

La valutazione dei metodi di analisi qualitativi utilizzando le linee guida contenute nel documento NCCLS EP12-P “User protocol for evaluation of qualitative test performance; Proposed guideline”

F. Poltronieri

Laboratorio Analisi Chimico Cliniche ed Ematologiche, Azienda Ospedaliera, Verona

Introduzione

Gli esami diagnostici di tipo qualitativo sono molto importanti nella Medicina di Laboratorio. Utilizzati nello screening, diagnosi e monitoraggio di numerose patologie. In passato, per la valutazione delle prestazioni delle analisi di questo tipo venivano impiegati protocolli non uniformi, con conseguenti interpretazioni diverse dei risultati ottenuti.

La pubblicazione, nel luglio del 2000, da parte del National Committee for Clinical Laboratory Standards (NCCLS) del documento EP12-P “*User Protocol for Evaluation of Qualitative Test Performance; Proposed Guideline*” si poneva come obiettivo quello di promuovere l’uniformità nella valutazione delle performance dei test qualitativi tra laboratori clinici anche di aree specialistiche diverse, produttori di kit diagnostici (le linee guida danno suggerimenti sulle modalità di presentazione delle caratteristiche tecniche) e agenzie di accreditamento. Da queste linee guida è possibile ottenere un protocollo di valutazione per gli esami di tipo qualitativo che hanno solo due possibili *outcome*.

Sono previsti aggiornamenti del documento per comprendere gli esami che possono avere più di due risposte. Il documento è indirizzato al personale dei laboratori clinici, utilizzatore finale di questi esami. L’applicazione del protocollo di studio consente di verificare le performance del test e la soddisfazione delle esigenze di tipo analitico e clinico richieste dal laboratorio, oltre che essere di supporto alla redazione dei documenti necessari al mantenimento del Sistema Qualità.

Un esame di tipo qualitativo può essere implementato in laboratorio per numerose ragioni, fondamentalmente, deve essere di facile utilizzo e interpretazione e fornire un vantaggio anche di tipo economico.

L’introduzione in routine di un nuovo metodo deve essere documentata con la valutazione ed elaborazione delle performance analitiche del test.

La documentazione necessaria non si limita al solo confronto con un altro metodo.

Tutto il personale deve possedere familiarità con il metodo e devono essere documentate le procedure di controllo di qualità.

Sebbene siano difficilmente realizzabili, a causa delle talvolta notevoli differenze fra i diversi metodi, linee guida universali per i dosaggi di tipo qualitativo esistono, tuttavia, alcune caratteristiche comuni.

Gli esami qualitativi devono fornire all’utilizzatore risultati riproducibili e concordanti con le indicazioni cliniche; devono inoltre essere allestiti studi di riproducibilità e comparazione con campioni di cui sia accertata la diagnosi. Come per i metodi quantitativi, l’utilità clinica del test si determina dalla verifica della sensibilità e specificità, dal valore predittivo e dalla valutazione dell’efficienza del metodo insieme all’indice di prevalenza della malattia.

Gli esami di tipo qualitativo possono essere divisi in base al loro scopo in tre classi.

- *Esami di screening*: usati di solito su larga scala nella popolazione per rilevare la presenza di un fattore o una causa di malattia. È auspicabile che questo tipo di esame sia dotato di una elevata sensibilità al fine di individuare anche le basse positività. La specificità può essere minore rispetto agli esami di conferma, scopo dell’esame è infatti quello di evitare la mancata individuazione di soggetti positivi. La necessità sottoporre a esame di conferma gli eventuali positivi è comunque preferibile rispetto al rischio sociale di incorrere in falsi negativi.
- *Esami diagnostici*: utilizzati nella diagnosi di particolari patologie o condizioni, basate sulla presenza di determinati presupposti clinici. Un esempio può essere l’uso di particolari test colturali in microbiologia. Questo tipo di esame deve essere rapido per essere impiegato clinicamente, anche se ciò può andare a discapito della sensibilità e specificità. La specificità richiesta all’esame può anche non essere eccellente perché comunque i campioni positivi sono sottoposti a esami di conferma.
- *Esami di conferma*: utilizzati per confermare esami di screening o di diagnosi eseguiti in precedenza. Sono gli esami che permettono al clinico di convalidare una diagnosi. Viene richiesta una elevata specificità a discapito, se necessario, della sensibilità e devono possedere un elevato valore predittivo per la malattia.

