuta infatti in parte alla produzione e al rilascio di molecole simili o uguali da parte di tessuti non neoplastici. Inoltre, come si è detto precedentemente[1], numerose condizioni diverse dal tumore possono aumentare anche considerevolmente le concentrazioni ematiche di alcuni marcatori. Infine, le concentrazioni circolanti del marcatore sono condizionate dal volume di distribuzione, dal metabolismo e dall'entità della "clearance".

La definizione del livello di un marcatore come positivo o negativo usando un valore soglia (cut-off, usando il termine anglosassone entrato nell'uso comune) è fatta in riferimento alla distribuzione del marcatore in soggetti apparentemente sani (ad es., media +2 oppure +3 DS nel caso di distribuzione gaussiana, 95° oppure 99° percentile nel caso di distribuzioni non gaussiane). Un caso particolare è rappresentato dalla stratificazione del livello soglia in base alle caratteristiche demografiche della popolazione nella quale si ritiene di applicare il marcatore. L'esempio più noto è rappresentato dai livelli soglia stratificati per età che sono stati proposti per il PSA. Il PSA presenta in genere concentrazioni più elevate nei soggetti più anziani, anche in assenza di patologia prostatica. Pertanto, usando livelli soglia più bassi nel giovane e più alti nell'anziano si privilegia la sensibilità nel primo caso, la specificità nel secondo. Questo tipo di approccio richiede però cautela. Nel caso del PSA, ad esempio, l'evidente associazione con l'età è in realtà secondaria all'incremento volumetrico della ghiandola che si verifica con l'invecchiamento. Infatti, quando i soggetti vengono stratificati per volume prostatico, il rapporto PSA ed età non si evidenzia più. Pertanto, in assenza di indicazioni precise sul volume della prostata è sconsigliabile utilizzare valori soglia differenziati per età in condizioni routinarie.

Impiego clinico: sensibilità, specificità, valore predittivo

L'uso di un valore soglia +/- consente di suddividere la popolazione in esame nei seguenti quattro gruppi.

Veri positivi:	soggetti con tumore e marcatore positivo
Veri negativi:	soggetti sani con marcatore negativo
Falsi negativi:	soggetti con tumore e marcatore negativo
Falsi positivi:	soggetti sani con marcatore positivo

L'identificazione di questi quattro gruppi permette di calcolare le prestazioni diagnostiche di un marcatore e precisamente, sensibilità, specificità, valore predittivo negativo (VPN) e positivo (VPP).

La sensibilità esprime la capacità di un marcatore di riconoscere la presenza di malattia in pazienti effettivamente malati (percentuale di risultati veri positivi sul totale dei soggetti malati).

$$\text{Sensibilità} = \frac{\text{veri positivi}}{\text{veri positivi + falsi negativi (totale malati)}}$$

La specificità esprime la capacità di un marcatore di riconoscere solo una data neoplasia e di restare negativo nei soggetti senza la malattia (percentuale di risultati veri negativi sul totale dei soggetti non malati).

$$\text{Specificità} = \frac{\text{veri negativi}}{\text{veri negativi + falsi positivi (totale soggetti non malati)}}$$

L'efficacia diagnostica esprime in modo complessivo l'efficacia del marcatore e rappresenta la quota complessiva di risultati corretti.

$$\text{Efficacia diagnostica} = \frac{\text{veri positivi + veri negativi (risposte corrette)}}{\text{veri positivi + veri negativi + falsi positivi + falsi negativi (totale soggetti indagati)}}$$

VPP e VPN esprimono l'attendibilità di un risultato rispettivamente positivo o negativo; in altre parole, indicano la percentuale di risultati veri positivi sul totale dei risultati positivi e dei risultati veri negativi sul totale dei risultati negativi.

$$\text{VPP} = \frac{\text{veri positivi}}{\text{veri positivi + falsi positivi (totale risultati positivi)}}$$

$$\text{VPN} = \frac{\text{veri negativi}}{\text{veri negativi + falsi negativi (totale risultati negativi)}}$$

