

Protocollo per la comparazione di due metodi analitici di laboratorio

Matteo Vidali¹, Michele Tronchin², Ruggero Dittadi³ per il Gruppo di Studio SIBioC - Medicina di Laboratorio "Statistica per il laboratorio"

¹Servizio di Immunoematologia e Medicina Trasfusionale, Ospedale SS. Trinità, Borgomanero, NO

²Abbott Diagnostics, Roma

³Laboratorio Analisi Chimico-Cliniche, Ospedale dell'Angelo, ULSS 12 Veneziana, Mestre, VE

ABSTRACT

Protocol for the comparison of two laboratory methods. Method comparison is one of the main concerns in Laboratory Medicine. With the introduction of new methods with potential advantages (e.g., cost reduction, improvement in efficiency, easier procedures and maintenance) the laboratory staff should investigate if the field method in use may be replaced by the new candidate method without compromising analytical quality nor resulting in a different medical decision or patient management. Several statistical approaches and graphical tools are available to investigate sources of analytical error and for decision-making. In this article, we present an operative protocol for the comparison of two quantitative analytical methods. All sequential steps, including experimental design, familiarization with the new method, quality assessment, sample selection, definition of acceptability criteria, sample measurement, data analysis and evaluation, final decision and reporting, are discussed and exemplified.

PREMESSA

L'esperimento di comparazione di due metodi analitici permette di stimare la differenza (o "bias") tra un metodo candidato e un metodo scelto come confronto. Sebbene tale attività interessi anche i produttori di diagnostici, il protocollo qui presentato è proposto specificatamente per i laboratori clinici. Il laboratorio si trova frequentemente nella necessità di confrontare un nuovo metodo analitico, sviluppato in proprio o validato e/o commercializzato da una ditta, con un metodo già in uso. Il laboratorista, nel condurre l'esperimento di comparazione, è interessato a valutare se il nuovo metodo può eventualmente sostituire il metodo attualmente in uso, consentendo uguali o migliori prestazioni analitiche, senza compromettere il significato e l'impatto clinico del dato prodotto oppure se è necessario stabilire nuovi intervalli di riferimento e nuovi livelli decisionali.

La comparazione tra metodi analitici costituisce parte integrante ed essenziale dell'attività di valutazione dei metodi analitici esplicitamente richiesta sia dallo standard ISO 15189, che contiene i requisiti per la gestione della qualità dei laboratori medici, sia dalla Direttiva IVD 98/79 EC, che definisce i requisiti per i dispositivi diagnostici in vitro (IVD) destinati al mercato

europeo. Secondo le norme, tale attività di valutazione deve essere attentamente pianificata e documentata in ogni sua fase.

Un numero crescente di pubblicazioni ha evidenziato come molti studi di comparazione tra metodi siano stati condotti in modo non corretto, relativamente alla fase del disegno sperimentale e a quella dell'analisi statistica e dell'interpretazione dei dati (1-3). Inoltre, sebbene siano disponibili in letteratura linee guida e protocolli *ad hoc*, questi spesso risultano difficilmente applicabili nella pratica di laboratorio a causa dell'elevato tecnicismo e della scarsità di indicazioni pratiche e/o criteri interpretativi (4).

SCOPO DEL DOCUMENTO

Scopo di questo documento è fornire ai laboratori clinici un protocollo operativo per la comparazione di due metodi analitici quantitativi. Pur cercando di garantire un adeguato rigore scientifico, abbiamo preferito evitare tecnicismi estremi al fine di elaborare un protocollo che risultasse relativamente agevole e facilmente applicabile nelle varie realtà laboratoristiche, fornendo, oltre a informazioni teoriche, anche indicazioni pratiche e criteri interpretativi atti a guidare l'operatore in tutte le fasi

Corrispondenza a: Matteo Vidali, Servizio di Immunoematologia e Medicina Trasfusionale, Ospedale SS. Trinità, Via Mons. Cavigioli 5, 28021 Borgomanero (Novara). Tel. 0322848357, Fax 0322845005, E-mail matteo.vidali@gmail.com

Ricevuto: 02.11.2015

Accettato: 13.11.2015

Pubblicato on-line: 05.05.2016

DOI: 10.19186/BC_2016.006

dell'esperimento di comparazione: dalla scelta dei campioni, alla verifica della qualità del nuovo metodo, alla definizione dei criteri di accettabilità, alla conduzione dell'esperimento, alle metodologie statistiche (compreso il calcolo e l'interpretazione dei risultati), alla valutazione dell'accettabilità, fino alla produzione di un documento riassuntivo finale. Il protocollo è corredato da esempi didattici, inclusi calcoli statistici, funzioni matematiche disponibili sui comuni fogli elettronici, grafici e reportistica. Inoltre, nella sezione del Gruppo di Studio SIBioC "Statistica per il laboratorio" del sito SIBioC (<http://www.sibioc.it>) è disponibile per i soci un modulo "software", che consente l'esecuzione di tutti i calcoli statistici descritti in questo protocollo.

FASI OPERATIVE

Poiché l'esperimento di comparazione rappresenta un'attività che richiede notevoli risorse, è necessaria un'attenta pianificazione. Si possono individuare le seguenti fasi operative:

1. definizione dello scopo dell'esperimento di comparazione;
2. raccolta delle informazioni e della documentazione relativa alle strumentazioni utilizzate e ai metodi investigati;
3. familiarizzazione con il nuovo metodo e verifica delle prestazioni analitiche;
4. selezione dei campioni;
5. definizione dei limiti di accettabilità;
6. conduzione dell'esperimento;
7. analisi grafica e statistica dei dati;
8. valutazione dell'accettabilità del nuovo metodo;
9. conclusioni e produzione del rapporto finale.

Definizione dello scopo dell'esperimento di comparazione

In questa prima fase occorre brevemente descrivere lo scopo dell'esperimento di comparazione, che dovrebbe essere quello di valutare l'entità dell'errore sistematico introdotto con l'impiego del metodo in prova e/o stabilire se il nuovo metodo è accettabile per l'uso previsto, potendo quindi essere utilizzato come sostituto/alternativa di quello attualmente in uso (Tabella 1). Per errore sistematico si intende uno scostamento o differenza tra il valore della grandezza ritenuto "vero" e il valore assunto con la misura. In quanto sistematico, lo scostamento si presenta sempre nello stesso verso, cioè

si ottengono risultati sempre maggiori o sempre minori del valore considerato "vero". Si identifica un errore sistematico costante, quando l'entità dello scostamento è indipendente dalla concentrazione dell'analita, e un errore sistematico proporzionale, quando lo scostamento varia all'aumentare della concentrazione dell'analita.

L'accettabilità è stabilita in base a differenti criteri, quali l'imprecisione combinata dei due metodi, l'errore analitico totale massimo ammissibile e/o altri criteri descritti in letteratura o stabiliti da commissioni di esperti (vedi sotto).

Raccolta delle informazioni e della documentazione relativa alle strumentazioni utilizzate e ai metodi investigati

In questa seconda fase occorre ricercare e documentare le informazioni più importanti relative alle strumentazioni e ai metodi utilizzati nell'esperimento di comparazione. In particolare, è necessario indicare la strumentazione, la/e matrice/i utilizzata/e o utilizzabili (compresa l'eventuale sostanza anticoagulante, accelerante o conservante), la quantità necessaria di campione, le condizioni di conservazione del campione, il principio di reazione, i calibratori utilizzati, le eventuali interferenze conosciute o limitazioni, le prestazioni analitiche riportate nella scheda tecnica, tra cui imprecisione, inesattezza, ambito misurabile, limite di rilevabilità e/o quantificazione (LOD e/o LOQ). Se la strumentazione esegue automaticamente la diluizione del campione oltre un certo livello di concentrazione di analita, questa informazione deve essere riportata (Tabella 2).