Definizioni

Per meglio comprendere il documento ufficiale è utile ricordare alcune definizioni dei termini più di frequente utilizzati.

Accuratezza/Misura accurata/ Accuratezza di misura: vicinanza tra valore misurato e valore vero. Viene espressa nella stessa unità di misura del risultato, come differenza tra il valore vero e quello trovato.

Analita: ogni sostanza o costituente per cui il laboratorio esegue una ricerca (si può trattare di ione, molecola, cellula, agente infettivo od organismo o attività).

Sensibilità analitica: concentrazione alla quale il 95% delle analisi risulta positivo.

Specificità analitica: per i test quantitativi è la capacità del metodo di determinare solo il componente che si propone di dosare o, in senso lato, è il modo in cui il metodo risponde alla presenza del solo analita specificato.

Sensibilità clinica: la porzione di pazienti con patologia ben definita in cui l'esame mostra risultati positivi o che eccedono il limite decisionale definito. La presenza della malattia deve essere definita con criteri indipendenti dal risultato del test.

Specificità clinica: porzione di soggetti che non presentano segni di malattia. I risultati dell'esame sono negativi o all'interno dei limiti decisionali definiti.

Controllo /Materiale di controllo: indifferentemente una soluzione o un liofilo ricostituito utilizzato nei processi di controllo di qualità (QC). Il risultato atteso può essere noto all'operatore e va verificato durante il dosaggio. Il materiale di controllo non deve essere utilizzato come calibratore nella stessa seduta in cui si utilizza come controllo.

Cut Off: punto al di sotto del quale il risultato di un test qualitativo è considerato negativo e al di sopra del quale viene considerato positivo (o viceversa).

Falsi negativi: pazienti in cui il valore trovato risulta negativo, ma in cui è accertata la presenza dell'analita ricercato.

Falsi positivi: pazienti in cui il valore trovato risulta positivo, ma in cui è accertata l'assenza dell'analita ricercato.

Gold Standard: termine aspecifico, indica che il processo o il materiale è il più possibile vicino alla perfezione. L'abuso del termine è sconsigliato.

Valore predittivo negativo: valore entro cui è verosimile che un individuo in cui l'esame risulti negativo, non presenti malattia.

Valore predittivo positivo: valore entro cui è verosimile che un individuo in cui l'esame risulti positivo, presenti malattia.

Prevalenza: espressa come rapporto fra il numero di soggetti affetti da patologia, rispetto al totale dei soggetti appartenenti allo stesso gruppo specifico.

Test qualitativi: gruppo di esami in cui è possibile ottenere solo due tipi di risposta (+/-, Si/No, ecc.); in verità alcuni esami qualitativi sono basati su un unico punto discriminante, in alternativa, alcuni esami classificati come qualitativi derivano dalla dicotomia di test quantitativi o con scala ordinaria.

Riproducibilità: vicinanza dei valori trovati sullo stesso campione in una quantità rilevante di misure successive. I parametri del metodo non devono subire modifiche durante le ripetizioni.

Vero negativo: risultato negativo dell'esame in soggetto senza segni o condizioni della patologia.

Vero positivo: Risultato positivo dell'esame in soggetto con segni o condizioni della malattia.

Periodo di familiarizzazione e addestramento

Il periodo di addestramento con il nuovo metodo di analisi deve essere il più completo possibile, comprendere le modalità di esecuzione, di preparazione e conservazione dei reagenti e campioni, simulazioni di analisi e interpretazione dei risultati dei campioni e l'analisi critica dei dati dei QC. L'addestramento deve comprendere, oltre a una sessione dimostrativa, anche una parte pratica in cui siano i futuri utilizzatori ad eseguire ed interpretare l'analisi e i dati relativi. La durata del periodo di valutazione deve essere almeno di dieci giorni.