Sensibilità e specificità da un lato, VPP e VPN da un altro rappresentano due approcci molto diversi per valutare un esame diagnostico. Sensibilità e specificità sono infatti valutazioni mirate all'efficienza del marcatore in condizioni cliniche note. Il valore predittivo esprime invece la probabilità che un valore positivo o negativo abbia un'associazione con la condizione indagata. La

[1]Gion M, Trevisiol C, Pregno S, et al. Guida all'uso clinico dei biomarcatori in oncologia: premesse e generalità. Biochim Clin 2011;35:97-106.

Tabella 1 presenta un esempio che illustra come un marcatore con date caratteristiche di sensibilità e specificità possa avere un valore predittivo profondamente diverso in condizioni di differente prevalenza della forma morbosa ricercata.

Il confronto tra diversi marcatori: la curva ROC

Un approccio molto utilizzato per valutare le prestazioni di un marcatore e, soprattutto, per confrontare in modo oggettivo marcatori diversi è l'uso delle curve ROC. La curva ROC è un grafico che si ottiene ponendo, per ogni valore disponibile dell'esame, la percentuale di veri positivi (sensibilità) in ordinata e la percentuale di falsi positivi (1 - specificità) in ascissa. L'ampiezza dell'area sottesa dalla curva è proporzionale all'efficacia diagnostica del marcatore.

La potenziale utilità clinica di un marcatore: il rapporto di verosimiglianza

Un altro parametro importante per valutare l'efficacia diagnostica di un marcatore è il rapporto di verosimiglianza ("likelihood ratio"), espresso dal rapporto fra i risultati veri positivi e falsi positivi. Tale parametro esprime la validità con cui l'esame identifica correttamente la malattia in oggetto.

$$\text{Rapporto di verosimiglianza} = \frac{\text{veri positivi}}{\text{falsi positivi}}$$

Il rapporto di verosimiglianza indica di quanto il risultato del marcatore modifica la probabilità che una persona con date caratteristiche cliniche abbia la malattia. In altre parole, indica la capacità di orientare la diagnosi dopo il test (probabilità post-test) nei confronti di quella che si avrebbe senza eseguire il test (probabilità pre-test). Maggiore è il valore del rapporto di verosimiglianza, maggiore è la probabilità che l'esame indichi la presenza di malattia; se il rapporto di verosimiglianza è 1 la probabilità pre-test non viene modificata. Sono stati riportati in letteratura alcuni limiti arbitrariamente scelti su base empirica per definire il potenziale impatto del marcatore nel processo decisionale basato sul rapporto di verosimiglianza (Tabella 2).

Note per la pratica clinica

Il giudizio di positività su una singola determinazione di un marcatore va considerato sempre come una classificazione probabilistica che deve essere valutata con prudenza ed eventualmente confermata prima di ulteriori decisioni cliniche.
Il valore soglia dipende in parte dal metodo di dosaggio impiegato in un dato laboratorio e dalle caratteristiche del gruppo di controllo e dei pazienti valutati. Sarebbe pertanto auspicabile che, almeno nei laboratori che hanno un ruolo di riferimento, il valore soglia venisse calcolato sulla base delle caratteristiche della popolazione sana di afferenza e delle patologie di più frequente riscontro.

Tabella 1
Sensibilità e specificità vs. valore predittivo

Scenario 1. Diagnosi di ricaduta dopo trattamento del tumore primitivo

Sensibilità: 85%; specificità: 90%

Soggetti esaminati: 100; probabilità di ricaduta: 40/100; malati attesi nel campione: 40

Veri positivi attesi: 34; veri negativi attesi: 54; falsi positivi attesi: 6; falsi negativi attesi: 6

$$\text{VPP} = \frac{\text{VP}}{\text{VP + FP}} = \frac{34}{34 + 6} = 0{,}85$$

$$\text{VPN} = \frac{\text{VN}}{\text{VN + FN}} = \frac{54}{54 + 6} = 0{,}90$$

Circa 8 valori positivi su 10 sono associati alla presenza di malattia.

