

Il metodo di Bland e Altman per confrontare metodi di misurazione

Luisa Zanolla
Università di Verona

ABSTRACT

The Bland and Altman method to compare measurement methods

The Bland and Altman (B&A) approach is a graph-based technique to assess the agreement between two measurement methods. The great value of this method is that the assessment is not based on statistical inference. Instead, it defines limits of agreement between the two measurements: it is left to the expert user to evaluate if the concordance is good enough to allow the measurements to be used interchangeably. The graphic presentation of the B&A method, namely the difference between the two measurements plotted over their average, assumes that the limits of agreement are the same over the entire range of values, i.e., that the variability is independent of the measurement magnitude. This assumption, however, is often not valid in clinical chemistry: it can be assessed by regressing the difference of the two measurements over their average. If the correlation is significant, the methods do not agree equally through the range. In this case, a logarithmic transformation could correct the problem. If even this procedure is unsuccessful, one should use more complex statistical techniques. Nonetheless, the B&A method is simple and straightforward, and allows for the assessment of many comparisons. This makes it valuable, and its use should be more widely applied in clinical studies

INTRODUZIONE

È comparso di recente su *Clinical Chemistry and Laboratory Medicine* un articolo di Cesana e Antonelli (1), due docenti italiani di statistica, che affrontano aspetti metodologicamente complessi del metodo di Bland e Altman (B&A). Un grafico di B&A, utilizzato in chimica analitica o in biologia e medicina, è un metodo di rappresentazione grafica dei dati che è alla base di una valutazione di concordanza tra due diversi metodi di misura. Nonostante tale metodo sia considerato uno standard per le valutazioni di concordanza, il suo utilizzo non è così diffuso come sarebbe auspicabile. La motivazione non è sicuramente la complessità statistica: B&A stessi, nel primo lavoro pubblicato sul loro metodo (2), sottolineavano esplicitamente la necessità di un confronto tra metodiche che fosse semplice, soprattutto quando il risultato deve essere presentato ed interpretato da lettori non necessariamente esperti di statistica, quali ad esempio clinici e laboratoristi. B&A si proponevano infatti di fornire un metodo semplice da utilizzare per i

ricercatori medici, che richiede solo software statistico di base (3); il metodo fornisce stime facili da interpretare e nella stessa unità di misura delle osservazioni originali.

In biochimica clinica è spesso necessario confrontare un nuovo metodo con un metodo già in uso, per verificare se la concordanza tra le misure sia tale che il nuovo metodo possa sostituire il vecchio. Una serie di campioni di pazienti viene valutata con entrambi i metodi e i dati comparativi sono sottoposti ad analisi statistica. I calcoli statistici non sono in grado di fornire risposte sull'accettabilità di un metodo; tuttavia, possono fornire stime specifiche sul tipo e sull'entità degli errori.

Sembra quindi tutt'ora necessario segnalare l'utilità del metodo di B&A per il confronto di metodi in *Biochimica Clinica* e auspicare un suo più diffuso utilizzo. Considerato inoltre che l'articolo recentemente pubblicato (1) non sembra adatto a questo scopo, per la sua trattazione di aspetti statistici complessi, questa Opinione si propone di rivedere i principi dell'analisi di B&A, e gli aspetti metodologici cui prestare attenzione.

Corrispondenza a: Luisa Zanolla Via Albere 73 37138 Verona, E-Mail: luisa.zanolla@univr.it

Ricevuto: 02/07/2024

Rivisto: 29/08/2024

Accettato: 29/08/2024

Published on-line: 22.10.2024

DOI: 10.19186/BC_2024.047

IL METODO DI BLAND E ALTMAN

La concordanza tra due metodi di misurazione clinica o laboratoristica può essere valutata utilizzando le differenze tra osservazioni effettuate applicando i due metodi sugli stessi soggetti.