Familiarizzazione con il nuovo metodo e verifica delle prestazioni analitiche

Prima di procedere con l'analisi, è essenziale che tutti gli operatori coinvolti nell'esperimento di comparazione siano addestrati e conoscano adeguatamente entrambi i metodi e le strumentazioni utilizzate (uso dei calibratori, operatività, manutenzioni, principali cause di errore). Inoltre, la qualità analitica di entrambi i metodi deve essere stata verificata prima di procedere con l'esperimento di comparazione. In particolare, è necessario verificare almeno la ripetibilità del metodo in prova, anche se già espressamente dichiarata nella scheda tecnica fornita dal produttore. La

Tabella 1

Esempio di come documentare la definizione dello scopo dell'esperimento

In questo esperimento sono stati comparati due metodi per la determinazione di *...(indicare l'analita)...* nel siero/plasma (o altra matrice) di soggetti umani. Il primo metodo (da ora definito "metodo di riferimento"/"metodo in uso"), venduto dalla ditta *...(indicare la ditta)...* e commercializzato in Italia dalla ditta *...(indicare la ditta)...*, è implementato su strumentazione *...(indicare la strumentazione)...* ed è attualmente in uso presso questo laboratorio da *...(indicare la data)...* Il secondo metodo (da ora definito "metodo in prova"), venduto dalla ditta *...(indicare la ditta)...* e commercializzato in Italia dalla ditta *...(indicare la ditta)...*, è implementato su strumentazione *...(indicare la strumentazione)...* ed è stato validato dalla stessa ditta (si allega documentazione tecnica). L'accettabilità sarà valutata sulla base dell'imprecisione combinata dei due metodi e dell'errore analitico totale massimo ammissibile descritto in letteratura per questo analita (*riportare la voce bibliografica*).

Tabella 2

Esempio di come documentare la raccolta delle informazioni prima dell'inizio del protocollo

Metodo di riferimento/in uso (indicare il metodo)

Parametri operativi:

Strumentazione:
 Principio di reazione:
 Matrice:
 Campione (quantità):
 Condizioni di conservazione del campione:
 Calibratori:
 Interferenze:
 Diluizione:
 Prestazioni analitiche riportate dal produttore:
 Imprecisione totale ("intra- e inter-assay"):
 Inesattezza:
 Intervallo di misura:
 LOD e/o LOQ:

Metodo in prova (indicare il metodo)

Parametri operativi:

Strumentazione:
 Principio di reazione:
 Matrice:
 Campione (quantità):
 Condizioni di conservazione del campione:
 Calibratori:
 Interferenze:
 Diluizione:
 Prestazioni analitiche riportate dal produttore:
 Imprecisione totale ("intra- e inter-assay"):
 Inesattezza:
 Intervallo di misura:
 LOD e/o LOQ:

LOD, limite di rilevabilità; LOQ, limite di quantificazione.

verifica della qualità del metodo attualmente in uso dovrebbe essere già disponibile prima della sua implementazione e controllata grazie a CQI e VEQ.

Verifica della ripetibilità del metodo in prova

Se il nuovo metodo è già stato precedentemente validato da soggetti terzi (ad es., ditta produttrice), è sufficiente eseguire un protocollo rapido per verificare quanto dichiarato. Sebbene il protocollo ancora maggiormente utilizzato in ambito internazionale sia rappresentato dallo schema 3x5 (3 replicati per 5 giorni) (5, 6), la letteratura più recente suggerisce di aumentare il numero di replicati e/o di giorni al fine di ottenere una stima più realistica della ripetibilità (7). È quindi consigliabile utilizzare uno schema 5x5 (5 replicati per 5 giorni) o uno in cui il prodotto replicati x giorni sia ≥ 20 (7). In questo protocollo di comparazione (e nel "software" messo a disposizione) vengono quindi riportati i calcoli relativi allo schema 3x5 e 5x5. In alternativa, per i metodi non validati da soggetti terzi ma sviluppati in proprio, è necessario adottare un protocollo più esteso (in genere si utilizza uno schema 2x20, cioè 2 replicati per 20 giorni). La verifica della ripetibilità è inoltre utile per valutare nelle fasi successive l'accettabilità del nuovo metodo.

Protocollo 3x5 e 5x5 per la verifica della ripetibilità

Tempistiche e scelta dei campioni. E' consigliabile distribuire i 5 giorni del protocollo per la verifica della ripetibilità su un intervallo temporale relativamente lungo (almeno 2 settimane). In tal modo, la ripetibilità stimata terrà conto di eventuali fattori che potrebbero cambiare nelle operazioni ordinarie (calibrazioni, temperatura, operatori, lotto reagenti). La verifica andrebbe effettuata utilizzando almeno 2 materiali di controllo o "pool" di pazienti con concentrazioni vicine sia a eventuali valori decisionali, sia a quelle utilizzate dalla ditta nella fase di validazione e riportate nella scheda tecnica.

Procedura operativa.

- Per ogni livello di concentrazione selezionato (almeno due), analizzare una volta al giorno per 5 giorni, 3 o 5 replicati indipendenti dello stesso campione. Per replicato non si intende lo stesso campione misurato più volte, ma aliquote indipendenti dello stesso: le 3 o 5 aliquote vanno cioè preparate e analizzate come se fossero 3 o 5 campioni indipendenti (se il metodo prevede estrazioni o altre fasi preparative occorre trattare in questo modo tutte le aliquote del campione e non misurare una singola aliquota per 3 o 5 volte).
- In ogni seduta analitica verificare la qualità analitica (tramite l'utilizzo di carte di controllo); se necessario fare più calibrazioni tra una seduta analitica e l'altra (mettersi cioè in condizioni simili a quelle dichiarate dalla ditta produttrice).
- Analizzare i dati (vedi formule seguenti ed esempio).

Esempio di calcolo per la verifica della ripetibilità totale per il protocollo 3x5 (numero di giorni, 5; numero di replicati, 3) (Tabella 3).

- Media giorno: rappresenta la media dei 3 replicati di ogni giorno (ottenibile in MS Excel e in OpenOffice con la funzione *MEDIA()*).
- DS giorno: rappresenta la DS dei 3 replicati di ogni giorno (ottenibile in MS Excel e in OpenOffice con la funzione *DEV.ST()*).
- Varianza giorno: è la DS giorno elevata al quadrato (ottenibile o elevando la DS giorno al quadrato o in MS Excel e in OpenOffice con la funzione *VAR()*).
- Media generale (media delle 5 medie giorno): nell'esempio in Tabella 3, 141,333.
- Varianza entro i giorni (media delle 5 varianze giorno): nell'esempio in Tabella 3, 0,400.
- Varianza tra i giorni (varianza delle 5 medie giorno): nell'esempio in Tabella 3, 4,611.
- Varianza totale o varianza entro il laboratorio: nell'esempio in Tabella 3, 4,878.

La varianza totale è ottenuta sommando la varianza tra i giorni (4,611) e la varianza entro i giorni (0,400) moltiplicata per il coefficiente $(N-1)/N$ dove N rappresenta il numero dei replicati [nel nostro caso $(3-1)/3$, quindi: $4,611+0,400 \cdot 2/3 = 4,878$].

La ripetibilità totale (o ripetibilità entro il laboratorio) si può esprimere come:

Tabella 3

Esempio di dati e calcoli per il protocollo 3x5 di verifica della ripetibilità

Giorno	Replicato 1	Replicato 2	Replicato 3	Media giorno	DS giorno	Varianza giorno
1	140	140	140	140,000	0,000	0,000
2	138	139	138	138,333	0,577	0,333
3	143	144	144	143,667	0,577	0,333
4	143	143	142	142,667	0,577	0,333
5	142	143	141	142,000	1,000	1,000

Tabella 4

Esempio di dati e calcoli per il protocollo 5x5 di verifica della ripetibilità

Giorno	Replicato 1	Replicato 2	Replicato 3	Replicato 4	Replicato 5	Media giorno	DS giorno	Varianza giorno
1	140	139	138	138	140	139,000	1,000	1,000
2	140	143	141	143	137	140,800	2,490	6,200
3	140	138	136	141	136	138,200	2,280	5,200
4	141	144	142	143	144	142,800	1,304	1,700
5	139	140	141	138	141	139,800	1,304	1,700

- a) DS entro il laboratorio (ottenuta calcolando la radice quadrata della varianza entro il laboratorio): 2,209;
 b) CV (ottenuto calcolando il rapporto tra la DS entro il laboratorio e la media generale): 2,209/141,333=0,01563 o 1,563%.

Esempio di calcolo per la verifica della ripetibilità totale per il protocollo 5x5 (numero di giorni, 5; numero di replicati, 5) (Tabella 4).

- Media generale (media delle 5 medie giorno): 140,120.
- Varianza entro i giorni (media delle 5 varianze giorno): 3,160.
- Varianza tra i giorni (varianza delle 5 medie giorno): 3,172.
- Varianza totale o varianza entro il laboratorio: 5,700 (3,172+3,160·4/5).

La ripetibilità totale (o ripetibilità entro il laboratorio) si può esprimere come:

- a) DS entro il laboratorio: 2,387;
 b) CV: 2,387/140,120 = 0,01704 o 1,704%.

