Definizione dei traguardi di prestazione analitica sulla base dell'"outcome" clinico

George G. Klee
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, USA

Traduzione a cura di Ferruccio Ceriotti

ABSTRACT

Establishment of outcome-related analytic performance goals. Accrediting organizations require laboratories to establish analytic performance criteria that ensure their tests provide results of the high quality required for patient care. However, the procedures for instituting performance criteria that are directly linked to the needs of medical practice are not well established, and therefore alternative strategies often are used to create and implement surrogate performance standards. I reviewed six approaches for establishing outcome-related analytic performance goals: (a) limits defined by regulations and EQAS, (b) limits based on biologic variation, (c) limits based on surveys of clinicians about their needs, (d) limits based on effects on guideline driven medical decisions, (e) limits based on analysis of patterns for ordering follow-up clinical tests, and (f) limits based on formal medical decision models. Performance criteria were tabulated for 12 common chemistry analytes and four routine hematology tests. There is no consensus currently about the preferred methods for establishing medically necessary analytic performance limits. The various methods reviewed give considerably different performance limits. The analytic performance limits claimed by a laboratory should correspond to those limits that can be reliably maintained based on validated quality control monitoring systems. These limits generally are larger than the observed CVs and bias parameters collected for assay validation. There is a major need for increased communication among laboratorians and clinicians on this topic, especially when the analytic performance limits that can be consistently maintained by a laboratory are inconsistent with the expectations of health care providers.

INTRODUZIONE

La definizione di limiti oggettivi di prestazione rappresenta un requisito essenziale per la valutazione e il controllo efficace dei sistemi di laboratorio. Le regole del Clinical Laboratory Improvement Amendments (CLIA-88) richiedono che il direttore del laboratorio "assicuri che le metodologie scelte abbiano la capacità di fornire la qualità dei risultati necessaria per la cura dei pazienti e stabilire e mantenere livelli accettabili di prestazioni per ciascun sistema analitico" (1). Criteri per le prestazioni analitiche chiaramente definiti sono anche necessari per lo sviluppo di robusti sistemi di controllo di qualità (CQ) perché sono indispensabili per calcolare la frequenza di falsi negativi e falsi positivi (2). Tuttavia, le procedure per stabilire le specifiche di qualità analitica necessarie perché un esame possa raggiungere l'utilità clinica richiesta non sono ben definite.

Fraser ha pubblicato un approccio gerarchico alla classificazione delle strategie per definire specifiche di qualità in Medicina di Laboratorio (3) e questa lista gerarchica di strategie è stata approvata da una conferenza internazionale, "Strategies to set global quality specifications in Laboratory Medicine", e denominata "The Stockholm Conference Hierarchy" (4, 5). Le prime 4 strategie di Fraser erano: (a) valutazione dell'effetto sulle decisioni cliniche, (b) raccomandazioni professionali da parte di gruppi di esperti nazionali e internazionali, singoli esperti o gruppi istituzionali, (c) specifiche di legge o da parte di organizzatori di valutazioni esterne e (d) dati pubblicati relativi allo stato dell'arte.

Fraser notava che, benché le specifiche basate su come la qualità influenzi le decisioni cliniche fossero al top della gerarchia, questo approccio è difficile da applicare perché pochi esami sono usati in una singola e ben definita situazione clinica, con strategie mediche ampiamente accettate e standardizzate, che siano direttamente collegate al risultato dell'esame. Notava inoltre che l'analisi degli effetti della prestazione dell'esame sulle decisioni cliniche è strettamente dipendente dalle assunzioni

**Questo articolo è stato tradotto con il permesso dell'American Association for Clinical Chemistry (AACC). AACC non è responsabile della correttezza della traduzione. Le opinioni presentate sono esclusivamente quelle degli Autori e non necessariamente quelle dell'AACC o di Clinical Chemistry. Tradotto da Clin Chem 2010;56:714-22 su permesso dell'Editore.*
Copyright originale © 2010 American Association for Clinical Chemistry, Inc. In caso di citazione dell'articolo, riferirsi alla pubblicazione originale in Clinical Chemistry.

fatte sul modo in cui i risultati sono utilizzati dai clinici. Per questi motivi, la prima strategia è usata di rado.

Una recente analisi del "College of American Pathologists" sulla soddisfazione dei medici rispetto ai servizi forniti dai laboratori clinici ha dimostrato come la categoria scelta con maggior frequenza come la più importante era "qualità/affidabilità dei risultati" (6). È interessante notare come questa categoria, che riguarda l'esattezza e la precisione dei risultati, aveva uno dei livelli di soddisfazione più elevato, mentre altre categorie, come i tempi di risposta ("turnaround time"), l'adeguatezza del pannello di esami offerti e i servizi di trasporto dei campioni, avevano livelli di soddisfazione inferiori. Forse è più semplice per i clinici valutare queste ultime categorie, che hanno livelli di prestazioni attese ben definiti, piuttosto che valutare la qualità/affidabilità dei risultati, che invece non hanno livelli di prestazioni attese altrettanto ben definiti. In assenza di limiti oggettivi per le prestazioni analitiche che definiscano "la qualità di risultati necessaria per la cura del paziente" è anche difficile per i laboratoristi assicurare che i loro sistemi di CQ siano in grado di individuare in maniera attendibile le prestazioni inadeguate. Per altro, Plebani ha notato in modo convincente che, anche quando un laboratorio abbia specifiche di qualità analitica ben definite, questo può avere scarso effetto sulle decisioni cliniche, a meno che il clinico non sia chiaramente informato relativamente a queste specifiche (7).