Confrontando due metodi non possiamo in genere considerare nessuno di questi come il vero valore della quantità misurata, a meno della situazione rara e metodologicamente non scevra da problemi, in cui uno dei due metodi sia un metodo di riferimento; l'interesse è nel capire se i metodi diano risposte che siano, in un certo senso, comparabili. Un esempio classico è valutare se un metodo nuovo, più vantaggioso di quello in uso (più economico, rapido o sicuro) produca risposte che concordinino sufficientemente, per scopi clinici, con quelle di un metodo consolidato. Se i valori ottenuti con il nuovo metodo concordano sufficientemente bene con il vecchio, può essere effettuata la sostituzione. La questione non è quindi se i due metodi concordinino, ma quanto concordano.

Già nel primo articolo del 1986 (4) B&A argomentarono contro l'uso diffuso di valutare la concordanza utilizzando il coefficiente di correlazione lineare per questo scopo; peraltro, anche l'uso della regressione, nelle varie forme statistiche, non rappresenta strettamente uno strumento di valutazione della concordanza (5).

B&A avevano presentato l'approccio a studi di concordanza che da loro prende il nome, già nel 1983, in una rivista di statistica (2), ma fu la pubblicazione su Lancet nel 1986 (4) che portò l'attenzione sul metodo, risultando uno degli articoli più citati (56159 volte a inizio 2024) (6).

L'uso della tecnica di B&A per la comparazione di metodi analitici è peraltro tra i metodi suggeriti dalle istruzioni per Autori di Biochimica Clinica (7).

Il metodo descritto da B&A (2,4) propone una valutazione della concordanza tra due misure che riconduce i parametri considerati nella scala del clinico che valuta le misure, fornendo una stima sia del bias che della variabilità delle differenze. Il metodo si basa sulla rappresentazione grafica della differenza assoluta tra le due misure (differenza espressa con la stessa unità di misura delle osservazioni), rispetto alla media delle due misure stesse. Vediamo in Figura 1 l'esempio riportato da B&A (4), costituito da misure del picco di flusso espiratorio effettuate con uno spirometro tradizionale e con uno spirometro mini; va innanzitutto sottolineato come ciascuna coppia di misurazioni fosse effettuata in sequenza casuale, accorgimento che andrebbe sempre applicato.

Prima ancora di rappresentare graficamente i dati, è di estrema importanza definire la differenza tra le due misure che risulterebbe clinicamente accettabile. Nell'esempio, se le due misure fornissero una differenza che non superi, ad esempio, i 10 L/min, lo spirometro mini potrebbe venire usato al posto dello spirometro tradizionale, perché una differenza di entità così modesta non risulta suscettibile di influenzare la gestione del paziente. Per contro, se la differenza giungesse a 100 L/min, lo spirometro mini non risulterebbe clinicamente utilizzabile. Ovviamente

la definizione della differenza minima spetta agli esperti del settore. Più spesso, tuttavia, i ricercatori riportano semplicemente i limiti di concordanza senza confrontarli con un intervallo pre-specificato.

Il grafico di B&A prevede in ascissa la media delle due misurazioni, e in ordinata la differenza assoluta tra le due misurazioni stesse. La media delle due misurazioni costituisce la migliore approssimazione, dato che il valore "vero" non è conoscibile (4), salvo quando sia scelto di assumere come tale un metodo di riferimento accettato.

Possiamo definire quantitativamente la mancanza di concordanza calcolando il bias, stimato dalla media delle differenze tra le due misurazioni, corrispondente nell'esempio a -2,12 L/min, che può essere inserito nel grafico di B&A (linea continua in Figura 1). Se non esiste un bias sistematico tra i due metodi, la differenza media dovrebbe essere approssimativamente zero. Se l'intervallo di confidenza al 95% per la media delle differenze non include lo zero, si può affermare che esiste un bias statisticamente significativo tra i metodi; ma tale significatività statistica non costituisce il criterio per sostenere la concordanza, perché è possibile che non possa essere raggiunta per numerosità campionaria scarsa o variabilità elevata. È sempre un giudizio clinico quello che porta a valutare se l'entità del bias sia tollerabile per una applicazione pratica. È probabilmente per tale motivo che l'intervallo di confidenza del bias medio non venga di solito riportato, mentre rappresenta un parametro imprescindibile, poiché descrive i valori possibili del bias su cui basare il giudizio clinico.