Materiali per la valutazione

Controlli

Sono richiesti appropriati materiali di controllo per monitorare le performance del test. I materiali di controllo forniti dal produttore possono essere utilizzati, è possibile utilizzare altri materiali di controllo in cui siano stati escluse possibili interferenze dovute alla matrice del controllo stesso.

Nelle comparazioni di metodi si deve, quando possibile, utilizzare lo stesso materiale di controllo per tutti i metodi in valutazione. Di solito l'esecuzione quotidiana dei due livelli di controllo (positivo e negativo) è sufficiente, alcuni metodi particolari richiedono esecuzioni dei controlli più frequenti.

Raccolta dei campioni

I campioni per effettuare le prove di valutazione devono essere raccolti e conservati come da istruzioni del produttore del materiale diagnostico, per evitare la loro degradazione.

Studi di riproducibilità

Controlli positivi e negativi

I materiali di controllo devono essere dosati, in doppio, in ogni seduta analitica e di ogni dosaggio deve essere tenuta registrazione. Se la durata della valutazione è di 10 giorni al termine dello studio saranno disponibili almeno 20 dati; se il periodo di valutazione supera i 10 giorni è possibile dosare in singolo i controlli ad ottenere, comunque, almeno 20 risultati finali per ogni livello di controllo.

Se entrambi i materiali di controllo non rientrano nei limiti previsti la seduta analitica e i dati, relativi ai campioni, ottenuti nella seduta vanno eliminati e la seduta ripetuta. Se nel corso del periodo di valutazione più di una seduta produce risultati del materiale di controllo non conformi è necessario sospendere la prova e contattare il fornitore del kit diagnostico al fine di determinare le cause e di attuare eventuali azioni correttive.

Concentrazioni vicino al cut off

Gli studi di riproducibilità per gli esami qualitativi devono provvedere a stimare la precisione del metodo a concentrazioni vicine al cut off. Non è corretto procedere all'analisi di campioni la cui concentrazione sia molto lontana (in basso o in alto) dal livello decisionale.

Una comune definizione del punto di *cut off* per i test qualitativi è la *concentrazione in cui la ripetizione del test sullo stesso campione si rivela positiva nel 50% dei casi e negativa nell'altro 50%*. Il leggero incremento o decremento della concentrazione sposta di molto la percentuale positiva o negativa dei risultati ottenuti. Questo dimostra come a concentrazioni vicine al limite del cut off l'imprecisione nei metodi sia elevata.

La concentrazione sopra o sotto il livello di cut off alla quale il 95% dei risultati si dimostrano rispettivamente positivi o negativi viene definita "Intervallo al 95% del cut off" del metodo. La capacità del metodo di produrre su campioni rianalizzati un risultato compreso in tale intervallo è indice di buona affidabilità del metodo.

La concentrazione compresa all'interno dell'intervallo del 95% varia a seconda del metodo e dell'analita. Distinguere la differenza rientra tra le prove di valutazione del metodo stesso.

Gli esperimenti di riproducibilità non possono essere impiegati per definire sia intervallo del 95% che i limiti di concentrazione significativi per il metodo, ma possono dare indicazioni quando, questi intervalli, sono all'interno o meno del 20% dell'intervallo di concentrazione dal punto di cut off.

L'esperimento descritto presume che la curva di calibrazione (concentrazione vs segnale) sia lineare per campioni diluiti nell'intervallo di concentrazione prossimo al punto di cut off. Sappiamo che tale asserzione non è sempre valida come ad esempio nei dosaggi EIA di screening ottenuti con miscele di anticorpi anti HIV. Se dall'esperimento risulta che nell'intervallo del +/-

20% di concentrazione dal punto di cut off si descrive l'intervallo del 95% l'esame dimostra un CV del 10%. Una volta determinato l'intervallo del 95% il CV% ($CV = SD/\text{concentrazione cut off}$) si determina dividendo l'intervallo del 95% per due. Questo tipo di stima può essere fatto solo per analiti singoli come glucosio o HbsAg e non per analiti multipli.

