

Concordance entre deux méthodes de mesure d'une même variable : diagramme de Bland et Altman

Agreement between Two Methods of Measurement: Bland and Altman Plot

Y. Freund

Reçu le 12 novembre 2015 ; accepté le 28 janvier 2016
© SFMU et Lavoisier SAS 2016

Introduction

Une même grandeur peut être mesurée ou estimée de manières différentes. Par exemple, la pression artérielle systolique peut être mesurée de manière invasive en continu, ou par une mesure non invasive automatique par un tensiomètre. La mesure invasive est ici la méthode de référence et procurera les mesures les plus fiables. Pour pouvoir la remplacer par un tensiomètre, il faut s'assurer de la concordance des deux résultats. Autre exemple pour la mesure de l'hémoglobine plasmatique : avant d'utiliser un système de mesure capillaire, il faut s'assurer de la concordance des valeurs retrouvées avec la méthode de référence. L'objet de cet article est de donner les bases de compréhension à l'analyse de concordance entre deux méthodes, les critères à en attendre, et de détailler la méthode de Bland et Altman pour y répondre.

Concordance entre deux mesures

En 1986, Bland et Altman publient un article détaillant les écueils habituels dans ces analyses et proposant une solution simple et graphique pour évaluer la concordance entre deux mesures d'une même grandeur [1]. Brièvement, ils rappellent que la méthode trop souvent employée est celle du calcul de la corrélation. Cette méthode est inadaptée à ce problème particulier et ne permet pas de répondre à la question. La corrélation entre une variable et elle-même ne peut permettre d'évaluer la fiabilité des méthodes de mesure, et en

aucun cas une corrélation parfaite ne se traduira cliniquement par une substituabilité totale des deux méthodes.

L'évaluation de la fiabilité d'une nouvelle méthode de mesure par rapport à une méthode de référence peut se faire de différentes manières et doit répondre à plusieurs impératifs. Pour ce faire, il convient d'étudier la justesse et la précision de la nouvelle méthode de mesure. Prenons comme exemple le tir sur cible (Fig. 1). Une bonne justesse (traduction du terme consacré *accuracy*) revient à être proche de la cible, et une bonne précision correspond à une dispersion faible des valeurs. L'approche de Bland et Altman consiste à analyser et représenter graphiquement ces deux notions.

Le biais, pour évaluer la justesse

Le biais est la différence moyenne entre les deux mesures. Il est évident que pour que la nouvelle méthode soit fiable, le biais doit être le plus petit possible. Il n'existe pas de valeur maximale autorisée du biais, celle-ci doit s'estimer au cas par cas, selon la variable mesurée et la situation clinique envisagée. Par exemple, si on cherche à évaluer la fiabilité de la clairance de la créatinine calculée selon la formule de Cockcroft et Gault, le biais acceptable dépendra de la situation clinique envisagée. En effet, il s'agit de détecter la présence d'une insuffisance rénale parmi une population saine, un biais de 20 ml/min par exemple sera largement acceptable. En revanche, si on s'intéresse à une population de patients avec une insuffisance rénale terminale, un biais de 20 ml/min sera beaucoup trop important. Le biais étant la différence moyenne entre les deux mesures sur la population, il se calcule facilement, mais son utilité brute est faible. Un biais trop élevé remettra en question la nouvelle méthode, mais un biais très faible ne pourra pas la valider. En effet, la nouvelle méthode peut régulièrement surestimer et sous-estimer les valeurs de la grandeur mesurée, mais avec une moyenne nulle. La détermination de l'écart-type du biais est alors indispensable, de même que sa variation en fonction de la valeur de la mesure. En effet, un biais fixe peut

Y. Freund (✉)

Sorbonne Universités, UPMC Paris univ-06,
47-83, boulevard de l'Hôpital, F-75013 Paris, France
e-mail : Yonathan.freund@aphp.fr

Service d'accueil des urgences, hôpital Pitié-Salpêtrière,
Assistance publique-Hôpitaux de Paris,
47-83, boulevard de l'Hôpital, F-75013 Paris, France