In alcuni casi l'intervallo dei valori possibili è estremamente ampio; è stato proposto, in questa situazione (5) di esprimere la differenza tra le due osservazioni come percentuale del loro valore medio. Si evita in tal modo la presenza di valori di differenza molto piccoli per valori medi piccoli, e molto grandi per valori medi grandi.

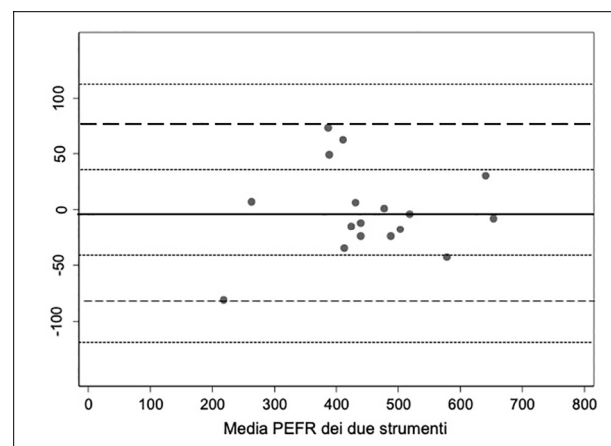


Figura 1

Esempio riportato da B&A (4), costituito da misure del picco di flusso espiratorio (L/min) effettuate con uno spirometro tradizionale e con uno spirometro mini. La linea continua rappresenta il bias (media delle differenze), le linee tratteggiate i limiti di concordanza. Il tratteggio più fitto rappresenta l'intervallo di confidenza dei limiti di concordanza. PEFR, picco di flusso espiratorio

La deviazione standard della differenza, nell'esempio precedentemente citato, è 38,8 L/min; per le proprietà della distribuzione gaussiana ci si attende che il 95% delle differenze si trovi all'interno dell'intervallo tra la differenza media $\pm 1,96$ deviazioni standard. B&A, per semplicità, indicano il valore arrotondato a 2. Questo intervallo è indicato come "limiti di concordanza", che possono essere sovrapposti al grafico di B&A (linee a tratteggio lungo in Figura 1). Anche l'ampiezza dei limiti di concordanza richiede una valutazione clinica in quanto resta il quesito se si tratti di un intervallo di valori accettabile per l'applicazione pratica. Nell'esempio i limiti di concordanza risultano essere da -79,7 a 75,5 L/min. Se ne conclude che le misurazioni effettuate con lo spirometro mini possono essere inferiori di 80 L/min o superiori di 76 L/min rispetto allo spirometro tradizionale, il che indica come lo strumento non sia utilizzabile per finalità cliniche.

Nonostante la formulazione dei limiti di concordanza appaia simile ad un intervallo di confidenza, e spesso vengano confusi, la definizione formale non coincide (nell'intervallo di confidenza si utilizza l'errore standard, non la deviazione standard, e tale grandezza viene moltiplicata per il percentile appropriato della distribuzione t di Student).

Questo tipo di grafico rende molto più semplice valutare l'entità del bias, la dispersione delle differenze ed individuare valori anomali (i cosiddetti outliers). È utile che il grafico abbia la stessa scala sia per ascissa che per ordinata (anche se questo accorgimento non è applicabile quando le differenze sono molto basse).

Il vantaggio del metodo di B&A è che restituisce la valutazione non ad un processo inferenziale statistico, ma all'esperto del dominio: saranno il laboratorista o il clinico a definire, preferibilmente in anticipo, quale sia la differenza che risulti clinicamente rilevante (su criteri biologici, analitici o altre considerazioni) (5) e non consenta quindi di considerare i due metodi concordanti ed intercambiabili. La definizione della differenza consente anche di effettuare una stima della numerosità necessaria per una valutazione della concordanza dotata di adeguata potenza statistica.