Scenario 2. Screening di tumore nella popolazione generale

Sensibilità: 85%; specificità: 90%

Soggetti esaminati: 100.000; prevalenza del tumore: 40/100.000; malati attesi nel campione: 40

Veri positivi attesi: 34; veri negativi attesi: 89964; falsi positivi attesi: 9996; falsi negativi attesi: 6

$$\text{VPP} = \frac{\text{VP}}{\text{VP + FP}} = \frac{34}{34 + 9996} = 0{,}003$$

$$\text{VPN} = \frac{\text{VN}}{\text{VN + FN}} = \frac{89964}{89964 + 6} = 0{,}999$$

Solo circa 1 ogni 300 valori positivi del marcatore è associato alla presenza di malattia.

VPP, valore predittivo positivo; VP, veri positivi; FP, falsi positivi; VPN, valore predittivo negativo; VN, veri negativi; FN, falsi negativi.

Tabella 2
Criteri empirici basati sul rapporto di verosimiglianza per valutare la potenziale utilità clinica di un marcatore

Rapporto di verosimiglianza	Effetto del marcatore sulla probabilità di malattia "post-test"
>10	Cambiamento consistente; impatto clinico importante
5-10	Cambiamento moderato; impatto clinico probabile
2-5	Cambiamento modesto; impatto clinico possibile
1-2	Cambiamento trascurabile; impatto clinico irrilevante

Valutazione di più referti successivi: il problema delle variazioni del valore nel tempo

Quando sono disponibili più referti successivi per uno stesso paziente, la valutazione dovrebbe basarsi sulle variazioni nel tempo delle concentrazioni del marcatore nel singolo soggetto, superando così, almeno in parte, le problematiche connesse con il valore soglia. Questo criterio può essere applicato in numerose situazioni cliniche, quali la valutazione di eventuale malattia residua dopo chirurgia, l'evidenziazione precoce della ricaduta della malattia e il monitoraggio dell'efficacia della terapia nella malattia avanzata.

La valutazione della variazione fra prelievi seriati può essere basata su criteri empirici o su criteri matematici. I criteri empirici sono largamente utilizzati. Per definizione, si basano su scelte arbitrarie legate ad esperienze appunto "empiriche", maturate in scenari particolari (ad es., patologia, metodica analitica, modalità di monitoraggio, ecc.). Caratteristicamente, i criteri empirici sono difficilmente implementabili in realtà diverse da quelle dove sono stati messi a punto. Nel monitoraggio del carcinoma ovarico e prostatico sono stati standardizzati dei criteri empirici, definendo i limiti di tolleranza di numerose variabili (numero di prelievi da eseguire, distanza di tempo fra i diversi prelievi, livello di partenza consentito per il marcatore, percentuale di variazione considerata significativa, variabilità analitica accettata per il metodo usato). Tali criteri sono stati poi validati in casistiche differenti da quelle utilizzate per metterli a punto e possono essere quindi presi in considerazione per l'applicazione clinica.

L'analisi matematica delle variazioni fra campioni seriati è certamente più promettente in vista di un'applicazione clinica riproducibile su larga scala. La Tabella 3 riassume alcuni degli approcci possibili per tale valutazione. Una disamina in dettaglio delle prospettive e delle problematiche di tali approcci esula dallo scopo di questa Guida. In sintesi si può affermare che a tutt'oggi non sono disponibili algoritmi standardizzati e ottimizzati per un'applicazione routinaria nella maggior parte delle neoplasie epidemiologicamente rilevanti.