Confronto della ripetibilità totale con quanto dichiarato dal produttore. La ripetibilità ottenuta con il protocollo 3x5 o 5x5 può essere confrontata con quella dichiarata dal produttore. Il produttore può fornire la ripetibilità come DS o come CV. Nel secondo caso, per ottenere la DS è sufficiente applicare la formula $DS=(CV/100) \cdot M$, dove M rappresenta la media generale.

Per il particolare analita considerato nell'esempio protocollo 3x5, il produttore ha fornito una DS=2,06. Se in alternativa avesse fornito un CV=1,5%, si sarebbe ottenuta una DS=1,5/100·141,333=2,12.

Si prosegue calcolando il parametro T (effettivi gradi di libertà) con la formula:

$$T = \frac{[(N-1) \cdot S_r^2 + (N \cdot S_b^2)]^2}{\left(\frac{N-1}{G}\right) \cdot (S_r^2)^2 + \left(\frac{N^2 \cdot (S_b^2)^2}{G-1}\right)}$$

con:

N = numero dei replicati (nell'esempio = 3);

G = numero dei giorni (nell'esempio = 5);

S_b^2 = varianza tra i giorni (nell'esempio = 4,611);

S_r^2 = varianza entro i giorni (nell'esempio = 0,400).

Utilizzando il parametro calcolato T e la tabella statistica della distribuzione chi-quadrato (Tabella 5), si cerca il valore critico C (si può anche utilizzare la funzione di MS Excel o di OpenOffice *INV.CHI(0,025;T)* o *INV.CHI.QUAD.DS(0,025;T)* sostituendo a T il valore calcolato precedentemente). Nel nostro esempio:

T = 4,470;

C (con T approssimato a 4) = 11,14;

DS dichiarata dalla ditta (ottenibile dalla documentazione fornita dal produttore) = 2,06.

Si calcola infine un valore limite (o valore di verifica) ottenuto con la formula:

$$V = DS \cdot \frac{\sqrt{C}}{\sqrt{T}}$$

dove DS rappresenta la DS totale dichiarata dalla ditta e C e T i parametri calcolati precedentemente. Nel nostro esempio: V=3,25.

Poiché la DS entro il laboratorio ottenuta sperimentalmente con il protocollo 3x5 (2,209) è inferiore al valore limite (3,25), la DS totale dichiarata dal produttore risulta verificata.

Se il protocollo dovesse evidenziare un'impresione (ripetibilità totale) peggiore (DS maggiore del valore limite) rispetto quanto dichiarato dal produttore sarebbe necessario fare ulteriori indagini prima di procedere con le fasi successive. Il laboratorista può riverificare la ripetibilità utilizzando lo stesso protocollo 3x5 o 5x5

Tabella 5*Tabella per la statistica chi-quadrato*

T	C	T	C	T	C	T	C
3	9,35	4	11,14	5	12,83	6	14,45
7	16,01	8	17,53	9	19,02	10	20,48
11	21,92	12	23,34	13	24,74	14	26,12
15	27,49	16	28,85	17	30,19	18	31,53
19	32,85	20	34,17	21	35,48	22	36,78
23	38,08	24	39,36	25	40,65	26	41,92
27	43,19	28	44,46	29	45,72	30	46,98
31	48,23	32	49,48	33	50,73	34	51,97
35	53,20	36	54,44	37	55,67	38	56,90
39	58,12	40	59,34	41	60,56	42	61,78

*T, effettivi gradi di libertà; C, valore critico.***Tabella 6***Valori di G in funzione del numero di dati (N) da utilizzare per il test di Grubbs*

N	G	N	G	N	G
5	1,7637	20	3,0008	70	3,6217
10	2,4821	25	3,1353	80	3,6729
15	2,8061	30	3,2361	90	3,7163
16	2,8521	35	3,3160	100	3,7540
17	2,8940	40	3,3807	120	3,8167
18	2,9325	50	3,4825	200	3,9777
19	2,9680	60	3,5599	300	4,0935

oppure un protocollo più esteso (ad es., 2x20). Eventualmente può essere necessario contattare la ditta fornitrice del kit per chiarimenti. Se la discrepanza rimane dopo ulteriori indagini, per le successive fasi bisogna utilizzare il valore calcolato dall'utilizzatore finale e non quello dichiarato dalla ditta, in quanto quest'ultimo potrebbe sottostimare la vera ripetibilità totale.

Identificazione e trattamento di eventuali valori estremi. Tra i dati raccolti per l'esperimento di verifica della ripetibilità è possibile, talvolta, rilevare la presenza di una o più osservazioni che si discostano in una certa misura da tutte le altre (valori estremi). Questi risultati possono essere il frutto di errori grossolani o, al contrario, essere osservazioni genuine che solo apparentemente sembrano discordare dalle rimanenti. Conservare le prime osservazioni (errori) potrebbe condurre a un falso aumento della ripetibilità, mentre eliminare le seconde (valori genuini) potrebbe falsamente ridurre la ripetibilità e condurre quindi a stime troppo ottimistiche della stessa. In questo caso, si consiglia di procedere in questo modo:

- verificare che il valore estremo non dipenda da errori di trascrizione ed eventualmente procedere alla correzione;
- applicare un test statistico per individuare i valori

estremi. In letteratura sono descritti numerosi test statistici utili all'identificazione di valori estremi ("outliers"). Di seguito viene riportato il test di Grubbs:

- calcolare media e DS di tutti i dati dell'esperimento di verifica della ripetibilità (15 nel protocollo 3x5, 25 nel protocollo 5x5);
 - individuare il valore di G nella Tabella 6;
 - calcolare i limiti dell'intervallo di accettabilità come $media \pm G \cdot DS$. Tutti i valori al di fuori di questo intervallo sono considerati valori estremi;
- c) calcolare la ripetibilità totale confrontandola con quella dichiarata dal produttore, conservando l'eventuale osservazione estrema. Se la ripetibilità ottenuta sperimentalmente è compatibile con quella dichiarata dal produttore, procedere con il protocollo di comparazione senza eliminare il valore estremo. Se la ripetibilità totale è significativamente maggiore di quella dichiarata dal produttore, eliminare tutti i replicati del giorno in cui si è verificata l'osservazione estrema e procedere a una nuova sessione giornaliera (3 replicati nel protocollo 3x5 o 5 replicati nel protocollo 5x5). In presenza di più di due valori estremi nell'intero esperimento, alcuni autori suggeriscono di ripetere l'intero studio di ripetibilità. La Tabella 7 riassume la modalità di documentazione di questa fase del protocollo.

Tabella 7

Esempio di come documentare la fase di familiarizzazione e verifica preliminare

Familiarizzazione

Tutti gli operatori adibiti al protocollo sono stati adeguatamente addestrati da un tecnico specialista della ditta produttrice nelle operazioni ordinarie, come uso dei calibratori, manutenzione strumento, gestione errori e verifica della qualità analitica.

Calcolo e verifica della ripetibilità

Prima di procedere all'esperimento di comparazione si è verificata la qualità analitica delle strumentazioni e dei kit attraverso l'utilizzo di carte di controllo.

Per la strumentazione in uso, il laboratorio effettua regolarmente/quotidianamente CQI e partecipa a VEQ (*riportare i dati*)

Per la verifica della ripetibilità del nuovo metodo dichiarata dal produttore, si è utilizzato un protocollo 3 (triplicati) x 5 (giorni). Come campioni sono stati utilizzati materiali di controllo della ditta ... (*indicare la ditta*)... con concentrazioni pari a ... (*indicare le concentrazioni utilizzate*)..., simili a quelle utilizzate dalla ditta produttrice del metodo nuovo.

I valori misurati e utilizzati per il calcolo della ripetibilità sono stati i seguenti:

concentrazione 1: ... (*indicare la concentrazione del primo campione utilizzato*)...

	repl1	repl2	repl3
giorno1	1a	1b	1c
giorno2	2a	2b	2c
giorno3	3a	3b	3c
giorno4	4a	4b	4c
giorno5	5a	5b	5c
giorno n	na	nb	nc

(*inserire i valori corrispondenti agli altri livelli di concentrazione utilizzati e i valori di ripetibilità calcolati*)

La presenza di dati estremi è stata valutata tramite test di Grubbs.

Conclusioni

La ripetibilità totale calcolata utilizzando il campione con concentrazione pari a ... (*indicare la concentrazione*)... è risultata essere pari a ... (*indicare la ripetibilità totale calcolata espressa o come DS o come CV*)..., inferiore al valore di verifica ottenuto sulla base della ripetibilità dichiarata dal produttore.

Selezione dei campioni

Particolare attenzione deve essere posta nella scelta dei campioni, perché da essi dipende in larga misura l'attendibilità e l'utilità delle conclusioni dell'esperimento di comparazione.