Questo lavoro passa in rassegna alcuni dei metodi per definire le specifiche di qualità che sono stati pubblicati e discute pregi e difetti di questi approcci e le loro implicazioni nel miglioramento della cura. Questa rassegna mostra che non c'è un'unica strategia per definire specifiche di prestazioni analitiche sulla base dell'"outcome" clinico e possono essere necessari più approcci tra loro correlati. Nel 1997, Werner notava che il rapporto fra traguardi analitici e strategie cliniche è reciproco, tale per cui l'esito clinico può essere ottimizzato sia adattando le strategie cliniche alle prestazioni analitiche esistenti oppure ottimizzando le prestazioni analitiche in base alle esigenze cliniche (8). Sono stati proposti 6 approcci che possono servire come base per stabilire i limiti delle prestazioni analitiche: regolamentazioni e valutazioni esterne, variabilità biologica, indagini sulle necessità cliniche, decisioni basate su linee guida, analisi delle modalità cliniche di richiesta e modelli decisionali formalizzati.

APPROCCIO 1: LIMITI DI PRESTAZIONI DEFINITI DA DISPOSIZIONI REGOLATORIE E SPECIFICHE DI VALUTAZIONI ESTERNE

Agenzie di controllo e programmi di VEQ hanno definito limiti per la valutazione delle prestazioni nei confronti tra laboratori. Questi limiti sono basati soprattutto sullo stato dell'arte o sulla variabilità biologica. Essi sono usati come surrogati di prestazione clinica perché limiti ben definiti di prestazioni cliniche non sono generalmente disponibili (9).

Il CLIA-88 ha definito limiti di precisione interlaboratorio per molti esami. Benché le procedure usate per sviluppare questi limiti non siano state fornite in modo esplicito, i limiti sono implicitamente legati allo stato dell'arte della pratica di laboratorio del 1988. Solo pochi laboratori non sono stati in grado di rispettare i limiti posti in essere dal CLIA-88. Alcuni di questi limiti sono piuttosto ampi e si estendono in un ambito più ampio dell'intervallo di riferimento (Tabella 1); ad esempio, i limiti per il calcio plasmatico sono definiti come ±1 mg/dL, anche se differenze di minore ampiezza nei valori di calcemia potrebbero già avere implicazioni cliniche rilevanti.

Il programma di VEQ tedesco (G-EQAS) utilizza il concetto della radice quadrata della media dei quadrati delle differenze ["root mean square deviation" (RMSD)] per definire i limiti delle prestazioni (10). La RMSD è basata sull'eliminazione del 10% dei dati e sul calcolo della percentuale di errore rispetto a un valore bersaglio (10). Alcuni programmi di VEQ, come quello del Centro di Ricerche Biomediche (Castelfranco TV), hanno stabilito limiti di prestazioni derivati dall'errore totale analitico (TE) (11, 12). Questi limiti di TE combinano limiti di imprecisione e inesattezza per fornire una statistica combinata derivante dalla variabilità biologica intra-individuale (CV_I) e inter-individuale (CV_G). Come sarà discusso nella sezione successiva, il traguardo di imprecisione desiderabile è pari a $<0,5\ CV_I$ e quello per il "bias" è $<0,25\ (CV_I^2 + CV_G^2)^{1/2}$. Questi limiti di prestazioni di solito includono un fattore di copertura (k), che indica con che rigore i limiti sono applicati. Un fattore k di 1,65 corrisponde a una copertura pari al 95° percentile, mentre un fattore k di 1,96 corrisponde al 97,5° percentile. In altre parole, con k=1,65 è atteso che il 5% dei risultati possa superare il limite, mentre con k=1,96 solo il 2,5%. Esempi di limiti di prestazioni per CLIA, G-EQAS e TE sono presentati nella Tabella 1. In questa Tabella, i limiti di TE sono calcolati come: $k\ (0,5\ CV_I) + 0,25\ (CV_I^2 + CV_G^2)^{1/2}$, dove k=1,65.

Né CLIA-88, né i sistemi di VEQ hanno limiti per la valutazione dell'esattezza. Entrambi i sistemi usano metodi statistici per assegnare i valori bersaglio, che sono specifici per vari gruppi di metodi. L'assunto implicito è che i clinici compensino queste differenze fra metodi basandosi sugli specifici intervalli di riferimento e su informazioni interpretative fornite dai laboratori, ma non ci sono prove per convalidare questo assunto.

APPROCCIO 2: USO DELLA VARIABILITÀ BIOLOGICA PER FORMULARE LIMITI DI VARIABILITÀ ANALITICA ACCETTABILE

Si può dimostrare statisticamente che la variabilità totale di un risultato è la combinazione dell'imprecisione analitica (espressa come DS o CV) e della variabilità biologica (sempre espressa come DS o CV). La variabilità totale (assumendo l'indipendenza delle due fonti di variabilità) è la radice quadrata della somma dei quadrati delle DS analitica e biologica:

$$DS_{totale} = \sqrt{(DS^2_{analitica} + DS^2_{biologica})}$$

Tabella 1
Limiti di prestazioni analitiche utlizzati nel Clinical Laboratory Improvement Amendments (CLIA), nella VEQ tedesca (G-EQAS) e basati sui dati di variabilità biologica