Se esiste un bias consistente, è possibile correggerlo, qualora fosse necessario, sottraendo la differenza media dalle misurazioni del nuovo metodo. Nel caso, ad esempio, del confronto tra un metodo a gerarchia più elevata, come un metodo di spettrometria di massa, rispetto ad un metodo immunometrico, anche in presenza di un bias clinicamente rilevante, la considerazione non porta necessariamente all'esclusione dell'uso della nuova metodica, anche se la sua adozione nella pratica quotidiana di laboratorio probabilmente richiede lavoro aggiuntivo come, ad esempio, il calcolo di nuovi valori di riferimento o una diversa calibrazione. In generale è un problema molto più serio e non facilmente correggibile quando si dovesse osservare una notevole ampiezza dei limiti di concordanza.

È importante ricordare che il grafico di B&A rappresenta solo la prima fase del metodo per valutare la concordanza tra due misurazioni.

Si potrebbe obiettare che non è scontato che le

differenze tra due misure seguano una distribuzione gaussiana, ma l'uso della differenza di due variabili esclude larga parte della variabilità tra soggetti; inoltre non è strettamente richiesto che le misurazioni debbano seguire una distribuzione gaussiana. È comunque possibile una ulteriore verifica della distribuzione delle differenze disegnando un istogramma; l'ipotesi di distribuzione non gaussiana, oltre che graficamente, è saggiabile con un test statistico non parametrico: nell'esempio, il test di Shapiro-Wilk non segnala deviazione dalla distribuzione gaussiana ($p=0,614$).

Sono descritti metodi alternativi non parametrici per la stima del bias se la distribuzione devia dalla gaussiana, o per asimmetria o per presenza di code molto pronunciate.

La precisione dei limiti di concordanza

Occorre avere chiaro che il bias ed i limiti di concordanza del metodo di B&A rappresentano una stima campionaria del valore "vero" della popolazione (ovviamente non conoscibile). Lo stesso esperimento condotto su un altro campione di soggetti fornirebbe valori diversi.

I limiti campionari, soprattutto se stimati su un numero limitato di soggetti, possono deviare sensibilmente dal valore "vero".

B&A forniscono indicazioni su un metodo semplificato per calcolare l'intervallo di confidenza dei limiti di concordanza, assumendo che la distribuzione delle differenze non presenti una deviazione significativa dalla gaussiana (4).

L'errore standard dei limiti di concordanza è approssimato dagli Autori a

$$\sqrt{3 \cdot s^2/n}$$

dove s è la deviazione standard delle differenze.

In entrambi i casi occorrerà calcolare il percentile della distribuzione t di Student con (n-1) gradi di libertà (si potrà facilmente recuperare da Excel usando la funzione INV.T.2T(0,05;16) dato i 17 casi di questo esempio, pari a 2,12 e moltiplicarlo per l'errore standard per ottenere la grandezza da aggiungere e togliere dalla stima puntuale, per ottenere l'intervallo di confidenza:

$$\text{Errore standard} = \frac{\sqrt{3 \cdot (38,8)^2}}{17} = 16,3$$

$$\begin{aligned} -79,7 \pm 2,12 \times 16,3 &= -114,3 \text{ a } -45,1 \\ 75,5 \pm 2,12 \times 16,3 &= 40,9 \text{ a } 110,1 \end{aligned}$$

Questi limiti sono ovviamente molto più ampi dei limiti di concordanza, e tendono a fornire un quadro ancora meno ottimistico della concordanza tra i due metodi (la rappresentazione in Figura 1 è con il tratteggio fitto).

È stato quindi sottolineato come i limiti di concordanza non dovrebbero mai essere presentati o interpretati senza i relativi intervalli di confidenza, la cui inclusione dovrebbe diventare una pratica standard nella letteratura (8), mentre nei fatti è ancora estremamente raro vederli riportati. Considerazioni analoghe valgono per il bias.