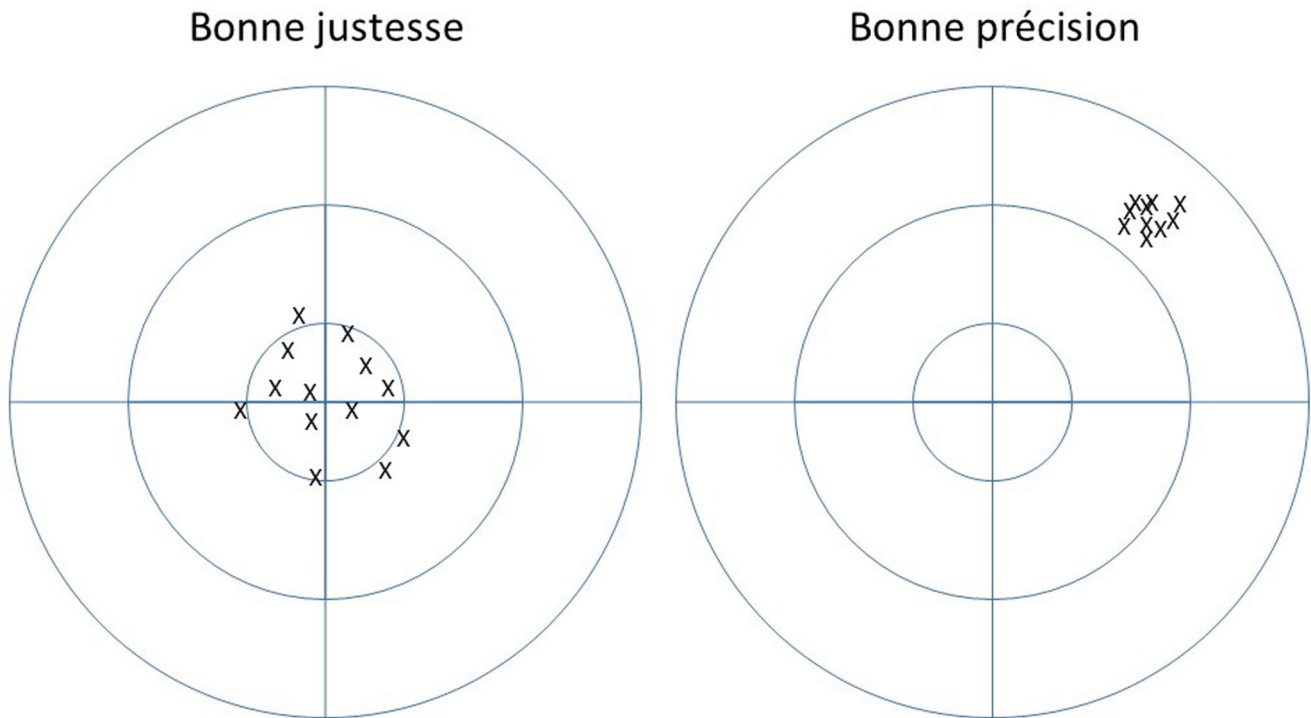


Fig. 1 Cible avec illustration d'une bonne justesse et d'une bonne précision

aisément se corriger à l'aide d'une constante, mais un biais variant avec la mesure sera plus difficile à interpréter et à corriger.

La valeur du biais doit être ramenée à la valeur de référence de la variable étudiée afin d'en évaluer sa pertinence. Ainsi, le biais relatif est égal au rapport entre biais absolu et valeur de référence.

Les limites de concordance, pour évaluer la précision

La précision estime la dispersion du biais absolu et représente le caractère reproductible de la méthode. Son évaluation est indispensable, car un biais nul peut être associé à de fréquents et larges écarts entre les deux mesures. Pour évaluer la précision du biais et la dispersion des valeurs, plusieurs variables peuvent nous apporter une information : l'écart-type, la valeur de certains percentiles, la largeur de l'interquartile... On détaillera ici les limites de concordance à 95 %.