	CLIA	G-EQAS	Variabilità biologica e limiti derivati[a]				
Analita			CV_I	CV_G	Imprecisione	Inesattezza	Errore totale
S-Bilirubina	0,4 mg/dL (o 20%)	13%	23,8%	39,0%	11,9%	11,4%	31,1%
S-Calcio	1 mg/dL	6%	1,9%	2,8%	0,8%	0,8%	2,4%
S-Colesterolo	10%	7%	5,4%	15,2%	4,0%	4,0%	8,5%
S-Cortisolo	25%	16%	20,9%	45,6%	12,5%	12,5%	29,8%
S-Creatinina	0,3 mg/dL (o 15%)	11,5%	5,3%	14,2%	3,8%	3,8%	8,2%
S-Glucosio	6 mg/dL (o 10%)	11%	5,7%	6,9%	2,9%	2,2%	6,9%
S-Ferro	20%	-	26,5%	23,2%	13,3%	8,8%	30,7%
S-Fosforo	0,3 mg/dL (o 10,7%)	9%	8,5%	9,4%	4,3%	3,2%	10,2%
S-Potassio	0,5 mmol/L	4,5%	4,8%	5,6%	2,4%	1,8%	5,8%
S-Sodio	4 mmol/L	3%	0,7%	1,0%	0,4%	0,3%	0,9%
S-Tiroxina	10 µg/L (o 20%)	12,5%	4,9%	10,9%	2,5%	3,0%	7,0%
S-Proteine totali	10%	6%	2,7%	4,0%	1,4%	1,2%	3,4%
S-Trigliceridi	25%	9%	20,9%	37,2%	10,5%	10,7%	27,9%
Ematocrito	6%	5%	2,8%	6,4%	1,4%	1,7%	4,1%
Emoglobina	7%	4%	2,8%	6,6%	1,4%	1,8%	4,1%
Leucociti	15%	6,5%	10,9%	19,6%	5,6%	5,6%	14,6%
Volume eritrocitario medio	3DS	-	1,3%	4,8%	0,7%	1,2%	2,3%

[a]*http://www.westgard.com/guest17.htm*
CV_I, coefficiente di variabilità biologica intraindividuale; CV_G, coefficiente di variabilità biologica interindividuale.

La scelta del valore di CV_I piuttosto che di CV_G dipende da come l'esame è utilizzato nella pratica clinica. Per monitorare le variazioni in un singolo paziente nel tempo, il CV_I è il più appropriato. Per classificare i pazienti in categorie diagnostiche o terapeutiche utilizzando intervalli di riferimento o limiti decisionali, il CV_G è più importante. Bisogna sottolineare il fatto che la maggior parte delle stime di CV_I e CV_G sono state ottenute in soggetti sani, mentre la maggior parte delle decisioni mediche sono prese su pazienti malati. In questi, le patologie spesso comportano concentrazioni diverse degli analiti e una diversa ampiezza delle loro variazioni.

Statisticamente, la variabilità biologica può essere utilizzata per formulare limiti di imprecisione analitica accettabile, perché se $DS_{analitica}$ è piccola rispetto a $DS_{biologica}$, la variabilità totale è aumentata in modo trascurabile. È stata proposta la seguente terminologia (11, 12):

- qualità minima: $CV_{analitico}$ <0,75 CV_I (aggiunge <25% al CV_{totale});
- qualità desiderabile: $CV_{analitico}$ <0,50 CV_I (aggiunge <12% al CV_{totale});
- qualità ottimale: $CV_{analitico}$ <0.25 CV_I (aggiunge <3% al CV_{totale}).

Alcuni aspetti piuttosto delicati influenzano questo approccio, ad esempio gli effetti sulla variabilità biologica delle variabili preanalitiche. Se la variabili preanalitiche, per esempio, ora del prelievo, postura, dieta, tempo di applicazione del laccio, somministrazione di farmaci, provenienza del prelievo (se i campioni vengono da ambulatori o da pazienti ospedalizzati) sono controllate, allora la variabilità biologica verosimilmente sarà inferiore.

Un ulteriore svantaggio di questo approccio si ha quando i limiti di imprecisione vengono espansi per fornire limiti di TE. Vari Autori hanno proposto che limiti di qualità per il "bias" analitico possano pure essere statisticamente derivati dalla variabilità biologica. Gowans et al. hanno proposto che, per poter condividere gli stessi intervalli di riferimento, i laboratori dovrebbero avere un "bias" <0,25 $(CV_I^2 + CV_G^2)^{1/2}$ (13). Fraser e Petersen, in seguito, hanno raccomandato la seguente terminologia per gli obiettivi del "bias" analitico, analoga a quella raccomandata per l'imprecisione (14):

- bias di qualità minima: <0,375 $\sqrt{(CV_I^2 + CV_G^2)}$;
- bias desiderabile: <0,25 $\sqrt{(CV_I^2 + CV_G^2)}$;
- bias ottimale: <0,125 $\sqrt{(CV_I^2 + CV_G^2)}$.

Un aspetto importante relativo a questi traguardi per il "bias" è che essi sono definiti per decisioni nei singoli pazienti, soggetti alla variabilità biologica; tuttavia, la presenza di un "bias" analitico introduce uno scostamento sistematico di tutti i risultati ottenuti con un metodo, che può comportare variazioni importanti nelle decisioni mediche per un gran numero di pazienti. La variabilità biologica può allargare la distribuzione dei risultati di un esame, ma non sposta il punto centrale della distribuzione, men-

tre un "bias" analitico modifica la posizione della distribuzione, potendo avere un grande effetto in prossimità dei livelli decisionali. Questo concetto è discusso ulteriormente nell'approccio 4, che tratta le caratteristiche delle prestazioni analitiche basate sugli effetti su decisioni mediche assunte in base a linee guida.

La variabilità biologica e i limiti per una qualità "desiderabile" relativi una selezione di analiti, basati sulle informazioni raccolte da Ricos et al. (15), sono riassunti nella Tabella 1.