La relazione tra differenza e media

Le stime del bias e dei limiti di concordanza sono correttamente interpretabili solo se è possibile presumere che essi siano uniformi in tutto il campo di misura, ipotesi che possono essere verificate visivamente nel grafico di B&A, osservando se la dispersione dei dati aumenta o diminuisce al crescere del valore medio. L'osservazione del grafico può consentire di valutare pattern particolari di dispersione delle differenze, quali un andamento diagonale quando i valori crescono o si riducono, al crescere del valore medio, o un pattern a ventaglio quando la dispersione delle differenze aumenta al crescere del valore medio. La differenza tra le due misurazioni non deve dipendere dall'ordine di grandezza (ossia dalla media) e la variabilità, analogamente, non deve dipendere dall'ordine di grandezza. Nell'esempio in Figura 1 si era osservato come le differenze non variassero in maniera sistematica al crescere dei valori in ascissa. È possibile anche verificare l'impressione visiva di un aumento del bias al crescere dei valori, mediante una tecnica statistica molto semplice, la correlazione tra la differenza dei valori e la media dei valori. Nell'esempio la correlazione è pari a $r=0,0837$, molto bassa e del tutto non significativa, confermando l'impressione ottenuta dall'ispezione del grafico.

Diversa la situazione nel grafico di B&A su dati ripresi dal primo lavoro degli Autori (2) relativo a misurazioni della pressione arteriosa ottenuti con due diversi strumenti (Figura 2 – pannello a sinistra).

L'impressione visiva è che le differenze aumentino al crescere dei valori in ascissa: in effetti esiste una correlazione pari a $r=0,4225$ ($p=0,035$). La differenza tra le misure aumenta al crescere dei valori di pressione, per cui i limiti di concordanza risulteranno troppo ampi per bassi di valori di pressione e insufficientemente ampi per valori elevati.

Questo è un aspetto dell'applicazione del metodo di B&A che viene spesso trascurato. Gli Autori sottolineano come sia una situazione che si presenta spesso in

Biochimica Clinica (9). Il primo suggerimento di B&A per risolvere questo problema è una trasformazione dei dati che corregga la proporzionalità tra differenze e medie. I dati vengono trasformati in forma logaritmica, e rappresentati con le stesse modalità, differenza rispetto a media (Figura 2 – pannello a destra).

La media è 0,025 e i limiti di concordanza da -0,012 a 0,066. L'impressione visiva di assenza di correlazione tra differenza e media è confermata dal basso coefficiente di correlazione e dall'assenza di significatività statistica ($r=0,265$, $p=0,199$).

Occorre però a questo punto riportare i limiti di concordanza nella scala di misura originale, calcolando gli antilogaritmi: dato che B&A utilizzano i logaritmi in base 10 (4), si eleva 10 alla potenza costituita dal valore da trasformare, ottenendo 0,972 e 1,164. Tuttavia, l'antilogaritmo della differenza tra due valori in una scala logaritmica è un rapporto adimensionale, per cui possiamo esprimerlo come un intervallo che preveda che nel 95% dei casi il metodo nuovo sarà nell'intervallo tra il 3% in meno e il 16% in più della misura ottenuta con il metodo di riferimento; i limiti di concordanza controtrasformati sono quindi interpretati come un rapporto. L'interpretazione del grafico in forma logaritmica risulta di immediata comprensione aspetto che costituisce una limitazione a tale metodo.

Di fatto gli Autori (4) suggeriscono la trasformazione logaritmica come la sola trasformazione ritenuta suggeribile, in quanto la contro-trasformazione fornisce un dato interpretabile, caratteristica non condivisa da trasformazioni quali radice quadrata o reciproci.

Si può ottenere lo stesso risultato della trasformazione logaritmica lavorando direttamente con i rapporti: invece delle differenze tra le due misure, si calcola il rapporto; quindi, si calcolano bias e limiti di concordanza direttamente su questi (3).

Se la trasformazione logaritmica non aiuta, il metodo di B&A potrebbe non essere appropriato, ma si rendono utili altri metodi.