Il s'agit des bornes de l'intervalle de confiance à 95 % du biais absolu et se détermine comme étant la valeur du biais, plus ou moins le produit $1,96 \times$ (écart-type). Si la distribution des valeurs suit la loi normale, 95 % des différences se trouveront dans cet intervalle. Notons que la précision peut être estimée en calculant la moyenne des valeurs absolues des différences.

Représentation graphique

Le diagramme de Bland et Altman permet une représentation graphique de ces données. L'ordonnée (axe Y) représente la différence entre les valeurs de la méthode de référence et la méthode testée. L'abscisse (axe X) représente la moyenne de ces deux valeurs. Ainsi, graphiquement, toutes les mesures d'un échantillon sont reproduites sur ce graphique, auquel on ajoute une droite représentant le biais ($y = \text{biais absolu}$) et deux droites pointillées représentant les limites de concordance ($y = \text{biais absolu} \pm 1,96 \times \text{écart-type}$).

Ainsi, en regardant un tel graphique, les informations nécessaires et précédemment décrites sont visibles :

- le biais absolu est représenté, et on évaluera donc la justesse de l'analyse ;
- la dispersion est représentée par l'écart entre les deux droites de limites de concordance. Si celles-ci sont proches, la précision est bonne. Si elles sont éloignées, elle l'est moins. On pourra même y faire figurer l'intervalle de confiance des limites de concordance, afin d'évaluer la fiabilité et reproductibilité de ces paramètres ;
- la répartition des valeurs en outre principalement dans l'intervalle de concordance confirme la normalité de la loi suivie ;
- les valeurs extrêmes sont facilement vues ;
- enfin, la tendance pour des moyennes plus fortes peut donner une information sur les « zones » dans lesquelles

la précision est meilleure. Les mesures peuvent être très justes pour des valeurs basses, et se disperser pour des valeurs plus élevées.

Illustrations

Sur la Figure 2 sont illustrés quatre exemples (Fig. 2A,B,C, D).

Sur la Figure 2A, on voit l'exemple d'une concordance excellente. En effet, on voit que le biais est quasi nul, et que les limites de concordance à 95 % sont très proches. Ainsi, il apparaît visuellement que la différence entre les deux mesures pour chaque point (la distance par rapport à l'axe $y = 0$) est très faible, et ce quelle que soit la valeur de la variable mesurée. *Une attention particulière doit être portée au choix de l'échelle de l'axe Y. En effet, avec une échelle très large, les limites de concordance peuvent paraître très proches.*

Sur la Figure 2B, on illustre le cas d'une méthode avec un biais quasi nul, mais une grande dispersion. On voit bien que même si la moyenne des différences est proche de zéro, les mesures diffèrent largement, et l'intervalle entre les limites

de concordance à 95 % est très large. Visuellement, on voit clairement que la méthode testée n'est pas fiable, malgré son excellent biais.

Sur la Figure 2C, le biais n'est pas négligeable, mais les différences ne varient pas beaucoup. L'intervalle entre les limites de concordance est très étroit. Ainsi, on a ici le cas d'une méthode avec un biais réel, mais fiable. En effet, on peut imaginer soustraire le biais connu de la valeur obtenue, et ainsi avoir une méthode fiable, avec un biais nul et des limites de concordance à 95 % proches. On peut dire que la méthode testée est fiable, mais mal calibrée.

Enfin, la Figure 2D illustre le cas non rare d'une méthode avec un biais nul par rapport à la méthode de référence, et une allure du diagramme de Bland et Altman en V couché. Cela traduit une perte de précision avec la croissance des valeurs. On pourra en conclure que la méthode testée est fiable jusqu'à une certaine limite.