APPROCCIO 3: VARIAZIONI ACCETTABILI DEI RISULTATI DI LABORATORIO NELL'OPINIONE DEI CLINICI

Vari studi hanno stimato l'entità della variazione del risultato di un esame di laboratorio che indurrebbe un clinico a modificare i suoi programmi di cura. In genere, questi studi hanno presentato specifiche situazioni con varie concentrazioni degli analiti in esame. I limiti di variabilità analitica tollerati dai clinici dipendono dalla loro esperienza: i clinici più esperti mostrano una maggiore tolleranza per le variazioni. Un lavoro molto citato sul significato medico degli esami di laboratorio è stato effettuato da Barnett negli anni '60 (16). Questo Autore utilizzò "l'opinione degli esperti" per stimare come la variabilità di un esame di laboratorio influenzasse le decisioni cliniche. In modo simile, Skendzel pubblicava negli anni '70 un rapporto relativo a un'indagine su 125 internisti, utilizzando una serie di casi clinici seguiti da una lista di risultati possibili di esami di laboratorio per valutare l'entità della variazione di un risultato che avrebbe modificato le decisioni del clinico (17). Skendzel ha poi effettuato un'indagine più ampia su medici dell'"American Medical Association", che è stata utilizzata per definire limiti di prestazioni analitiche da non superare per ottenere un'utilità clinica (18). Più di recente, Thue et al. hanno intervistato i medici di medicina generale norvegesi, definendo un limite di imprecisione analitica per l'emoglobina del 2,8% (19). In modo simile, Skeie et al. hanno intervistato pazienti che attuavano l'automonitoraggio della glicemia e hanno raccomandato limiti di imprecisione del 3,1%–5,0% basati sulle loro decisioni di variare il dosaggio dell'insulina (20).

Alcuni dei limiti di qualità ottenuti con questo approccio sono riassunti nella Tabella 2. Skendzel e colleghi hanno coniato il termine "coefficiente di variazione medico", calcolato usando un fattore di conversione statistico per trasformare la massima variazione accettabile in qualcosa di simile a una DS. La variazione "clinicamente significativa" della concentrazione di un analita è divisa per 1,645 e $\sqrt{2}$ per trasformarla in una DS; poi questo numero è trasformato in un CV moltiplicandolo per 100 e dividendolo per la media tra il valore originale dell'esame e il massimo valore di variazione considerato accettabile. I valori presentati nella Tabella 2 possono sembrare grandi

Tabella 2
Limiti di prestazioni analitiche basati sull'utilità clinica del risultato nell'opinione dei clinici e limiti di "bias" analitico basati sulle distribuzioni delle popolazioni

	Unità	Utilità medica			Limiti di "bias" della popolazione		
		Valore iniziale	Valore modificato	CV medico[a]	Livello decisionale	"Bias"[b]	"Bias" %
S-Bilirubina	mg/dL	0,8	1,4	23,4%	1,1	0,1	9,0%
S-Calcio	mg/dL	9,0	10,6	7,0%	10,2	0,1	1,0%
S-Colesterolo	mg/dL	210	280	12,3%	200	2,3	1,2%
S-Creatinina	mg/dL	1,00	1,50	17,2%	0,8	0,1	12,5%
S-Glucosio	mg/dL	100	130	11,2%	100	2,0	2,0%
S-Ferro	µg/L	150	100	17,2%	—	—	—
S-Fosfato	mg/dL	3,5	2,5	14,3%	2,5	0,1	4,0%
S-Potassio	mmol/L	3,8	3,4	4,8%	3,6	0,1	2,8%
S-Sodio	mmol/L	125	130	1,7%	134	1,5	1,1%
S-Tiroxina	µg/L	60	40	17,2%	50	4	8,0%
S-Proteine totali	g/L	70	85	8,3%	63	2	3,2%
S-Trigliceridi	mg/dL	130	190	16,1%	400	5,8	1,5%
Ematocrito	%	42	37	5,4%	35	0,7	2,0%
Emoglobina	g/L	150	138	3,6%	119	3	2,5%
Leucociti	10^9/L	6,0	3,4	16,4%	3,5	0,2	5,7%
Volume eritrocitario medio	fL	95	100	3,2%	81,5	0,7	1,0%

[a]CV medico = 100 x [(valore modificato – valore iniziale)/(1,645 x √2)]/[(valore modificato + valore iniziale)/2]).
[b]Limite di "bias" = 1 DS della variazione della distribuzione di frequenza cumulativa della popolazione.

e facili da mantenere, ma il CV analitico richiesto al laboratorio per garantire queste prestazioni anche solo ad un livello da 3 a 5 sigma, è pari a un terzo o un quarto di questi limiti di prestazione.

Poichè esistono metodi statistici per valutare la variabilità biologica e analitica, forse queste stime di cambiamenti potrebbero essere valutate meglio usando approcci più scientifici. Comunque, le opinioni di clinici esperti possono essere particolarmente utili nel valutare l'impatto clinico di un "bias" e quello di risultati aberranti. Le decisioni mediche sono in genere basate su un insieme di osservazioni e di misure per ciascun paziente. La verifica della presenza di incongruenze tra i vari fattori è uno strumento utile per identificare risultati errati prima che possano causare eventi avversi.

Domande provenienti da clinici esperti, soprattutto quelli che lavorano in ambiti specialistici, possono costituire un allarme precoce utile per identificare la presenza di problemi nelle prestazioni analitiche del laboratorio. Ad esempio, un incremento improvviso di referti con ipercalcemia potrebbe significare una sovrastima nelle misure del calcio sierico (21). Un aumento della discordanza tra i valori di tireotropina e di tiroxina libera potrebbe indicare la presenza di problemi analitici con l'uno o l'altro dei due esami (22). Allo stesso modo, l'aumentata discordanza tra volume eritrocitario medio e i valori di sideremia, ferritinemia, vitamina B_{12} e/o folati potrebbe essere un allarme precoce di problemi analitici (23). Anche i dubbi dei clinici relativamente alla specificità e ai limiti di rilevazione dei metodi immunologici per la misura di cortisolo e testosterone possono mettere in luce attese di prestazioni analitiche non raggiunte (24-28).