Riproducibilità nei metodi qualitativi per concentrazioni vicine al punto di cut off: procedura

- Stabilire la concentrazione di cut off per l'analita in studio. Se non riportato nella confezione del kit è necessario allestire una serie di diluizioni scallari e determinare il punto in cui il 50% dei risultati sia negativo e l'altro 50% positivo.
- La concentrazione corrispondente determinerà il punto di cut off.
- Preparare una serie di aliquote di campioni con concentrazione maggiore e minore del 20% al punto di cut off in quantità sufficiente all'esecuzione di almeno 20 replicati.
- Analizzare i campioni registrando la percentuale di risultati positivi e negativi
- Dai valori percentuali ottenuti sarà possibile ricavare:
 - a) Se il punto di cut off è accurato: il 50% di risultati deve essere positivo e il 50% negativo. Se questo non è verificato la stima del cut off non è adeguata o la curva di estrapolazione del metodo non è lineare in prossimità del punto di cut off.
 - b) Se il +20 o -20% dell'intervallo di concentrazione è compreso nell'intervallo del 95% del cut off del metodo. Se questo non si verifica, sarà necessario rideterminare l'intervallo del 95%.

Comparazione dei metodi

Gli studi di comparazione di due o più metodi prevedono l'analisi degli stessi campioni con i due o più metodi e la successiva comparazione dei risultati ottenuti. Il confronto di un metodo qualitativo può avvenire con un altro metodo qualitativo, con il metodo utilizzato presso il laboratorio che esegue la valutazione, con il metodo di riferimento, con un metodo quantitativo o con la diagnosi clinica.

Di seguito vengono evidenziati i punti maggiormente significativi che deve comprendere la valutazione.

Campioni da analizzare

Devono essere clinicamente significativi per l'analita da misurare, vanno raccolti in quantità sufficiente all'esecuzione di tutte le prove, devono essere conservati correttamente. Lo studio va eseguito, possibilmente, sulla stessa aliquota e contemporaneamente, per limitare variazioni imputabili al tempo e alle modalità di conservazione dei campioni.

Numero dei campioni

Il numero totale di campioni da analizzare dipende dagli intenti che si pone il valutatore. Se per esempio, si dosano 100 campioni per una patologia che ha un'incidenza sulla popolazione del 5%, ci sarà la probabilità di trovare 5 campioni positivi, contro 95 di negativi; ciò può essere soddisfacente se l'intento è quello di determinare i possibili falsi positivi, ma certamente risulta non soddisfacente per la valutazione del numero dei falsi negativi nella popolazione positiva (sensibilità). Una indicazione di massima può essere quella di testare un numero di campioni tale che almeno 50 campioni siano positivi, ciò per comparare i metodi e almeno altri 50 campioni risultino negativi per determinare così la specificità del metodo.

Il numero di campioni analizzati deve permettere oltre ad una accettabile analisi statistica anche la valutazione della variabilità biologica in soggetti con e senza la patologia.

Durata dello studio

Uno studio di comparazione deve svolgersi con sedute analitiche quotidiane per almeno 10-20 giorni al fine di ottenere un numero significativo di valori per la comparazione. I campioni analizzati andrebbero conservati sino alla fine dello studio per eventuali ulteriori analisi. Tutti i valori ottenuti vanno registrati immediatamente e valutati in maniera preliminare. In caso di evidenti discrepanze si cerca di individuare la possibile causa e i valori vanno eliminati. Se non è possibile identificare la causa della seduta aberrante i valori ottenuti non vanno eliminati.

Risultati discrepanti

La discrepanza fra risultati potrebbe essere imputata al metodo in studio o a quello utilizzato come riferimento. Per questo motivo sarebbe auspicabile utilizzare come riferimento il metodo ritenuto "gold standard". Nel caso si stia eseguendo la comparazione di un metodo in cui si ottengono valori numerici che successivamente vengono trasformati in risultati qualitativi i valori discrepanti vanno tabulati. È importante stabilire se le discrepanze sono prossime ai valori di cut off.