Tabella 3
Valutazione di referti seriati: alcuni esempi di approccio matematico

- Analisi della regressione dose/tempo
- Valutazione della differenza critica
- Tempo di dimezzamento (emivita, t½)
- Tempo di raddoppio
- Algoritmi complessi

PRINCIPI DI VALUTAZIONE ECONOMICA DELLA DIAGNOSTICA DI LABORATORIO

Negli ultimi cinquant'anni i nostri sistemi sanitari sono notevolmente cambiati nella forma organizzativa e per le innovazioni introdotte a favore dei pazienti con un contributo al miglioramento della morbilità e della mortalità. A partire dagli anni '90 è sempre più condiviso il concetto di valutare con attenzione le risorse impiegate per individuarne l'utilizzo più efficace.

Se nei mercati concorrenziali il valore di un bene è dato dal suo prezzo, che rappresenta un indicatore del valore assegnato dal cliente, questo non si realizza in sanità perché non sono presenti le condizioni di un mercato concorrenziale. Viceversa, è stata proposta la valutazione economica come strumento per ricercare la migliore allocazione delle risorse, confrontando i costi e l'efficacia di due o più alternative rendendo esplicito il valore delle risorse aggiuntive investite in un determinato trattamento sanitario. Va subito detto che nonostante la grande enfasi posta da una parte degli economisti sanitari sulla valutazione economica, la ricaduta reale sembra essere limitata.

Fasi di una valutazione economica

Per disegnare e leggere criticamente una valutazione economica sono previste diverse fasi.

Definizione del quesito dello studio e scelta dell'alternativa di confronto

Lo studio può rispondere a vari quesiti clinici, di diagnosi e trattamento di una patologia, della gestione organizzativa. La chiarezza del quesito è la condizione da cui partire per definire la relazione fra un'esposizione (a un trattamento, a un esame diagnostico) e l'esito di interesse. Il passaggio successivo è la definizione dell'alternativa a confronto con la tecnologia indagata.

Valutazione dei dati di efficacia

Partendo dall'affermazione di Sackett che se non si dispongono di solide stime di efficacia "ci si deve domandare se non si faccia meglio a fare qualcos'altro", per l'analisi dell'efficacia di un trattamento in una valutazione economica dovrà essere usato almeno uno studio clinico controllato, meglio se randomizzato e che rifletta la pratica clinica. Quando sono disponibili più studi è auspicabile l'utilizzo di revisioni sistematiche.

Una volta individuati gli studi clinici e i dati di efficacia si deve decidere quali indicatori usare per la loro sintesi.

Si possono utilizzare “end point” cosiddetti forti o surrogati. Con i primi si intendono esiti clinici importanti quali il numero di anni di vita guadagnati, le vite salvate, gli eventi evitati; con i secondi si intendono parametri intermedi quali la colesterolemia, la riduzione della pressione arteriosa o la densità minerale ossea. Le valutazioni economiche di buona qualità dovrebbero evitare di utilizzare “end point” surrogati che, così come negli studi clinici randomizzati e controllati, possono sovra o sottostimare la reale efficacia degli interventi.

Una revisione della letteratura sulle valutazioni economiche legate alla diagnostica ha mostrato come la analisi costi-efficacia fosse la più utilizzata. Questo perché l’efficacia dell’esame è misurata attraverso degli indicatori intermedi e quindi utilizzati nel rapporto incrementale.

Scelta della tecnica di valutazione economica

Dopo aver definito le alternative e l’efficacia di un trattamento, è possibile decidere il tipo di tecnica economica da utilizzare. Come mostrato in Appendice, se l’efficacia è misurabile solo con “end point” clinici la scelta ricadrà sull’analisi costi-efficacia o, in caso di pari efficacia, sull’analisi di minimizzazione dei costi. Solo nel caso in cui vi siano stime sulla qualità di vita per i pazienti si può ricorrere all’analisi costi-utilità. Al fine di poter valutare dal punto di vista economico anche programmi complessi come, per esempio, l’implementazione di linee guida, l’uso di un unico indicatore di risultato come il costo per evento evitato potrebbe risultare limitativo. In questo caso è stato proposto l’uso del “balance sheet approach” che consiste nell’elencare tutti i costi e tutti i benefici delle alternative considerate. Tale approccio è quello adottato anche nella metodologia GRADE per la presentazione dei dati economici legati alle raccomandazioni cliniche.