APPROCCIO 4: LIMITI DELLE PRESTAZIONI ANALITICHE BASATI SUGLI EFFETTI SU DECISIONI CLINICHE ORIENTATE DA LINEE GUIDA

Le strategie per standardizzare la pratica clinica di solito sono basate su una combinazione di opinioni di esperti e dati di letteratura. La maggior parte di queste "linee guida" sono state sviluppate da clinici che assumono che i risultati prodotti da laboratori accreditati siano uniformemente di alta qualità e armonizzati. Dettagli sulla standardizzazione delle misure e/o sui traguardi di prestazione sono raramente affiancati ai limiti decisionali proposti da queste linee guida.

Il "bias" analitico può avere un marcato effetto sulla percentuale di pazienti inclusi in ciascun gruppo della linea guida. Si è valutato l'impatto del "bias" analitico su tre linee guida: (a) uso del colesterolo sierico per identificare pazienti a rischio di patologia coronarica, (b) uso della tireotropina sierica per individuare l'ipotiroidismo primario e (c) uso dell'antigene prostata-specifico (PSA) sierico per lo screening del tumore prostatico (29). Un "bias" positivo del 3% aumentava il numero dei pazienti a rischio di patologia coronarica del 16,7% ad una concentrazione di colesterolo di 200 mg/dL. Un bias positivo del 6% dell'ormone tireostimolante causava un aumento del 26,6% dei pazienti con sospetto di ipotiroidismo a un livello decisionale di 5,0 mIU/L. Un "bias" positivo del 6% nel valore del PSA causava un incremento del 11,4% nel numero degli uomini sottoposti a screening, il cui risultato era positivo per tumore della prostata alla soglia di 4,0 µg/L. Percentuali simili di incremento erano trovate anche ad altri livelli di cut-off per ciascuno di questi esami. Benché il traguardo ideale per il "bias" analitico sia zero, ottenerlo potrebbe essere molto costoso e soprattutto non raggiungibile. Analisi relative all'impatto negativo di un eventuale "bias" sulle decisioni mediche, come quelle illustrate sopra, possono fornire utili informazioni di costo-beneficio. Il costo della riduzione del "bias" analitico può essere correlato ai benefici potenziali legati a una pratica migliore. Shermock et al. hanno studiato gli effetti della variazione dei risultati delle analisi di laboratorio nel rapporto di normalizzazione internazionale (INR) su una serie di azioni mediche guidate dall'INR (30). Questi Autori hanno trovato che l'influenza degli errori analitici sulle decisioni mediche dipendeva da quanto un valore era vicino ai limiti decisionali, suggerendo prudenza nell'utilizzare gli stessi limiti di accettabilità per tutti i livelli di concentrazione dell'esame.

Un procedimento possibile per definire limiti di accettabilità per il "bias" analitico è basato su modello delle variazioni a breve termine della distribuzione dei valori di un esame nella popolazione dei pazienti (31). La distribuzione di frequenza cumulativa di 20 gruppi di dati di circa 1000 risultati ciascuno (corrispondenti a pazienti visti alla "Mayo Clinic" ogni giorno) è mostrata nella Figura 1. La variazione tra giorni della popolazione può essere considerata, per le decisioni basate sulle linee guida, alla stessa stregua della variazione biologica intraindividuale per le decisioni sul singolo paziente. Per analogia, se il laboratorio mantiene piccolo il "bias" analitico in confronto con gli scostamenti della distribuzione di frequenza vista nei giorni, gli effetti sulla pratica clinica dovrebbero essere trascurabili. Questo modello propone che il "bias" analitico sia mantenuto a meno di 1 DS della variazione della popolazione. Nella Figura 1 l'ambito tra i limiti di -2 DS e +2 DS per il colesterolo ai limiti decisionali di 200 mg/dL e di 240 mg/dL è 9 mg/dL, quindi il limite di tolleranza del "bias" (1 DS) è 2,3 mg/dL. I limiti di tolleranza del "bias" per il colesterolo e altri analiti di chimi-

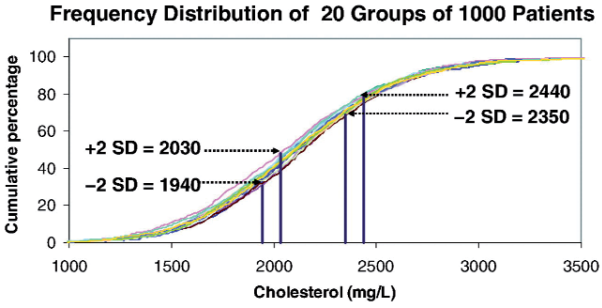


Figura 1
Uso delle variazioni delle distribuzioni cumulative dei risultati di ampie coorti di pazienti per definire i criteri di prestazioni per il "bias". Sono indicati i limiti ±2 DS per i livelli decisionali a 200 e 240 mg/dL. I limiti per il "bias" analitico sono posti a 1 DS.

ca ed ematologia, calcolati usando il modello illustrato, sono riportati nella Tabella 2 (31, 32). Questi limiti sono relativamente piccoli, ma indicano come piccole variazioni del "bias" analitico possano direttamente influire sulla cura del paziente.