Interprétation

Pour évaluer la pertinence et l'intérêt d'une méthode de mesure de substitution, il faut tout d'abord évaluer ce

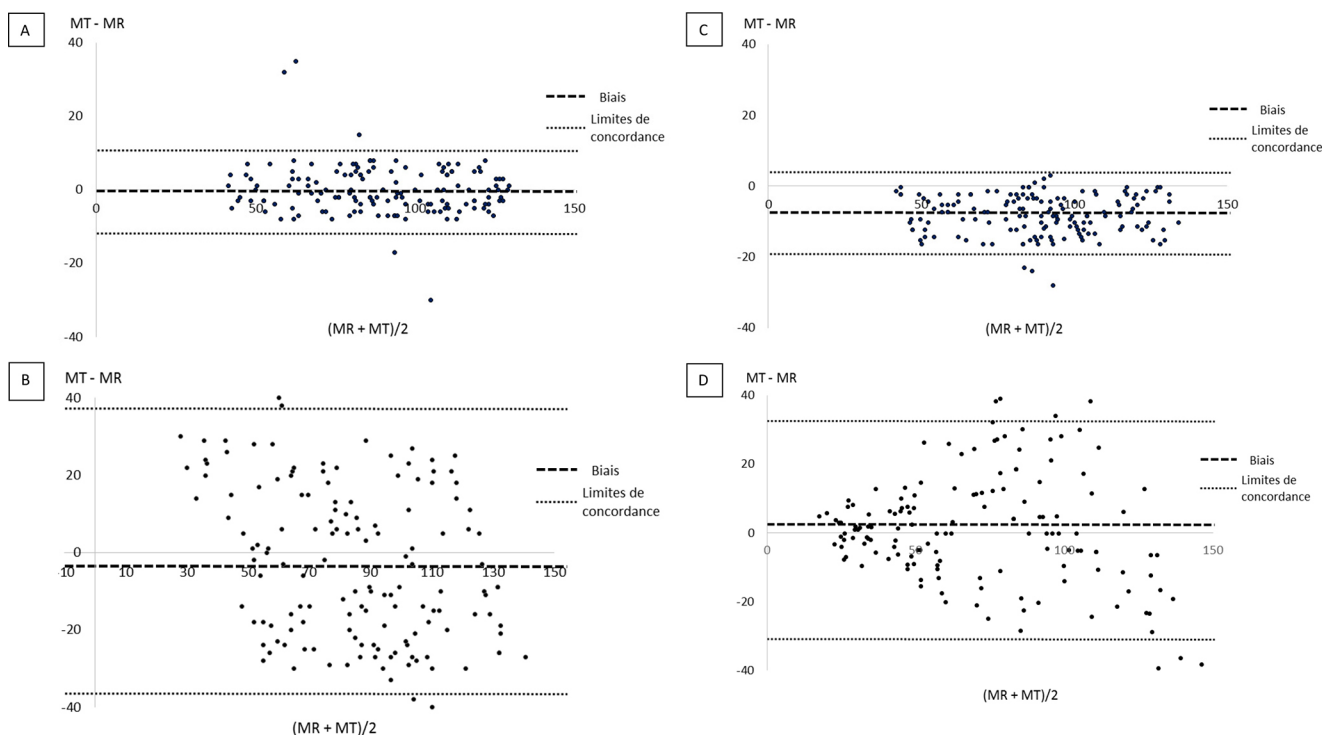


Fig. 2 Quatre exemples de diagramme de Bland et Altman. MR : méthode de référence ; MT : méthode testée. A. Biais = 0,5 – limites de concordance = ± 10 . Biais parfait et bonne précision, la méthode évaluée est une bonne méthode de substitution. B. Biais = 0,5 – limites de concordance = ± 37 . Biais parfait, mais limites de concordance très larges. Méthode imprécise ne pouvant pas se substituer à la méthode de référence. C. Biais = -9 – limites de concordance = ± 11 . Mauvais biais, méthode mal calibrée, mais bonne précision. À utiliser en incorporant le biais dans l'interprétation. D. Biais = 2,3 – limites de concordance = ± 37 . Limites de concordance larges, mais aspect en V couché. Méthode intéressante pour les valeurs basses