Numero

La maggior parte dei lavori e delle linee guida disponibili in letteratura suggeriscono un numero minimo di campioni pari a 40-60 (4, 8). È necessario che le concentrazioni di questi campioni coprano l'intero intervallo misurabile e l'intervallo dei valori osservabili in clinica. Inoltre, una maggior proporzione di campioni dovrebbe essere selezionata nell'intorno di valori decisionali per meglio stimare l'errore dei due metodi in prossimità di questi. In letteratura sono disponibili alcune linee guida pratiche su come scegliere i campioni per i vari analiti (4). Di conseguenza, il numero di campioni può essere subordinato al tipo di analita selezionato.

È preferibile utilizzare un maggior numero di campioni (60-100) quando si confrontano metodi analitici basati su principi di reazione diversa. Infatti, l'utilizzo di un maggior numero di campioni rende maggiormente probabile l'identificazione di possibili interferenti.

Aliquote singole o duplicate

L'uso di due aliquote indipendenti dello stesso campione sugli strumenti, misurate in ordine diverso, è senza dubbio preferibile all'utilizzo di una singola

aliquota per strumento. Infatti, l'uso di replicati dello stesso campione su ogni strumento consente di verificare che eventuali valori anomali non siano dovuti a errori grossolani. In presenza di differenze marcate tra il metodo in uso e quello in prova, è possibile verificare, tramite i replicati, l'attendibilità dei risultati di ciascun metodo. L'uso di campioni in singolo obbliga invece a verificare eventuali anomalie e/o discrepanze tra i due metodi al momento della determinazione per eventualmente ripetere la stessa. Se la procedura di confronto dei dati ottenuti dai due strumenti messi a confronto viene fatta al termine dell'esperimento, potrebbe non essere più materialmente possibile rianalizzare lo stesso campione.

Pur ritenendo migliore la prima strategia, in questa sede, allo scopo di facilitare la maggiore diffusione e il più ampio utilizzo di questo documento, viene proposto il secondo approccio, sottolineando tuttavia la necessità di verificare al momento della lettura strumentale i valori dei campioni e le differenze tra i due metodi, rappresentando i dati in un diagramma, non appena questi si rendono disponibili.

Durata dell'esperimento di comparazione

Maggiore è la durata dell'esperimento, migliore è la stima dell'errore sistematico, sia perché viene minimizzato l'errore sistematico di una particolare seduta e/o giorno, sia perché la stima ottenuta in un periodo più lungo tiene conto di differenti fattori che potrebbero essere trascurabili in un periodo breve, ma presenti comunque nelle operazioni ordinarie del laboratorio.

Appare conveniente suddividere, ad esempio, 40 campioni in un periodo di 5 giorni (8 campioni + controllo/i per 5 giorni). Nulla vieta di aumentare questo periodo e/o il numero di campioni per le ragioni riportate in precedenza. Infatti, come per il protocollo di verifica della ripetibilità, è consigliabile distribuire le sedute dell'esperimento di comparazione su un intervallo temporale più lungo (almeno due settimane), in modo che la stima tenga conto di eventuali variabili che possono cambiare nelle operazioni ordinarie (calibrazioni, temperatura, operatori, lotto reagenti). È necessario specificare nel rapporto finale dell'esperimento la durata dell'esperimento e le date delle singole sedute analitiche (Tabella 8).

Definizione dei limiti di accettabilità

Prima di iniziare l'esperimento, dovrebbe essere specificato il limite massimo di errore analitico consentito senza compromettere l'interpretazione del risultato dell'esame e di un'eventuale decisione clinica.

Un metodo "in prova" sarà considerato accettabile se le differenze sono comprese entro l'imprecisione combinata dei metodi e/o entro un determinato limite massimo di errore consentito.

Accettabilità basata sull'imprecisione combinata dei due metodi

L'imprecisione combinata dei due metodi (combinazione delle imprecisioni dei due metodi) è calcolata con la formula seguente:

$$\sqrt{(CV_1^2 + CV_2^2)}$$

dove CV_1 rappresenta il CV del metodo in uso e CV_2 il CV del metodo in prova, cioè la sua ripetibilità totale, espressa come CV e calcolata nella fase precedente (rapporto tra la ripetibilità espressa come DS e valore dichiarato del campione utilizzato per l'esperimento di ripetibilità o media delle concentrazioni ottenute).

Se si ipotizza che i due metodi non differiscano significativamente, allora il 95% delle differenze tra i due metodi dovrebbe essere compreso nell'intervallo $0 \pm 1,96$ x imprecisione combinata.

Tabella 8

Esempio di come documentare la fase sperimentale del protocollo

In questo esperimento sono stati utilizzati 40 campioni misurati in singolo su entrambi gli strumenti. L'intervallo di valori considerato corrisponde a quello misurabile dello strumento: ...*(indicare l'intervallo)*...

L'analisi dei 40 campioni è stata suddivisa in ...*(indicare il numero dei giorni)*... giorni, dal...*(data inizio)*... al...*(data fine)*...

Tabella 9

Esempio di come documentare l'accettabilità basandola sull'imprecisione combinata

CV metodo in uso: 5%

CV metodo in prova (calcolato nell'esperimento di ripetibilità): 3%

Imprecisione combinata: $\sqrt{(5^2+3^2)}=5,8\%$

Intervallo 95% delle differenze: $0 \pm 1,96 \times 5,8\%$, cioè tra -11,4% e +11,4%. Ad es., limiti delle differenze in corrispondenza di un cut-off teorico di 60 mg/dL: $0 \pm 1,96 \times (0,058 \times 60) = 0 \pm 6,8$ mg/dL

Se i CV dei due metodi cambiano lungo l'intervallo di concentrazione, occorre utilizzare diverse imprecisioni combinate per diversi intervalli di concentrazione (Tabella 9).

Accettabilità basata sulle specifiche di qualità analitica (massimo errore accettabile)

Le specifiche di qualità analitica, da cui ricavare il massimo errore accettabile, possono essere fissate sulla base di modelli diversi descritti in letteratura, che presentano differenti livelli gerarchici, tra cui:

- modello basato sulla valutazione dell'effetto delle prestazioni analitiche su "outcome" clinici in specifiche condizioni cliniche;
- modello basato sulla valutazione dell'effetto delle prestazioni analitiche su decisioni cliniche;
- modello basato sulla variabilità biologica dell'analita;
- modello basato su raccomandazioni pubblicate di esperti;
- modello basato sullo stato dell'arte attuale (di strumentazioni e/o metodi) derivato per esempio da programmi di VEQ.

Per un esempio sull'utilizzo di modelli differenti il lettore può consultare i lavori di Stockl et al. (9) e Bianchi et al. (10).

Se disponibili, è possibile fissare il limite massimo di errore accettabile utilizzando i dati di variabilità biologica. Occorre sottolineare che il dato di variabilità biologica non è sempre disponibile, o, se lo è, può essere poco attendibile, come ad esempio quando i risultati nei soggetti analizzati si posizionano mediamente intorno al LOD. In questi casi si può ricorrere a modelli di livello gerarchico inferiore (vedi elenco precedente).

La Tabella 10 mostra le formule per il calcolo dell'imprecisione massima, dell'inesattezza massima e dell'errore massimo accettabile. L'errore massimo ammissibile calcolato in questo modo è disponibile in letteratura espresso come percentuale (<http://www.westgard.com/>). Se espresso nell'unità di misura dell'analita, per convertirlo in percentuale è sufficiente dividere l'errore per la concentrazione corrispondente alla decisione clinica (concentrazione di analita scelta) e moltiplicare il risultato per 100 (si

Tabella 10

Formule per il calcolo dell'imprecisione massima e dell'inesattezza massima e dell'errore massimo accettabili

Per l'imprecisione: $I_{max} = 0,5 \cdot CV_I$, dove CV_I indica la variabilità biologica intraindividuale

Per l'inesattezza: $B_{max} = 0,25 \cdot \sqrt{CV_I^2 + CV_G^2}$ dove CV_I e CV_G indicano, rispettivamente, la variabilità biologica intra- e interindividuale

Per l'errore totale diversi criteri potranno essere scelti, in alternativa, a seconda della criticità dell'analita dosato:

$$TE_{max} = B_{max} + 1,65 \cdot I_{max} \text{ (errore più restrittivo)}$$

$$TE_{max} = B_{max} + 2 \cdot I_{max}$$

$$TE_{max} = B_{max} + 3 \cdot I_{max}$$

$$TE_{max} = B_{max} + 4 \cdot I_{max} \text{ (errore meno restrittivo)}$$

Tabella 11

Esempio di come documentare la definizione dei limiti di accettabilità

Da dati di letteratura (...citare fonte...) per l'analita in oggetto si sono ottenuti i seguenti limiti:

Massima imprecisione ammissibile: 4,8%

Massima inesattezza ammissibile: 6,4%

Massimo errore totale ammissibile: $6,4\% + 1,65 \times 4,8\% = 14,3\%$

[In alternativa si può riportare (esempio di errore massimo ammissibile selezionato per la comparazione di due metodi per il monitoraggio terapeutico di un farmaco anticonvulsivante): "L'errore massimo ammissibile in questo esperimento di comparazione è stato fissato al 15%, considerando l'errore massimo descritto in letteratura per farmaci simili per azione e/o struttura chimica (...citare la fonte...)"]

trasforma cioè una DS in un CV, percentuale come già visto in precedenza). Talvolta gli errori sono forniti sia come concentrazione sia come percentuale, a seconda della concentrazione dell'analita considerato (Tabella 11).