Gli stessi B&A propongono (3) un diverso metodo

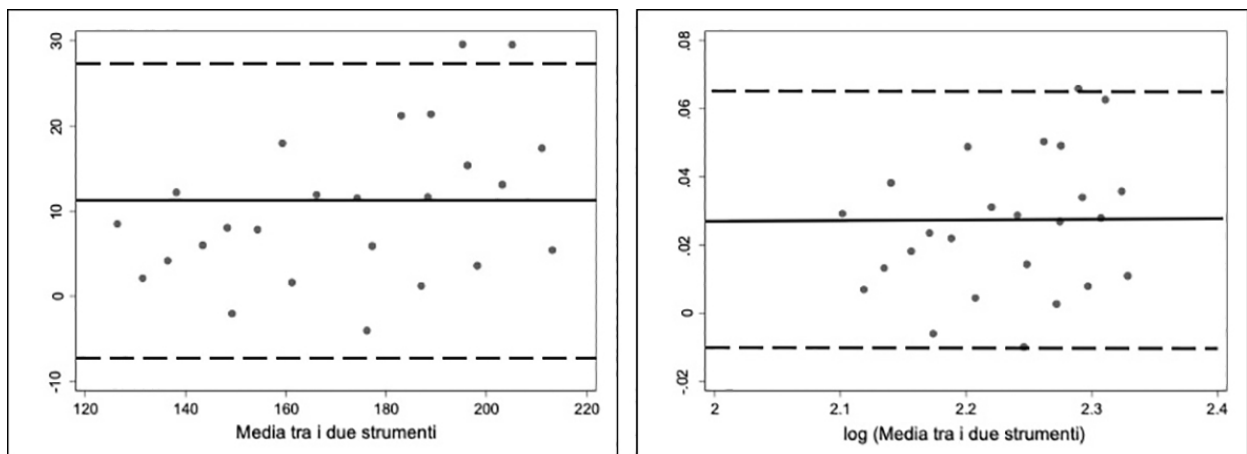


Figura 2

Nel pannello a sinistra l'esempio riportato da B&A (2) relativo a misurazioni della pressione arteriosa (mmHg) ottenuti con due diversi strumenti. La linea continua rappresenta il bias (media delle differenze), le linee tratteggiate i limiti di concordanza. Nel pannello a destra il grafico degli stessi dati dopo trasformazione logaritmica.

statistico per gestire queste situazioni: si tratta di un metodo più complesso, ma conviene conoscerne l'esistenza perché alcuni software statistici generano il grafico secondo tale modello, riconoscibile per il fatto che le rette che definiscono i limiti di concordanza non sono parallele,

Il grafico rispetto alla media versus il metodo standard

Nel confronto tra due metodi ci si trova molto spesso nella situazione in cui non è possibile considerare nessuno di questi come il vero valore della quantità misurata, che rappresenta una grandezza non conoscibile, motivo per cui il metodo di B&A prevede di utilizzare in sua vece la media delle due misurazioni.

È stato tuttavia obiettato che in alcuni casi il metodo di riferimento (S, misurazione al braccio nell'esempio) (Figura 3) rappresenta uno standard consolidato, presumibilmente più accurato rispetto al nuovo metodo proposto (T, misurazione al dito), che chiaramente deve offrire altri vantaggi rispetto allo standard. Sulla base di tali considerazioni è stato proposto di modificare il grafico di B&A, mettendo in ascissa il metodo standard (S) invece della media dei due metodi (10,11). Occorre comunque ricordare che anche il metodo di riferimento non è misurato senza errore. B&A, comunque, si esprimono molto negativamente sulla possibilità di riportare la misura delle differenze a quella considerata standard (12), sia con una formale dimostrazione matematica, che utilizzando la casistica già citata nel primo lavoro (2), ma con una numerosità più ampia (13): la relazione tra differenza media e la media delle due misure ha una pendenza positiva (Figura 3a), e la relazione tra differenza media e misurazione al dito (T) ha pure una pendenza positiva (Figura 3b), mentre la relazione tra differenza media e misurazione al braccio (S) ha una pendenza negativa (Figura 3c). Si conclude quindi, secondo gli Autori, che entrambi i grafici sono fallaci, con differenze statisticamente significative ma in direzioni opposte, a partire dagli stessi dati.

Va inoltre considerato che il valore della misurazione con il metodo standard compare sia in ascissa, che nella differenza in ordinata, tra metodo test e metodo standard; è una situazione che costituisce un noto artefatto matematico; anche in una simulazione con numeri casuali (14), la correlazione tra un valore (S) e la sua differenza con un altro valore (TS) determinava una elevata correlazione, statisticamente significativa ($r=0,74, p<0,0001$).