Scelta del punto di vista e delle risorse considerate

La metodologia dell’analisi ha implicazioni sul tipo di informazioni da raccogliere per la valutazione desiderata. Si potrà cioè optare per il punto di vista del terzo pagante (per esempio il Servizio Sanitario Nazionale o un’assicurazione) prendendo in considerazione solo i costi diretti, sia sanitari (ricoveri, farmaci, visite) che non sanitari (assistenza domiciliare al paziente). Oppure per il punto di vista della società, che include anche i costi indiretti, ossia le perdite di produttività del paziente e dei suoi familiari dovuti alla patologia.

Infine, a causa della preferenza di posticipare eventi negativi e anticipare quelli positivi è necessario tener conto della diversa distribuzione temporale degli eventi (risorse o benefici). Quindi, nel caso in cui la valutazione economica presenti un orizzonte temporale superiore all’anno dovranno essere attualizzati sia i costi sia i benefici.

Calcolo dei risultati e dell’analisi di sensibilità

Una volta raccolti i dati di efficacia e di costo è possibile calcolare i risultati della valutazione economica. Nel confronto fra un trattamento nuovo (A) rispetto a uno vecchio (B) ci si può trovare di fronte, schematicamente, a quattro diversi scenari:

1. A è meno efficace di B e più costoso: A si dice dominato e non viene adottato;
2. A è più efficace di B e meno costoso: A è detto dominante e viene sempre adottato;
3. A è meno efficace di B e meno costoso: è un caso particolare in cui, per esempio, un’alternativa di poco meno efficace potrebbe essere scelta se molto meno costosa. Qui si pone un problema etico se accettare alternative meno efficaci;
4. A è più efficace e più costoso. Questo è il caso più frequente, ad esempio per le nuove terapie, e si dovrà calcolare il rapporto incrementale, ossia un rapporto fra le differenze di costo (costi incrementali) e le differenze di efficacia (benefici incrementali). Come già detto, il risultato è espresso, ad esempio, come costo per anno di vita salvato.

Se si utilizza l’analisi costi-efficacia si possono confrontare i risultati solo con quelli di altri lavori che riportino lo stesso parametro per esprimere l’efficacia (per esempio il costo per anno di vita salvato). Nell’analisi di minimizzazione dei costi il confronto sarà soltanto sulle risorse considerate nello studio. Il confronto fra patologie diverse potrebbe in teoria essere svolto con le cosiddette “league table” che confrontano risultati di diversi interventi attraverso il costo per “quality adjusted life years” (QALY) guadagnato. Ogni decisore sanitario vorrebbe poter disporre di una classifica degli interventi di questo tipo in modo da poter stabilire, fissato un certo budget, quali interventi finanziare o meno.

Ma quando un trattamento viene definito costo-efficace? Vi sono diversi valori di riferimento, ma quello convenzionalmente più usato è pari a 50.000 dollari per anno di vita salvato o di 30.000 sterline per QALYs secondo il NICE inglese: entro tali valori i trattamenti sono definiti accettabili. Se da un lato i sistemi sanitari pubblici dovranno sempre più tener conto dei risultati delle valutazioni economiche per rendere sostenibili bilanci e domanda di salute, dall’altro permangono dubbi sulla solidità di queste valutazioni, sia per la bassa qualità metodologica, sia per la presenza di forti conflitti di interesse.