Conduzione dell'esperimento

In questa fase vanno documentate le modalità con cui è stato condotto l'esperimento, i controlli giornalieri (riportare il valore teorico, l'intervallo $\pm 2DS$, il valore misurato), eventuali errori/problemi riscontrati, calibrizioni effettuate e ripetizioni.

La pratica di scegliere per l'esperimento di comparazione i campioni valutati in routine con il metodo di riferimento (o in uso) e campioni refrigerati o congelati per il metodo in valutazione non è corretta, a meno che sia dimostrata la stabilità dei campioni a tali temperature o a meno di rianalizzare nuovamente gli stessi campioni anche con il metodo in uso. Occorre ricordare, tuttavia, che la conservazione dei campioni, in alcune situazioni, potrebbe introdurre nuove variabili differenti da quelle evidenziabili nella misura quotidiana dei campioni. L'approccio migliore sarebbe quello di analizzare i campioni con il metodo in uso e con quello in prova entro due ore (o meno nel caso di analiti con stabilità inferiore) l'uno dall'altro (Tabella 12).

Analisi grafica e statistica dei dati raccolti

Optando per la valutazione in singolo su ogni strumentazione, un grafico preliminare dovrebbe già essere stato costruito durante il periodo di acquisizione dei dati per identificare eventuali differenze estreme tra i due metodi. Due sono gli strumenti grafici e statistici maggiormente utilizzati (entrambi implementati nel "software"): il grafico della retta di regressione e il

grafico delle differenze (Bland-Altman plot) nelle sue varianti.

Grafico e analisi di regressione

I dati ottenuti con il metodo di riferimento o in uso (asse x) sono riportati in un grafico verso i dati ottenuti con il metodo in prova (asse y) unitamente alla linea di identità ($y=x$) e alla retta di regressione. Se questa operazione viene fatta progressivamente, non appena nuovi dati sono disponibili, è possibile individuare subito eventuali discrepanze e rianalizzare un campione per confermare o correggere i risultati non corretti.

Per individuare la retta di regressione è possibile utilizzare diversi modelli di regressione. Il comune modello di regressione lineare semplice (o ordinaria), ampiamente utilizzato, ha il merito della semplicità di calcolo e della disponibilità in molti "software". Tuttavia, non sempre è corretto utilizzarlo. Infatti, la regressione lineare semplice assume che il metodo in uso (quello sull'asse delle ascisse) sia privo di errore e che l'errore del metodo in prova (asse delle ordinate) sia distribuito normalmente e sia costante lungo tutto l'intervallo di concentrazioni studiato. Queste assunzioni raramente sono soddisfatte nella pratica. Alcuni autori utilizzano la regressione lineare ordinaria quando il coefficiente di correlazione (r) tra i due metodi risulta $>0,99$ (per ampi intervalli di concentrazioni) o $>0,975$ (per intervalli non estesi) oppure suggeriscono di adottare una versione modificata della stessa (11). Oggi sono disponibili a basso prezzo "software" grafici e/o statistici che consentono modelli di regressione alternativi, come quello di Deming (o Deming pesato) o quello non parametrico di Passing-Bablok (12). In questo protocollo, la preferenza viene data a quest'ultimo, in

Tabella 12

Esempio di diario per documentare la conduzione dell'esperimento

Giorno 1 (data: ...):

- metodo in uso: controlli entro 1 DS (*indicare livello*); nessuna calibrazione effettuata; determinazione di 50 campioni e selezione di 8 campioni (la selezione è fatta per ottenere un numero di campioni che coprono l'intero intervallo di misura); nessun errore strumentale; nessuna diluizione del campione;
- metodo in prova: controlli entro 1 DS (*indicare livello*); nessuna calibrazione effettuata; determinazione di 8 campioni entro 2 ore dall'esecuzione con il metodo in uso; nessun errore strumentale; nessuna diluizione del campione; ispezione grafica dei risultati.

Giorno 2 (data: ...):

- metodo in uso: controlli tra 1 e 2 DS (*indicare livello*); nessuna calibrazione effettuata; determinazione di 60 campioni e selezione di 8 campioni; nessun errore strumentale; nessuna diluizione del campione;
- metodo in prova: controlli entro 1 DS (*indicare livello*); nessuna calibrazione effettuata; determinazione di 8 campioni entro 2 ore dall'esecuzione con il metodo in uso; nessun errore strumentale; nessuna diluizione del campione; ispezione grafica dei risultati: identificazione di un risultato anomalo (valore metodo in uso: ...; valore metodo in prova: ...); rianalisi entro 1 ora dello stesso campione con il metodo in prova (valore: ...).

Proseguire con i dati relativi alle giornate successive.

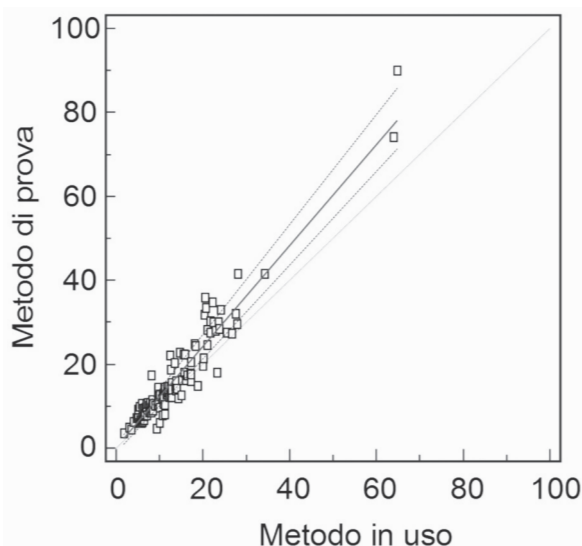
quanto il modello non parametrico di Passing-Bablok richiede il minor numero di assunzioni. Infatti, nella regressione di Passing-Bablok i valori estremi ("outliers") possono essere inclusi, è consentita l'imprecisione in entrambi i metodi e non è necessario né che l'errore sia normalmente distribuito né che sia costante lungo l'intervallo di concentrazioni. Il "software" di comparazione messo a disposizione sul sito SIBioC permette di applicare il modello di regressione di Passing-Bablok.

L'analisi di regressione consente di valutare l'errore sistematico costante (intercetta) e proporzionale (pendenza). È sempre necessario riportare non solo questi due parametri, stime puntuali della vera intercetta e della vera pendenza, ma anche i relativi intervalli di confidenza al 95% (stima intervallare), cioè gli intervalli in cui abbiamo una certa confidenza (al 95%) si possano trovare la vera intercetta e la vera pendenza. In particolare, per dimostrare l'assenza di errore sistematico costante e proporzionale, i relativi intervalli di confidenza devono includere 0 (zero) per l'intercetta e 1 (uno) per la pendenza.

L'impatto clinico di un errore sistematico determinato statisticamente andrà comunque interpretato sulla base dei criteri di accettabilità stabiliti; è infatti possibile che piccoli errori sistematici proporzionali o costanti, identificati con l'analisi di regressione, pur statisticamente significativi, non precludano l'utilizzo di un determinato metodo. In questo protocollo di comparazione non vengono fornite indicazioni relativamente all'indice di correlazione; anzi il suo uso è fortemente sconsigliato. Infatti, l'indice di correlazione permette di valutare l'associazione tra due variabili, ma non l'accordo (Figura 1).