Diagnosi clinica

Dettagliate informazioni sulla determinazione di sensibilità e specificità clinica sono descritte nel documento dell'NCCLS GP-10 "Valutazione dell'accuratezza clinica dei test di laboratorio utilizzando le curve ROC". In esso sono riportate le prove di valutazione dell'accuratezza dei test comparate con lo stato clinico dei pazienti. Il presente documento fornisce solo alcuni accenni.

I campioni usati per gli studi di valutazione devono essere rappresentativi dei diversi stati clinici e delle diverse età. Di ogni campione utilizzato dovrebbero essere disponibili adeguate informazioni cliniche e anamnestiche in modo da spiegare eventuali discordanze negli studi di comparazione. Anche se è comune ritrovare nella letteratura studi di comparazione fra

metodi commerciali diversi di una analisi qualitativa, rimane comunque la diagnosi clinica il "gold standard" contro cui i metodi andrebbero comparati.

Analisi dei dati

I dati ottenuti sono valutati con ricerche di tipo statistico di non semplice descrizione sintetica. Ci si limiterà quindi ad un cenno sintetico dei calcoli da eseguire rimandando al documento originale per una descrizione approfondita. Il calcolo di specificità e sensibilità è relativamente semplice quando la diagnosi clinica è nota.

In una tabella a doppia entrata (Tab. I) vengono comparati i risultati ottenuti con le diagnosi cliniche dei campioni. All'interno di ogni cella è riportato il numero di campioni che corrispondono alle etichette della tabella stessa.

Tabella I. Comparazione diagnosi – metodo.

Metodo x	Diagnosi clinica		
	Positivo (malato)	Negativo (sano)	Totale
Positivo	A	B	A + B
Negativo	C	D	C + D
Totale	A+C	B+D	N

Sensibilità = $A/(A+C)$; Specificità = $D/(B+D)$; Efficienza del metodo (totale dei risultati in accordo con la clinica rappresentati nella percentuale dei risultati veri sia positivi che negativi) = $(A+D)/N$.

Quando un test qualitativo è derivato dalla dicotomia di un test quantitativo o su scala ordinaria le performance vengono descritte utilizzando le curve ROC. (NCCLS document GP10-A Assessment of Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plot; Approved Guideline).

Prevalenza e valore predittivo

La prevalenza è il valore espresso in frazione della frequenza della malattia nella popolazione. Utilizzando campioni con diagnosi nota i valori di prevalenza teorici e osservati devono essere simili.

Il valore predittivo di un esame combina la prevalenza della malattia con sensibilità e specificità. Il valore predittivo positivo di un esame è il numero di risultati veri – positivi diviso il numero degli esami risultati positivi (veri e falsi positivi). Il valore predittivo negativo è il numero di veri –negativi diviso il numero totale dei risultati negativi.

Conclusioni

Non è semplice trarre conclusioni dallo studio di un documento NCCLS. È importante comunque sottolineare come tramite le indicazioni contenute sia possibile ottenere una traccia da seguire per procedere alla valutazione di un metodo analitico. Da parte di

tutto il personale del laboratorio deve esserci uno sforzo nell'abituarsi alla consultazione di questi documenti ufficiali e di adeguare le proprie procedure alle indicazioni riportate. Tutte le procedure sono in linea con la documentazione richiesta dal Sistema Qualità di un moderno laboratorio. Sono stati trattati i punti del documento maggiormente significativi, alcuni si sono esclusivamente accennati, in particolare le elaborazioni di tipo matematico e sta-

tistico, in quanto richiederebbero una discussione esclusiva.

Bibliografia

National Committee for Clinical Laboratory Standards (NCCLS). User protocol for evaluation of qualitative test performance; proposed guideline. NCCLS Document EP12-P. Wayne, PA: NCCLS 2000, Vol. 20, n. 15.