qu'on considère comme cliniquement pertinent. La taille de l'intervalle entre les limites de concordance est importante, mais la proportion de mesures supérieures (ou inférieures) à une limite cliniquement pertinente l'est aussi. Par exemple, si on veut évaluer la fiabilité d'une méthode de dosage du lactate capillaire par rapport à la mesure sur gaz du sang artériel, il faut se demander avant l'étude ce que l'on considère comme acceptable. La réponse n'est pas toujours univocale et souvent sujette à discussion. Pour cet exemple, on peut par exemple dire que des mesures qui diffèrent de plus de 0,5 mmol/l sont aberrantes. Mais on pourra tout aussi bien justifier que la barre doit être plus stricte, et que des limites de concordance à 95 % qui seraient à $\pm 0,5$ mmol/l sont trop larges. On pourra recommander de fixer le seuil qui paraît raisonnable à l'avance, et le comparer aux limites de concordance obtenues après analyse [2].

On pourra suggérer une analyse en sept points devant la lecture d'un diagramme de Bland et Altman :1)

quel est le paramètre mesuré ?2)

Quelle est la technique de mesure de référence et quelle est la technique de mesure évaluée ?3)

Quelle est la valeur maximale des limites d'agrément au-delà de laquelle le clinicien ne peut plus considérer que les deux mesures sont concordantes ?4)

Les différences sont-elles réparties de façon gaussienne ?5)

Quelle est la valeur du biais ? Est-il stable sur l'intervalle de mesure ?6)

Quelles sont les valeurs des limites de concordance ? Excèdent-elles celles définies au point 3 ? Ces limites s'évasent-elles vers les valeurs basses ou hautes de mesure ?7)

Quelle est la précision du biais et des limites de concordance (leur intervalle de confiance à 95 %) ?

Pour aller plus loin

Quelques subtilités dans l'analyse de Bland et Altman méritent d'être évoquées.

La différence observée entre les deux mesures peut provenir d'une imprécision de la méthode de référence et non seulement de la méthode testée. Ainsi, la précision de la méthode de référence doit être connue et devrait contrebalancer d'éventuels mauvais résultats, et le choix de la méthode de référence doit se faire de manière extrêmement précautionneuse. Critchley et Critchley ont proposé une méthode alter-

native pour s'affranchir de cette limite, laquelle est aussi remise en cause dans certains cas... [3,4].

Une attention particulière doit être posée sur la faisabilité de la nouvelle méthode et donc des potentielles valeurs manquantes. Si une méthode de substitution a un biais et une précision intéressants, mais qu'elle s'avère incapable de donner une mesure dans un nombre substantiel de cas, alors son intérêt peut s'en trouver très limité [4]. Le nombre de valeurs manquantes ou mesures impossibles est donc une variable primordiale à analyser, dont le report est nécessaire.

En cas de mesures multiples avec la mesure de substitution, l'estimation du biais pourrait être plus précise, mais les limites de concordance se trouveraient artificiellement étroites. Myles et Cui ont proposé une méthode pour prendre en compte les mesures multiples, qu'on laissera le soin au lecteur intrépide de découvrir [5].

Points clés

- Le diagramme de Bland et Altman permet d'évaluer graphiquement si pour une variable donnée, une nouvelle méthode de mesure peut remplacer la méthode de référence ;
- le biais est la différence moyenne entre les deux méthodes ;
- la précision est estimée par les limites de concordance à 95 %.

Il est indispensable de préciser quelle différence est acceptable cliniquement.

Références

1. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327:307–10
2. Mantha S, Roizen MF, Fleisher LA, et al (2000) Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesth Analg* 90:593–602
3. Critchley LA, Critchley JA (1999) A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. *J Clin Monit Comput* 15:85–91
4. Le Manach Y, Collins GS (2015) Disagreement between cardiac output measurement devices: which device is the gold standard? *Br J Anaesth* [Epub ahead of print]
5. Myles PS, Cui J (2007) Using the Bland-Altman method to measure agreement with re-peated measures. *Br J Anaesth* 99:309–11