Grafico di Bland-Altman e determinazione del "bias"

Un altro grafico utile è quello di Bland-Altman, dove le concentrazioni dell'analita (esprese come media dei due metodi) sono graficate verso le differenze tra i due metodi (13, 14). Il grafico, quindi, presenterà sull'asse



Numero campioni analizzati: 123

Metodo in uso: media (DS), 13,4 (9,2) mg/dL

Metodo in prova: media (DS), 16,3 (12,0) mg/dL

Regressione di Passing-Bablok:

Intercetta: $-0,02$ (intervallo di confidenza al 95%: $-1,07/0,84$)

Pendenza: $1,21$ (intervallo di confidenza al 95%: $1,12/1,31$)

Equazione: $y = -0,02 + 1,21x$

Commento: presenza di errore sistematico proporzionale (infatti la pendenza è >1 e 1 non è compreso nell'intervallo di confidenza al 95%).

Figura 1

Esempio di come documentare l'analisi di regressione dei dati.

delle ascisse la media dei due metodi e sull'asse delle ordinate la loro differenza. In alternativa, la media dei due metodi (asse delle ascisse) può essere graficata verso le differenze dei due metodi espresse in percentuale rispetto alla media, cioè $\text{metodo1} - \text{metodo2} / \text{media} * 100$ (asse delle ordinate). Questo metodo è utile soprattutto quando l'intervallo di misura è ampio.

Si procede valutando l'eventuale presenza di errore

sistematico calcolando il "bias", ottenuto come media di tutte le differenze e il relativo intervallo di confidenza al 95%. Vi sarà quindi presenza di un errore sistematico significativo ("bias" significativo) quando il valore 0 (zero) non sarà compreso nel relativo intervallo di confidenza al 95%. In assenza di errore sistematico, i punti corrispondenti alle differenze tra i due metodi dovrebbero accumularsi casualmente intorno alla linea dello zero, cioè dovrebbero osservarsi differenze positive e negative prossime allo zero per ogni livello di concentrazione. È possibile osservare anche l'eventuale presenza di particolari tendenze delle differenze e/o l'apertura a ventaglio delle stesse, segno della dipendenza delle differenze dal livello di concentrazione. Nella Figura 2 sono riportate alcune situazioni di comune riscontro.

Oltre alla differenza media (e al relativo intervallo di confidenza al 95%) è necessario anche calcolare l'intervallo che comprende il 95% di tali differenze utilizzando $\text{media} \pm 1,96\text{DS}$, dove media e DS sono rispettivamente la media ("bias") e la DS delle differenze trovate. Questo intervallo, spesso trascurato e non riportato, ha invece una notevole importanza nella corretta interpretazione dell'esperimento di comparazione, perché permette di identificare l'entità della maggior parte (95%) delle differenze riscontrate. Un intervallo, ad esempio, da -66,7% a 40% (differenze espresse come percentuale della media delle concentrazioni) indica che un metodo può fornire un valore che è la metà dell'altro oppure un valore che è 150% più alto di quello ottenuto con l'altro metodo.

Come si può osservare nella Tabella 13, le differenze tra metodi espresse come percentuale della media dei due metodi possono non essere di immediata comprensione. In alternativa, è possibile calcolare l'intervallo che comprende il 95% delle differenze anche con metodi non parametrici, ad esempio come intervallo

compreso tra i percentili 2,5° e 97,5° (ottenibili in MS Excel e in OpenOffice utilizzando la funzione *PERCENTILE()*). Questa procedura alternativa è consigliabile quando le differenze tra metodi non presentano una distribuzione normale e/o simmetrica. Infatti, in presenza di una distribuzione asimmetrica, il calcolo $\text{media} \pm 1,96\text{DS}$ potrebbe sovrastimare l'intervallo al 95% delle differenze (Figura 3).

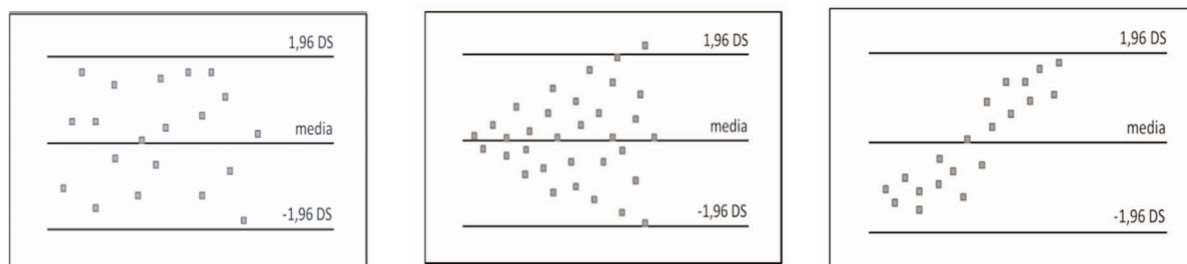
È consigliabile utilizzare ambedue le metodologie di confronto (regressione e grafico delle differenze), che dovrebbero fornire risultati congruenti.

Valutazione dell'accettabilità delle prestazioni del nuovo metodo

Accettabilità basata sull'imprecisione combinata dei due metodi

Nello stesso grafico delle differenze è possibile aggiungere due linee che rappresentano l'intervallo $0 \pm 1,96 \times \text{CV}$, dove CV rappresenta l'imprecisione combinata dei due metodi calcolata nella fase di definizione dei limiti di accettabilità. Come descritto in precedenza, se la ripetibilità varia lungo l'intervallo di concentrazioni (situazione abbastanza comune), occorre utilizzare CV differenti lungo l'intervallo di concentrazioni. Il "software" di comparazione consente di impostare fino a 3 CV differenti.

Se i risultati dei due metodi sono identici entro l'imprecisione combinata, le loro differenze dovrebbero essere simmetricamente distribuite intorno allo zero e il 95% di queste dovrebbero essere comprese nell'intervallo delimitato dalle due linee di riferimento. Il grafico nella Figura 4 è stato realizzato utilizzando due imprecisioni: fino a 10 mg/dL (metodo 1: 6%, metodo 2: 6,5%; imprecisione combinata: 8,8%) e >10 mg/dL (metodo 1: 4%, metodo 2: 4,5%; imprecisione



Nessuna dipendenza delle differenze dalla concentrazione: i punti si addensano casualmente intorno alla linea della media delle differenze.

Dipendenza tra la concentrazione (x) e le differenze (y): le differenze aumentano all'aumentare della concentrazione. In questo caso si ha una sovrastima dell'intervallo corrispondente al 95% delle differenze se si utilizza la formula $\text{media} \pm 1,96\text{DS}$. In alternativa è possibile utilizzare metodi più robusti (con il calcolo dei percentili 2,5° e 97,5°).

Dipendenza tra la concentrazione (x) e le differenze (y): un metodo produce in media risultati inferiori all'altro metodo per livelli di concentrazione bassi e mediamente superiori per livelli di concentrazione alti. Questa evenienza rappresenta una situazione da indagare ulteriormente.

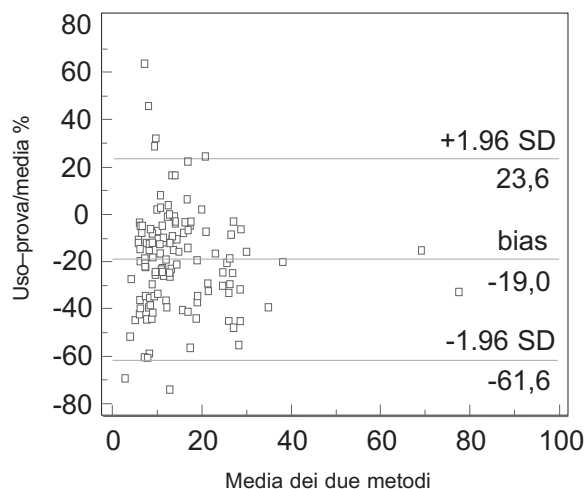
Figura 2

Esempi di grafici di Bland-Altman e relativa interpretazione.