Le valutazioni economiche negli studi diagnostici

Un aspetto economico rilevante riguarda le conseguenze economiche degli errori diagnostici. Ad esempio, diagnosi errate positive o negative possono portare, oltre a problemi di tipo clinico e di qualità della vita del paziente, anche un importante utilizzo inappropriato di risorse.

In alcuni casi le valutazioni economiche diagnostiche utilizzano come misura di efficacia il numero dei casi veri positivi. Tuttavia una valutazione economica che utilizza solo questo esito non definisce i costi e gli esiti clinici dei pazienti diagnosticati in modo erroneo. Anche i falsi negativi dovrebbero essere quindi considerati nell’analisi, potendo portare a importanti conseguenze in termini di efficacia, sicurezza e costi. Ad esempio, si pensi ai problemi nella gestione di pazienti con dolore toracico,

diagnosticati erroneamente come infarto acuto del miocardio.

Come già evidenziato, vi sono diversi studi di valutazione economica su esami diagnostici che purtroppo presentano una qualità metodologica bassa; se si vuole che questo strumento si accompagni all'espansione di utilizzo degli esami diagnostici, questo problema deve essere superato. Diversi metodi sono stati proposti in letteratura per tener conto delle specificità diagnostiche.

Nella revisione delle linee guida esistenti sui marcatori oncologici circolanti oggetto di questa Guida si è ricercata nei documenti analizzati anche la presenza di indicazioni di tipo economico nelle raccomandazioni prodotte. Poichè l'aspetto economico è ormai divenuto un imperativo per tutti i sistemi sanitari, pubblici e privati, l'attesa era di trovare nelle raccomandazioni prodotte per i clinici indicazioni sul consumo di risorse. Dall'analisi dei documenti non è emersa invece alcuna valutazione economica utilizzata per formulare raccomandazioni legate ai marcatori oncologici circolanti. Questo potrebbe significare che non sono presenti analisi economiche sul tema specifico o che lo strumento della valutazione economica risulta di fatto inadeguato allo scopo, perché chi produce raccomandazioni non lo conosce o non lo utilizza.

APPENDICE

LE TECNICHE DI VALUTAZIONE ECONOMICA

Condizione necessaria per avere una valutazione economica in senso stretto è aver considerato almeno due alternative e disporre di un confronto delle stesse in termini di costi (espresse al numeratore del rapporto) e di efficacia (al denominatore).

- Nelle analisi costi-efficacia, l'efficacia degli interventi è espressa in unità naturali (ad es., anni di vita, eventi evitati) e i risultati sono riportati come costo per anno di vita salvato. Nella analisi di minimizzazione dei costi consideriamo la pari efficacia fra le due alternative e confronteremo solo i costi delle stesse.
- Nelle analisi costi-utilità, gli anni di vita vengono pesati per la qualità di vita dei pazienti con la creazione di un nuovo indicatore, il "quality adjusted life years" (QALY) e i risultati sono espressi come costo per QALY.
- Nelle analisi costi-benefici, l'efficacia viene espressa in termini monetari. Questa tecnica permette di fare confronti anche al di fuori della sanità (viene per esempio usata anche nel campo dei trasporti) e tuttavia in sanità si presta a forti critiche a causa della monetizzazione dei benefici.

Per un approfondimento sulle valutazioni economiche:

- Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes, 3rd ed. Oxford: Oxford University Press, 2004.
- Laking G, Lord J, Fischer A. The economics of diagnosis. Health Econ 2006;15:1109-20.
- Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. Med Decis Making 1988;8:279-89.
- Scott MA, Raftery JP. Evidence-based laboratory medicine: Health economics, reimbursement, and the value proposition. In: Price CP, Christenson RH, eds. Evidence-based laboratory medicine: Principles, practice, and outcomes, 2nd ed. Washington: AACC Press, 2007.
- Severens JL, van der Wilt GJ. Economic evaluation of diagnostic tests: a review of published studies. Int J Technol Assess Health Care 1999;15:480-96.