Uscita di sicurezza: metodi non parametrici

Le differenze tra i metodi non hanno sempre una distribuzione gaussiana: nella maggior parte dei casi l'impatto sulla stima dei limiti di concordanza è modesto. Quando tuttavia la deviazione dalla distribuzione gaussiana diviene rilevante, anche B&A suggeriscono un approccio non parametrico (3). Riprendendo i dati di Figura 3 sulla misurazione della pressione arteriosa con due diversi strumenti, e non assumendo una distribuzione

gaussiana delle differenze, è possibile definire, su base clinica, un limite di accettabilità (ad esempio ± 10 mmHg rispetto alla differenza media) e calcolare la proporzione di differenze che escono dai limiti di concordanza così definiti. Nell'esempio citato, 6 misure sono fuori dall'intervallo, corrispondenti al 24% delle coppie di misurazioni. Gli Autori riportano che questo metodo, pur nella sua semplicità, è stato adottato dal protocollo della British Hypertension Society per la valutazione di

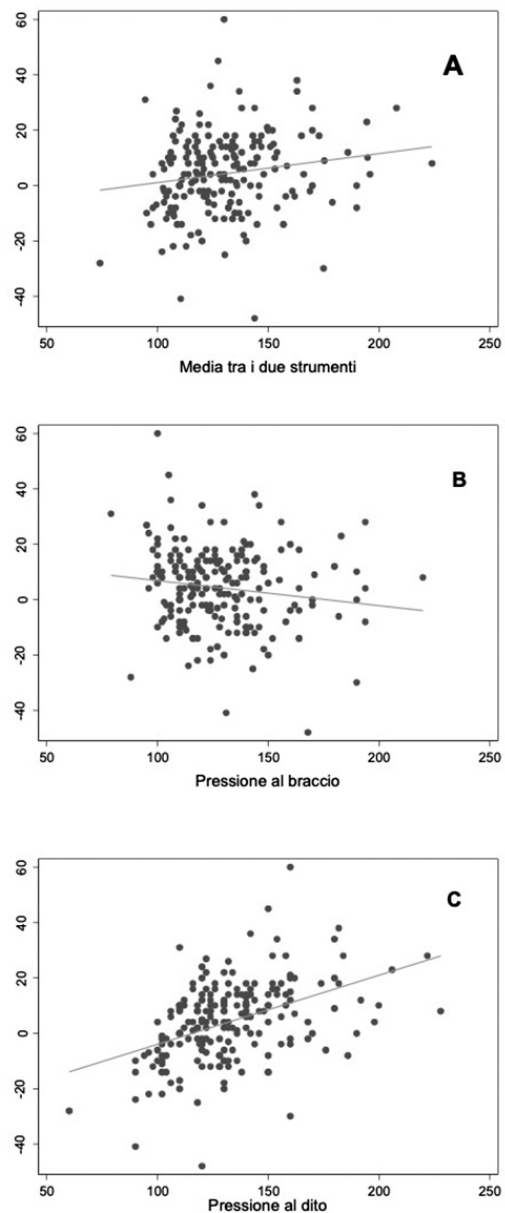


Figura 3
Esempio riportato da B&A (12) costituito da misure di pressione arteriosa ottenuta al braccio e al dito (mmHg). Il pannello A rappresenta il grafico classico di B&A, con in ascissa la media delle due misurazioni. Nel pannello B in ascissa la misurazione ottenuta al braccio. Nel pannello C la misura ottenuta al dito.

apparecchi semiautomatici per la misura della pressione arteriosa.

Una metodica alternativa, sempre secondo B&A, può essere costituita dal definire limiti di concordanza tali che una predefinita percentuale di differenze (ad esempio il 10%) cada al di fuori. B&A suggeriscono semplicemente di ordinare le osservazioni e prendere l'intervallo di valori rimanenti dopo che una data percentuale (nell'esempio il 5%) del campione viene rimossa da ciascuna estremità (3).