APPROCCIO 5: CARATTERISTICHE DELLE PRESTAZIONI ANALITICHE IN RELAZIONE ALLE RICHIESTE DI SUCCESSIVI ESAMI CLINICI

Le decisioni cliniche sono basate su molti fattori che includono i pazienti che presentano i problemi, la loro storia precedente, la storia familiare, i risultati di indagini di laboratorio e le preferenze e convinzioni dei clinici. In una struttura sanitaria complessa ci saranno molte differenze nell'utlizzo degli esami di laboratorio, quindi è difficile analizzare la relazione esatta tra i risultati di laboratorio e le azioni mediche. Per esempio, la misura nel siero dell'ormone paratiroideo (PTH) è la logica conseguenza in un paziente con un'ipercalcemia di nuova insorgenza. Se tutti i medici rispondessero nello stesso modo, ci si aspetterebbe un aumento della frequenza di richieste di PTH in funzione degli aumenti della concentrazione di calcio al di sopra del limite superiore di riferimento. Nella cura dei singoli pazienti c'è però una considerevole variazione nelle modalità di richiesta della misura del PTH, dovute sia alle differenze tra i pazienti che tra i clinici. D'altra parte, analizzando la cura di un gran numero di pazienti, c'è una relazione ben definita tra la concentrazione di calcio sierico e la frequenza relativa con cui il PTH è richiesto entro un breve tempo dal ricevimento del risultato del calcio (33).

La relazione fra i risultati di un esame e la frequenza delle successive procedure messe in atto in un particolare centro medico può essere definita combinando due serie di dati che sono di frequente computerizzati: i referti di laboratorio e i codici di fatturazione delle procedure. Poiché la relazione tra le due variabili spesso non è lineare, possono essere necessari programmi matematici di "curve-fitting" per definire queste relazioni. Nell'esempio riportato nella Figura 2, un totale di circa 100.000 referti di calcio sierico sono stati suddivisi in 4 categorie (maschi e femmine con pagamento Medicare e non-Medicare). Per ciascuna categoria è stato contato il numero di pazienti entro ciascun intervallo di 0,1 mg/dL di calcemia ed è stata calcolata la frequenza relativa di successivi codici CPT4 ["Current Procedural Terminology, 4th edition", modalità di codifica dei test diagnostici, non solo di laboratorio (NdT)]. Si è trovato che molti codici CPT4, oltre al PTH (n=83.970), si associavano alle concentrazioni di calcio sierico, inclusi una seconda misura di calcio sierico (n=82.310), calcio urinario (n=82.340), fosfatasi alcalina sierica (n=84.075), radiografia del torace (n=71.020) e scintigrafia delle paratiroidi (n=78.070) (33).

Queste relazioni matematiche fra i risultati di analisi di laboratorio e le successive procedure di approfondimento non forniscono in modo esplicito limiti di prestazioni, ma solamente un potenziale meccanismo per analizzare gli effetti di un "bias" analitico. Se si assume che la relazione statistica sia, almeno in parte, una relazione causale, allora queste curve che mettono in relazione concentrazione e frequenza di richieste successive possono essere utilizzate per simulare l'effetto di un "bias" analitico. Per esempio, nella Figura 2 se la misura del calcio sierico è sovrastimata di 0,2 mg/dL, i pazienti con un calcio di 10,3 mg/dL riceveranno un risultato di 10,5 mg/dL. Se le richieste di PTH sierico sono riferite al valore di calcemia, il 30% di questi pazienti avrebbe una richiesta di PTH invece del 20% atteso. L'integrazione sull'intero spettro dei valori di calcio fornirebbe una stima dell'impatto di questa sovrastima dei risultati di laboratorio sulla richiesta di procedure cliniche.

APPROCCIO 6: LIMITI DELLE PRESTAZIONI ANALITICHE BASATI SU MODELLI DECISIONALI USATI IN SISTEMI ESPERTI

Sono stati sviluppati sistemi di supporto alla decisione per assistere i clinici nelle scelte relative alla cura dei pazienti. Di solito si usano due tipi di algoritmi in questi sistemi decisionali: algoritmi appresi dalla macchina, basati sui dati, oppure algoritmi decisionali ingegnerizzati basati sulle conoscenze disponibili derivate dall'opinione di esperti e studi di letteratura. Gli algoritmi appresi dalla macchina generalmente hanno un'incertezza intrinseca associata ai dati, mentre i modelli ingegnerizzati sulle conoscenze usano spesso dei punti decisionali fissi e non hanno ambiti di incertezza, a meno che non siano esplicitamente incorporati (34, 35).

Un'assunzione base in questi modelli decisionali è che i sistemi analitici di misura sono stabili e non influenzati da imprecisione e "bias", in particolare ai livelli decisionali. Il "bias" aumenta i risultati falsi-positivi e riduce i falsi-negativi (o viceversa), mentre l'imprecisione accresce entrambi (36). Sono stati usati studi di simulazione per analizzare gli effetti di imprecisione e "bias" sui modelli decisionali. Questi studi aggiungono un errore controllato ai "database" e ne analizzano l'impatto sulle decisioni. Questo tipo di studi può essere utilizzato per aiutare a definire traguardi di prestazioni analitiche (37, 38).

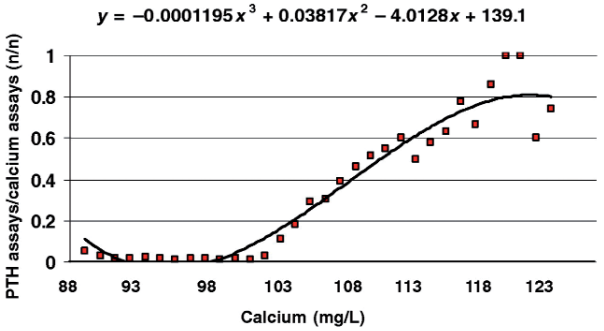


Figura 2
Relazione tra le concentrazioni di calcio sierico e la frequenza di successive richieste di PTH per pazienti di sesso femminile in Medicare. Adattata da Gallaher et al. (33).