Tabella 13

Significato delle differenze tra metodi espresse come percentuale della media dei due metodi

Differenza/media %	Commento
-100%	Il valore ottenuto con il metodo 1 è ~33% del valore ottenuto con il metodo 2 Es: metodo 1: 100 mg/L; metodo 2: 300 mg/L; $\text{diff}(\text{metodo 1} - \text{metodo 2}) = 100 - 300 = -200$ mg/L media: $(100+300)/2=200$; $(\text{diff metodi}/\text{media metodi})\% = (-200/200)\% = -100\%$
-90%	Il valore ottenuto con il metodo 1 è ~38% del valore ottenuto con il metodo 2
-80%	Il valore ottenuto con il metodo 1 è ~43% del valore ottenuto con il metodo 2
-70%	Il valore ottenuto con il metodo 1 è ~48% del valore ottenuto con il metodo 2
-60%	Il valore ottenuto con il metodo 1 è ~54% del valore ottenuto con il metodo 2
-50%	Il valore ottenuto con il metodo 1 è ~60% del valore ottenuto con il metodo 2
-40%	Il valore ottenuto con il metodo 1 è ~67% del valore ottenuto con il metodo 2
-30%	Il valore ottenuto con il metodo 1 è ~74% del valore ottenuto con il metodo 2
-20%	Il valore ottenuto con il metodo 1 è ~82% del valore ottenuto con il metodo 2
-10%	Il valore ottenuto con il metodo 1 è ~90% del valore ottenuto con il metodo 2
+10%	Il valore ottenuto con il metodo 1 è ~110% del valore ottenuto con il metodo 2
+20%	Il valore ottenuto con il metodo 1 è ~122% del valore ottenuto con il metodo 2
+30%	Il valore ottenuto con il metodo 1 è ~135% del valore ottenuto con il metodo 2
+40%	Il valore ottenuto con il metodo 1 è ~150% del valore ottenuto con il metodo 2
+50%	Il valore ottenuto con il metodo 1 è ~167% del valore ottenuto con il metodo 2
+60%	Il valore ottenuto con il metodo 1 è ~186% del valore ottenuto con il metodo 2
+70%	Il valore ottenuto con il metodo 1 è ~208% del valore ottenuto con il metodo 2
+80%	Il valore ottenuto con il metodo 1 è ~233% del valore ottenuto con il metodo 2
+90%	Il valore ottenuto con il metodo 1 è ~264% del valore ottenuto con il metodo 2
+100%	Il valore ottenuto con il metodo 1 è ~300% del valore ottenuto con il metodo 2



Numero campioni analizzati: 123

"Bias" medio: -19,0% (intervallo di confidenza al 95%: -22,8/-15,1)

DS delle differenze: 21,7%

Intervallo 95% delle differenze ("bias" $\pm 1,96 \times \text{DS}$):

limite inferiore: -61,6 (intervallo di confidenza al 95%: -68,2/-54,9)

limite superiore: 23,6 (intervallo di confidenza al 95%: 17,0/30,3)

Commento: presenza di "bias" significativo (l'intervallo di confidenza al 95% del "bias" non comprende lo zero). Il 95% delle differenze dei due metodi è compreso nell'intervallo tra -61,6% e 23,6%.

Figura 3

Esempio di grafico delle differenze espresse come rapporto percentuale con la media.

combinata: 6,0%). In questo esempio, i due metodi risultano identici entro l'imprecisione combinata (solo il 5% delle differenze sono esterne all'intervallo). Occorre interpretare con molta cautela le differenze tra i metodi per valori bassi di concentrazioni. Infatti, a concentrazioni molto basse anche piccole differenze assolute possono tradursi in grandi differenze percentuali senza comportare un impatto clinico significativo (Figura 5).

Accettabilità basata su specifiche di qualità prefissate

Per valutare l'accettabilità basata su specifiche di qualità prefissate si identifica l'errore totale massimo ammissibile come descritto in precedenza e si costruisce un diagramma (MEDx chart), graficando l'imprecisione (asse x) verso lo scostamento sistematico ("bias") (asse y) e rappresentando come un punto le prestazioni del metodo in prova (15). Le coordinate del punto sono ottenute utilizzando per l'ascissa l'imprecisione (calcolata dall'esperimento di ripetibilità) e per l'ordinata il "bias" (usando l'equazione della retta di regressione applicata a un certo livello di concentrazione scelto). Nel grafico sono anche rappresentate 4 zone corrispondenti a diverse specifiche di qualità (non accettabile, marginale, buona, ottima) (Figura 6).

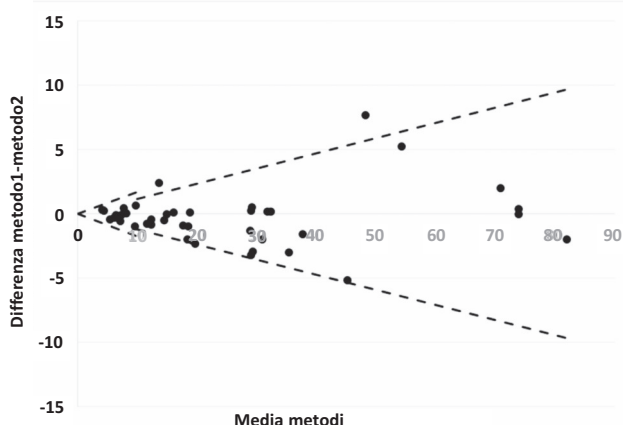
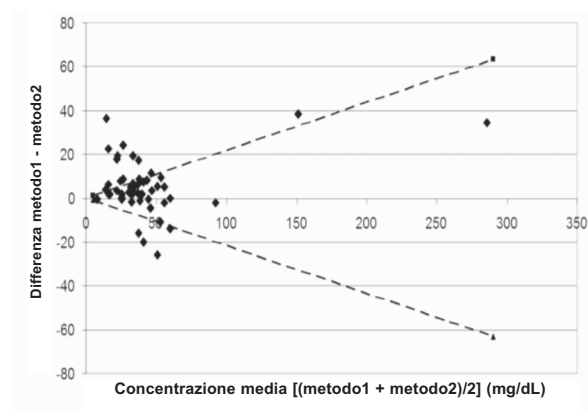


Figura 4
Grafico di accettabilità con in ascissa la media dei due metodi e in ordinata la loro differenza.



CV metodo 1: 7,8%
CV metodo 2: 8%
Imprecisione combinata: $\sqrt{(7,8^2+8^2)}=11,2\%$
Linee corrispondenti all'intervallo:
 $0 \pm 1,96 \times 11,2\% = 0 \pm 21,9\%$
Es.: a 150 mg/dL, intervallo da -32,9 a 32,9
(calcolo: $0 - 1,96 \times 0,112 \times 150$ e $0 + 1,96 \times 0,112 \times 150$)
Commento: i due metodi non sono identici entro l'imprecisione combinata (36% delle differenze esterne all'intervallo, rispetto al 5% atteso).
Nota: per questo esempio di grafico si è ipotizzato che le due imprecisioni (7,8% e 8%) rimanessero costanti lungo tutto l'intervallo di concentrazioni.

Figura 5
Esempio di come documentare la valutazione dell'accettabilità del nuovo metodo mediante imprecisione combinata.

Esempio di costruzione del grafico MEDx chart.

- Sull'asse delle ordinate si fissa il punto corrispondente all'errore massimo ammissibile trovato in letteratura o stabilito sulla base di specifici criteri: nell'esempio della Figura 6, 14,3.
- Sull'asse delle ascisse si fissano 3 punti corrispondenti ai vari criteri di qualità: errore max/2, errore max/3, errore max/4: nell'esempio, $14,3/2=7,15\%$, $14,3/3=4,77\%$, $14,3/4=3,56\%$.
- Si tracciano le 3 linee utili per suddividere il grafico nelle 4 zone.

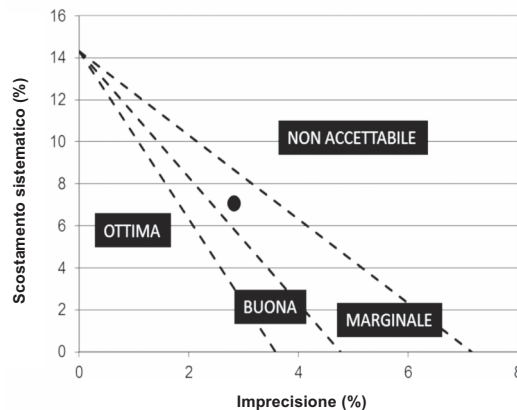
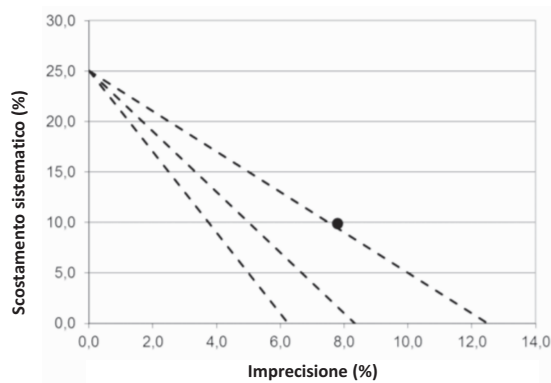


Figura 6
Esempio di grafico MEDx chart.