Se invece si vogliono calcolare limiti di confidenza, occorre ricorrere al metodo del bootstrap. Il metodo non parametrico è estremamente semplice, ma è stato usato solo raramente. È possibile che proprio per la sua estrema semplicità fornisca l'impressione di non costituire una valida metodica di analisi.

CONCLUSIONI

Il lavoro di Cesana et al. (1) ha il grande pregio di riportare l'attenzione sulle problematiche metodologiche della valutazione di concordanza. Il metodo di B&A ha grande popolarità, ma Cesana et al. ricordano al lettore come il grafico per cui il metodo di B&A è noto, non esaurisce l'analisi sulla complessità del problema.

Ma nella pratica clinica un approccio semplice all'analisi, un metodo "semplice da fare e da interpretare" (4) come auspicato da B&A, può essere comunque il modo più rivelatore di guardare ai dati (2).

I passi da seguire sono di qui seguito schematizzati.

- Definire quale sia una differenza clinicamente accettabile, che consenta di valutare i due metodi come intercambiabili.
- Calcolare sia la differenza tra le misurazioni effettuate con i due metodi a confronto che la loro media; costruire quindi il grafico con la media in ascissa e le differenze in ordinata; calcolare poi i limiti di concordanza. L'esame del grafico consente anche di valutare come le singole differenze si distribuiscono rispetto al bias medio. La media delle differenze ed i limiti di concordanza vengono quindi aggiunti al grafico; l'ispezione visiva dello stesso indicherà poi se vi sia una differenza della variabilità al crescere del valore della media. Tale impressione può essere verificata calcolando la correlazione tra differenze e medie. Se la correlazione presenta un basso valore del coefficiente e risulta non significativa, si completa il grafico riportando gli intervalli di confidenza dei limiti di concordanza.
- Se la correlazione risulta elevata e significativa, provare una trasformazione logaritmica dei valori delle misurazioni ripetendo la procedura con i valori trasformati. Se la correlazione tra differenze e medie nei valori trasformati risulta bassa e non significativa, si può operare la contro-trasformazione, avendo chiaro che i valori non sono più nella scala di misura originale, ma sono grandezze adimensionali.
- Se la correlazione risulta ancora elevata e significativa, occorrerà passare a metodi non parametrici o a metodologie di calcolo più complesse.

Nella larga maggioranza dei casi i passaggi al secondo punto esauriranno la valutazione, costituendo l'approccio semplice che B&A hanno proposto ed auspicato, e che nonostante i suoi pregi stenta ancora a trovare una presenza dominante nella letteratura scientifica.

BIBLIOGRAFIA

1. Cesana BM, Antonelli P. Bland and Altman agreement method: to plot differences against means or differences against standard? An endless tale? *Clin Chem Lab Med.* 2023;62:262-9.
2. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *J R Stat Soc Ser D: Statistician* 1983;32:307-17.
3. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135-60.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-10.
5. Giavarina D. Understanding Bland Altman analysis. *Biochem Med* 2015;25:141-51.
6. Google Scholar – (ultimo accesso: maggio 2024)
7. <https://biochimicaclinica.it/norme-autori/> (ultimo accesso: maggio 2024)
8. Hamilton C, Stamey J. Using Bland-Altman to assess agreement between two medical devices--don't forget the confidence intervals! *J Clin Monit Comput* 2007;21:331-3.
9. Altman DG, Bland JM. Commentary on quantifying agreement between two methods of measurement. *Clin Chem* 2002;48:801-2.
10. International Committee for Standardization in Haematology (ISCH). Protocol for evaluation of automated blood cell counters. *Clin Lab Haematol* 1984;6:69-84.
11. Kringle RO. Statistical procedures. In: Burtis CA, Ashwood ER, eds. *Textbook of clinical chemistry*. 2nd ed. Philadelphia: WB Saunders, 1994: 384-453.
12. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995;346:1085-7.
13. <https://www-users.york.ac.uk/~mb55/datasets/fingsys.dct> - (ultimo accesso: maggio 2024)
14. Gill JS, Beevers DG, Zezulka AV, Davies P. Relation between initial blood pressure and its fall with treatment. *Lancet* 1985;325:567-9.