L'analisi delle curve delle decisioni cliniche è uno strumento per valutare alcuni dei parametri degli esami diagnostici (39). Questa tecnica non richiede una definizione esplicita di costi, benefici e preferenze, tipicamente utilizzati nelle tecniche di analisi decisionale. Molti sistemi tradizionali di supporto alle decisioni usano semplici diramazioni sì/no a livelli decisionali predefiniti, ad esempio ematocrito <30% o albumina superiore al valore mediano (40, 41). Questi tipi di modelli sono meno utili per valutare l'impatto delle prestazioni analitiche.

Si possono sviluppare modelli statistici parametrici per la diagnosi di specifiche malattie usando la distribuzione dei risultati in pazienti con e senza quella specifica patologia. Anche sistemi diagnostici semplificati che impiegano solo un esame e tre stati di salute (funzione ridotta, normale e aumentata) richiedono insiemi multipli di dati e varie assunzioni. Comunque, una volta sviluppati, questi modelli possono fornire un utile mezzo per analizzare gli effetti di "bias" e imprecisione sulle decisioni mediche.

Il valore potenziale di questi modelli decisionali per valutare traguardi analitici è illustrato con un esempio che utilizza il TSH per classificare i pazienti come ipo-, normo- o ipertiroidei (42). La distribuzione dei valori di TSH nei tre diversi stati è stata sviluppata assieme alla stima della prevalenza di ipotiroidismo (5%) e ipertiroidismo (2%). Si è assunto che il "costo" di una falsa diagnosi fosse uguale per i due tipi di patologia, assegnandogli un valore arbitrario di 10 unità. Il termine "costo" si riferisce a tutte le conseguenze negative di decisioni di errata diagnosi e le unità sono usate solo per fornire una scala relativa. Si è anche assunto che il costo di una mancata diagnosi dipendesse da quanto distante fosse il vero risultato del TSH dal livello decisionale. Per l'ipotiroidismo si è assunto che il costo di un risultato falso-negativo fosse direttamente correlato al valore del TSH, con un fattore di proporzionalità di 3 (costo = 3 x valore di TSH). La Figura 3 mostra che il costo di una diagnosi errata raggiunge il punto più basso (nadir) per valori di TSH compresi fra 5,0 e 10,0 mIU/L. Questa parte relativamente piatta della curva può essere usata per valutare l'impatto clinico dell'incertezza analitica delle misure del TSH in termini di effetti sulla diagnosi di ipotiroidismo. Per l'ipertiroidismo il costo di un risultato falso-negativo era considerato inversamente correlato al valore del TSH (costo maggiore per TSH falsamente bassi) e la curva dei costi mostrava un nadir relativamente piatto per valori di TSH compresi fra 0,1 e 0,2 mIU/L. Questi tipi di modelli potrebbero essere potenzialmente utilizzati per aiutare a definire limiti di prestazioni analitiche collegate all'utilità medica degli esami.

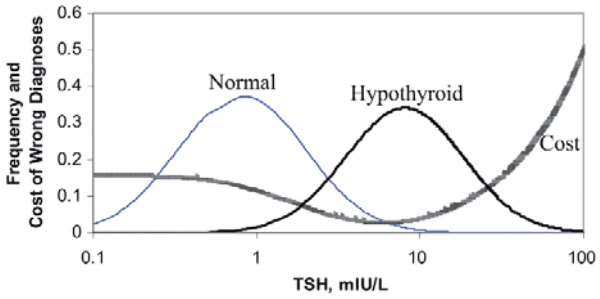


Figura 3
Distribuzione di frequenza e funzione decisionale per l'ottimizzazione dei costi per la diagnosi di ipotiroidismo utilizzando le misure di ormone tireotropo (TSH). Il nadir della funzione dei costi si trova a concentrazioni di TSH comprese tra 5,0 e 10,0 mIU/L. Queste curve potrebbero essere usate per valutare l'effetto delle prestazioni analitiche sulle decisioni mediche. Riprodotta da Klee (42) con il permesso dall'Editore.

DISCUSSIONE

Lo scopo di tutti i laboratori clinici è quello di fornire la qualità dei risultati necessaria per una buona cura del paziente. Tuttavia, la qualità è difficile da valutare se non ci sono specifici criteri di valutazione delle prestazioni. Vari metodi sono stati proposti per definire questi criteri. Quello più ampiamente utilizzato usa la variabilità biologica intra-individuale e tra gruppi di soggetti per fornire limiti entro cui la variabilità analitica è mascherata dalla variabilità biologica. Questi limiti funzionano bene per valutare l'imprecisione, specialmente per le decisioni mediche che riguardano il singolo paziente, ma non risolvono il problema dell'effetto che un "bias" analitico potrebbe avere sulle prestazioni complessive relative ad ampie coorti di pazienti. Questa rassegna riassume alcuni approcci alternativi utili a definire traguardi di prestazioni analitiche in modi che siano correlati al loro effetto sulle decisioni cliniche.