Errore massimo ammissibile (TE_{max}) ottenuto dalla letteratura (...citare fonte...): 25%
Criteri di accettabilità (dalla linea più esterna a quella più interna):
 $TE_{max}/2=12,5\%$
 $TE_{max}/3=8,3\%$
 $TE_{max}/4=6,25\%$
Imprecisione metodo in prova (ottenuto dall'esperimento di ripetibilità): 7,8%
"Bias" metodo in prova al valore decisionale di 30 $\mu\text{g/L}$ (ottenuto inserendo il valore 30 nella retta di regressione e calcolando l'errore percentuale rispetto al valore atteso di 30 $\mu\text{g/L}$):
equazione di regressione: $y=0,87+1,07x$
se $x=30$ allora $y=0,87+1,07 \times 30=32,97$.
Rispetto al valore atteso di 30 l'errore è: $(32,97-30)/30=0,099=9,9\%$ ("bias").
Il punto ha quindi come coordinate: $x=7,8\%$ e $y=9,9\%$
Commento: i due metodi non sono identici entro una qualità analitica prefissata.

Figura 7
Esempio di come documentare i risultati del protocollo.

- Si rappresenta un punto utilizzando come coordinate le prestazioni del metodo in prova (ascissa=imprecisione misurata con l'esperimento di ripetibilità; ordinata=inesattezza ("bias") valutata con l'analisi di regressione): nell'esempio, il protocollo rapido 3x5 ha confermato una imprecisione pari al 3%; l'equazione di regressione trovata con l'analisi di regressione di Passing-Bablok è risultata essere

Tabella 14*Esempi di report finale del protocollo*

Esempio 1: L'esperimento di comparazione ha evidenziato che il nuovo metodo, pur in presenza di un minimo errore sistematico proporzionale (...riportare pendenza e intervallo di confidenza al 95%...), risulta accettabile entro l'errore massimo ammissibile derivato dalla letteratura.

Esempio 2: L'esperimento di comparazione ha evidenziato sia un errore sistematico costante (...riportare intercetta e intervallo di confidenza al 95%...) che proporzionale (...riportare pendenza e intervallo di confidenza al 95%...). Inoltre, la prestazione del nuovo metodo è risultata non accettabile sia entro l'imprecisione combinata dei metodi, sia rispetto al livello di qualità prefissato. Pertanto esso non può, allo stato attuale, sostituire il metodo in uso in laboratorio.

$y=0,92x+1,5$ (cioè pendenza=0,92 e intercetta pari a 1,5). A un livello di concentrazione scelto di 150 mg/dL, la retta stima una concentrazione pari a $150 \cdot 0,92 + 1,5 = 139,5$ mg/dL, con un errore quindi pari a $(150 - 139,5) / 150 = 7\%$. Il punto avrà quindi coordinate $x=3\%$ e $y=7\%$, dimostrando per il metodo in prova prestazioni di qualità marginale (Figura 6).

Conclusioni e produzione del rapporto finale

Questa fase integra i dati di regressione con quelli della valutazione dell'accettabilità. Queste due analisi, sebbene complementari, costituiscono due approcci differenti e possono condurre a risultati discordanti. In tale eventualità, noi suggeriamo di preferire il criterio dell'accettabilità basata su specifiche di qualità prefissata, valutando se la differenza tra i due metodi è maggiore dell'errore totale massimo ammissibile per l'analisi in esame.

La regressione, evidenziando un errore sistematico proporzionale o costante, può suggerire, ad esempio, una differente specificità dei due metodi, un diverso tipo di standardizzazione o differenze dovute alla matrice, ecc. Il grafico di regressione e, particolarmente, il grafico di Bland-Altman, permettono inoltre di visualizzare e comprendere facilmente la dispersione dei dati. Questi due strumenti forniscono quindi importanti informazioni che il laboratorista deve valutare attentamente; tuttavia, la presenza di un piccolo, seppur significativo, errore sistematico può non precludere l'introduzione di un metodo nel laboratorio (in sostituzione dell'attuale), se tale errore non compromette il significato e l'impatto clinico del dato prodotto.

In generale, se il metodo in prova è risultato accettabile, esso può sostituire il metodo attualmente in uso e i dati dell'esperimento devono essere inseriti nei documenti del sistema della qualità del laboratorio insieme alle considerazioni conclusive (Tabella 14). Al contrario, se i due metodi non sono identici entro l'imprecisione combinata o entro specifiche prefissate di qualità, si conclude che non è possibile sostituire il metodo attuale con quello candidato e occorre procedere in altro modo (Figura 7). In primo luogo occorre revisionare tutti i dati dell'esperimento di comparazione per evidenziare possibili errori grossolani o metodologici.

È anche opportuno ribadire che la non accettabilità di un metodo non significa che il metodo candidato presenti necessariamente prestazioni e/o caratteristiche inferiori

al metodo in uso, ma unicamente che non è possibile, allo stato attuale, rimpiazzare il metodo in uso con quello candidato senza ulteriori interventi. A tale proposito, con la recente introduzione di strumentazioni analitiche tecnologicamente avanzate nel laboratorio clinico, metodi automatizzati in uso vengono confrontati con metodi di gerarchia analitica superiore (ad es., spettrometria di massa) e frequentemente si osservano valori molto discordanti (ad es., nel dosaggio di vitamina D o farmaci immunosoppressori). È chiaro che in questi casi il metodo candidato può presentare caratteristiche diverse da quello attualmente in uso in termini di specificità analitica. In questi casi la non accettabilità, risultato dell'esperimento di comparazione, impone una rivalutazione analitica più approfondita per accertare che effettivamente le differenze siano da addebitare al metodo in uso e, accertato questo, eventualmente procedere alla definizione di nuovi intervalli di riferimento.

Infine, sono possibili situazioni nelle quali, a fronte di un "bias" minimo, si osserva una dispersione dei dati molto elevata (evidenziata soprattutto dal grafico di Bland-Altman e dall'analisi dell'imprecisione combinata). In questo caso, i due metodi, sebbene mediamente confrontabili, forniscono dati contraddittori per molti dei campioni valutati nell'esperimento di comparazione. A questo punto, la scelta di implementare il nuovo metodo richiederebbe valutazioni di tipo differente, in particolare relative alle prestazioni cliniche del metodo stesso, che tuttavia esulano dagli scopi del presente protocollo (16).

CONFLITTO DI INTERESSI

Nessuno.

BIBLIOGRAFIA

- Altman DG, Bland JM. Measurement in Medicine: the analysis of method comparison studies. *The Statistician* 1983;32:307-17.
- Mantha S, Roizen MF, Fleisher LA, et al. Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesth Analg* 2000;90:593-602.
- Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method-comparison studies. *Clin Chem* 1973;19:49-57.
- CLSI. Measurement procedure comparison and bias estimation using patient samples; Approved guideline, 3rd ed. CLSI document EP09-A3. Wayne, PA: Clinical and

- Laboratory Standards Institute, 2013.
5. Chesher D. Evaluating assay precision. *Clin Biochem Rev* 2008;29:S23-6.
 6. Clinical and Laboratory Standards Institute. User verification of performance for precision and trueness; Approved guideline, 2nd ed. CLSI document EP15-A2. Wayne, PA: Clinical and Laboratory Standards Institute, 2005.
 7. Clinical and Laboratory Standards Institute. User verification of precision and estimation of bias; Approved guideline, 3rd ed. CLSI document EP15-A3. Wayne, PA: Clinical and Laboratory Standards Institute, 2014.
 8. Linnet K. Necessary sample size for method comparison studies based on regression analysis. *Clin Chem* 1999;45:882-94.
 9. Stockl D, Sluss PM, Thienpont LM. Specifications for trueness and precision of a reference measurement system for serum/plasma 25-hydroxyvitamin D analysis. *Clin Chim Acta* 2009;408:8-13.
 10. Bianchi V, Locatelli M, Vidali M, et al. Vitamina D: valutazione di un metodo cromatografico liquido ultra veloce (UPLC) e confronto con tecnica immunometrica e cromatografia liquida ad alta prestazione (HPLC) tradizionale. *Riv Ital Med Lab* 2012;8:138-48.
 11. Sotgia S, Zinellu A, Pinna GA, et al. A new general regression-based approach for method comparison studies *Clin Chem Lab Med* 2008;46:1046-9.
 12. Passing H, Bablok W. A new biometrical procedure for testing the quality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem* 1983;21:709-20.
 13. Petersen PH, Stockl D, Blaabjerg O, et al. Graphical interpretation of analytical data from comparison of a field method with a reference method by use of difference plots. *Clin Chem* 1997;43:2039-46.
 14. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135-60.
 15. Westgard JO. A method evaluation decision chart (MEDxchart) for judging method performance. *Clin Lab Sci* 1995;8:277-83.
 16. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.