BIBLIOGRAFIA

1. CDC. Current CLIA regulations (including all changes through 01/24/2004). http://www.cdc.gov/clia/regs/toc.aspx.
2. Westgard JO, Klee GG. Quality management - 4th ed. St. Louis: Elsevier Saunders, 2006.
3. Fraser CG. General strategies to set quality specifications for reliability performance characteristics. Scand J Clin Lab Invest 1999;59:487-90.
4. Fraser CG, Kallner A, Kenny D, et al. Introduction: strategies to set global quality specifications in laboratory medicine. Scand J Clin Lab Invest 1999;59:477-8.
5. Dhatt GS, Agarwal MM, Bishawi B, et al. Implementing the Stockholm Conference hierarchy of objective quality criteria in a routine laboratory. Clin Chem Lab Med 2007;45:549-52.
6. Jones BA, Bekeris LG, Nakhleh RE, et al. Physician satisfaction with clinical laboratory services: a College of American Pathologists Q-probes study of 138 institutions. Arch Pathol Lab Med 2009;133:38-43.
7. Plebani M. Quality specifications: self pleasure for clinical laboratories or added value for patient management? Clin Chem Lab Med 2007;45:462-6.
8. Werner M. Linking analytic performance goals to medical outcome. Clin Chim Acta 1997;260:99-115.
9. Sciacovelli L, Secchiero S, Zardo L, et al. External quality assessment schemes: need for recognised requirements. Clin Chim Acta 2001;309:183-99.
10. Westgard QC. Rilibak: quality goals the German way. http://www.westgard.com/rilibak-2.htm.
11. Fraser CG, Hyltoft Petersen P, Libeer JC, et al. Proposals for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8-12.
12. Sciacovelli L, Zardo L, Secchiero S, et al. Quality specifica-

tions in EQA schemes: from theory to practice. Clin Chim Acta 2004;346:87-97.
13. Gowans EM, Hyltoft Petersen P, Blaabjerg O, et al. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. Scand J Clin Lab Invest 1988;48:757-64.
14. Fraser CG, Petersen PH. Analytical performance characteristics should be judged against objective quality specifications. Clin Chem 1999;45:321-3.
15. Ricos C, Alvarez V, Cava F, et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59:491-500.
16. Barnett RN. Medical significance of laboratory results. Am J Clin Pathol 1968;50:671-6.
17. Skendzel LP. How physicians use laboratory tests. JAMA 1978;239:1077-80.
18. Skendzel LP, Barnett RN, Platt R. Medically useful criteria for analytic performance of laboratory tests. Am J Clin Pathol 1985;83:200-5.
19. Thue G, Sandberg S, Fugelli P. Clinical assessment of haemoglobin values by general practitioners related to analytical and biological variation. Scand J Clin Lab Invest 1991;51:453-9.
20. Skeie S, Thue G, Sandberg S. Patient-derived quality specifications for instruments used in self-monitoring of blood glucose. Clin Chem 2001;47:67-73.
21. Bais R. What information should manufacturers provide on their procedures? Clin Chem 2006;52:1624-5.
22. Burman KD. Commentary: discordant measurements of serum triiodothyronine (T4), thyroxine (T3) and thyroid-stimulating hormone (TSH). Clin Chem 2008;54:1246.
23. Snow CF. Laboratory diagnosis of vitamin B12 and folate deficiency: a guide for the primary care physician. Arch Intern Med 1999;159:1289-98.
24. Cohen J, Ward G, Prins J, et al. Variability of cortisol assays can confound the diagnosis of adrenal insufficiency in the critically ill population. Intensive Care Med 2006;32:1901-5.
25. Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: better than a guess? Clin Chem 2003;49:1250-1.
26. Matsumoto AM, Bremner WJ. Serum testosterone assays: accuracy matters. J Clin Endocrinol Metab 2004;89:520-4.
27. Rosner W, Auchus RJ, Azziz R, et al. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab 2007;92:405-13.
28. Vesper HW, Botelho JC, Shacklady C, et al. CDC project on standardizing steroid hormone measurements. Steroids 2008;73:1286-92.
29. Klee GG, Schryver PG, Kisabeth RM. Analytic bias specifications based on the analysis of effects on performance of medical guidelines. Scand J Clin Lab Invest 1999;59:509-12.
30. Shermock KM, Connor JT, Lavallee DC, et al. Clinical decision-making as the basis for assessing agreement between measures of the International Normalized Ratio. J Thromb Haemost 2009;7:87-93.
31. Klee G. A conceptual model for establishing tolerance limits for analytic bias and imprecision based on variations in population test distributions. Clin Chim Acta 1997;260:175-88.
32. Klee GG, Schryver P. Quality assurance for basic haematology cell counts. In: Rowan RM, van Assendelft OW, Preston FE, eds. Advanced laboratory methods in haematology. London: Edward Arnold Publishers, 2002:3-17.
33. Gallaher PM, Mobley RL, Klee GG, et al. The impact of calibration error in medical decision making. Final report. Gaithersburg, MD: National Institute of Standards and Technology, Chemical Science and Technology Laboratory; 2004. Report nr: Planning report 04-1.
34. McNair P, Brender J, Talmon J. Computer-aided test selection and result validation: opportunities and pitfalls. Clin Chim Acta 1998;278:243-55.
35. Sutton AJ, Cooper NJ, Goodacre S, et al. Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. Med Decis Making 2008;28:650-67.
36. Petersen PH, de Verdier CH, Groth T, et al. The influence of analytical bias on diagnostic misclassifications. Clin Chim Acta 1997;260:189-206.
37. Egmont-Petersen M, Talmon JL, Hasman A. Robustness metrics for measuring the influence of additive noise on the performance of statistical classifiers. Int J Med Inform 1997;46:103-12.
38. McNair P, Brender J. Information enhancement in clinical decision making by controlled data generation. Scand J Clin Lab Invest 1990;suppl 202:112-9.
39. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008;8:53.
40. Chu A, Ahn H, Halwan B, et al. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. Artif Intell Med 2008;42:247-59.
41. Ritchie RF. A knowledge-based system to aid with the clinical interpretation of complex serum protein data. Clin Chem Lab Med 2001;39:1045-53.
42. Klee GG. Clinical interpretation of reference intervals and reference limits: a plea for assay harmonization. Clin Chem Lab Med 2004;42:752-7.