A new statistical methodology overcame the defects of the Bland & Altman method

Short title: **A new statistical methodology to assess bias and precision**

Patrick Taffé, Patricia Halfon, and Matthieu Halfon

Center for Primary Care and Public Health, University of Lausanne, Switzerland (P Taffé PhD, P Halfon MD)
Division of Nephrology and Hypertension, Lausanne University Hospital, Lausanne, Switzerland (M Halfon MD)
Transplantation Center, Lausanne University Hospital, Lausanne, Switzerland (M Halfon MD)

Correspondence to: Dr Patrick Taffé, Center for Primary Care and Public Health, University of Lausanne, Division of Biostatistics, Biopôle 2, Route de la Corniche 10, 1010 Lausanne, Switzerland
patrick.taffe@unisante.ch

Conflict of interest: none

Abstract

Background and Objectives: The Bland and Altman‘s limits of agreement (LoA) method is the most commonly used statistical method to assess bias and precision of a new measuring device (it has been cited over 40’000 times as of March 2019). What is less known is that the LoA method can be dramatically misleading.

Methods: A new statistical methodology, which circumvent these deficiencies, has recently been published and made available in the R and Stata statistical packages. We aimed at introducing and illustrating with a small data set on blood pressure (BP) measurements, taken by two different oscillometric devices, the use of this new methodology to a clinical audience.

Results: For DBP, the LoA method was particularly misleading as it identified differential and proportional biases of opposite signs compared to the new methodology. Regarding SBP, the LoA method strongly overestimated both the differential and proportional biases, for both devices.

Conclusion: The LoA method may be dramatically misleading and does not allow one to estimate the precision of each measurement method. We recommend the use of the newly developed statistical methodology instead.

Keywords: Bias, precision, limits of agreement, blood pressure, oscillometric device.

What is new ?

Background
The widely used Bland and Altman limits of agreement (LoA) method to assess bias and precision of a new measuring device is challenged.

Key findings
It is shown, based a small data set on blood pressure measurements, that the LoA method can be dramatically misleading.

What this adds to what was known ?
A new statistical methodology, which resolves these issues, has recently been published in the statistical literature and made available in the R and Stata statistical packages. However, due to its technicality this paper may not have attracted clinician's attention.

What is the implication and what should change now ?
We recommend clinicians and researchers to use the newly developed statistical methodology.

Introduction

The Bland and Altman's limits of agreement (LoA) method is arguably one of the most widely used statistical tool in medical research to assess bias and precision of a measuring device with respect to a reference standard (the 1986 paper, published in the Lancet journal [1], has been cited over 40'000 times as of March 2019, google search). However, what is less known is that the LoA method can be dramatically misleading. Indeed, there are settings where the LoA method shows a positive or a negative bias and there is no genuine bias, whereas in others despite an apparently zero bias there is genuinely a bias [2]. In addition, it allows one to estimate the precision of the differences but not of each measurement method separately.

In its original formulation [1,3], the Bland and Altman method allows one to estimate the average bias and assess the agreement between two devices (i.e. the precision of the differences) by considering the width of the two parallel limits of agreement lines. This approach, however, relies on two restrictive assumptions:

- First, it assumes that the bias between the device and the standard is constant (therefore, the average of the differences is computed). However, there are no good reasons to make this assumption. Actually, the reverse is more likely to be true in clinical practice with a possibly a negative bias in a certain range and a positive bias in another [4]. To see this the bias need to be decomposed into a differential and a proportional bias.

- Second, it assumes that the variability of the measurement errors is uniform throughout the whole range of the measurements. Again, this assumption often turns out to be unrealistic and contradicted by empirical results. Indeed, empirical results often confirm the biological intuition that whenever the level of the latent trait is low measurement errors are low, whereas when it is high measurement errors are higher [4].

In their 1999 paper [5], Bland & Altman extended their 1983/1986 methodology to account for non-constant bias by regressing the differences on the averages of the two measures. The obtained regression line was supposed to illustrate the amplitude and sign of the bias. From the coefficient estimates (intercept and slope), one can derive the differential and the proportional biases (they are transformations of the intercept and slope). However, as mentioned above these estimates are biased and may be misleading. They also developed a methodology to take into account non-uniform variability of the measurement errors (with blood pressure the variability of the measurement errors is expected to increase with larger values), thereby producing non-parallel (i.e. oblique) LoA lines. However, it allows one to assess the variability of the differences, but not of each device separately.

A recently developed statistical method corrects for the deficiencies of the Bland & Altman method, and should be used instead [2]. While this new approach is based on mathematical models, interpretation of the results is made relatively simple by producing two new plots, the "bias" and "precision" plots. These plots should replace the standard LoA plot, and allow the researcher and clinician to easily quantify the amount of bias of the new measurement method and assess the precision of each measurement method.

We aimed at introducing and illustrating with a small data set on blood pressure (BP) measurements the use of this new methodology to a clinical audience.

Methods

Data

As an applied example, and to illustrate the differences between the two statistical approaches, we used the data from a study conducted at the Lausanne University Hospital in Switzerland approved by the Vaud cantonal ethics committee [6]. Briefly, eligible subjects were recruited during their hospitalization in the intensive care unit. Inclusion criteria were having AF at the time of measurement and an arterial indwelling catheter allowing the monitoring of invasive arterial blood pressure (IBP). Exclusion criteria were the presence of a pacemaker, of an arteria-venous fistula, signs of infection at the site of BP measurement, and age <18 years old. Two commonly used oscillometric blood pressure devices were evaluated: the Omron HEM907 (OHEM) and the microlife watchHomeBP (WBP).

For each participant, two separate sequences of pairs of BP measurements were performed (OHEM+IBP and WBP+IBP), resulting in 10 repeated pairs of BP measurements per individual, for each device. The two sequences were randomized (i.e., which pair of BP devices was to be used first) and separated by a 5 minutes pause. If a device failed to indicate a BP value twice consecutively, the measurement was stopped.

Statistical Analyses

We compared the estimates of bias and precision obtained for each oscillometric device (OHEM and WBP), against IBP measurements, by the "standard" (B&A) [1,3] and "extended" (eB&A) Bland and Altman methods [5], and the new statistical method developed by Taffé (Ta) [2,7].

The B&A method estimates only the average bias, whereas the eB&A and Ta methods allow the bias to be non-constant and depend on the level of true BP. For this, it is useful to decompose the bias (i.e. the systematic difference between the measured SBP and the true latent or unobserved

SBP) into two components, the differential and proportional biases (see Appendix B for the details). For estimating the true SBP, Bland and Altman use the mean of the two measurements given by the test device and the reference, which assumes that they are equally valuable estimates of the true SBP. Taffé uses an empirical Bayes method to computes the Best Linear Unbiased Prediction of the latent trait (BLUP). The BLUP uses for each subject the reference measurements of that subject to estimate his or her true value and additionally borrows information from the reference measurements of the other subjects in the sample. Using only reference measurements makes sense, as the aim is to assess how far the test device lies from the reference, i.e. the reference is assumed unbiased (Appendix B).

The mean is simple to compute but unfortunately provides a biased estimate, whereas the empirical Bayes method performs much better and allows one to get an unbiased estimate.

Results regarding the bias are graphically represented in the standard LoA plot and in the newly developed Bias plot. The Ta method provides an additional plot called precision plot. It is a plot with the standard deviation of the measurement's errors of the device on the y-axis and the true blood pressure (i.e. BLUP of x) on the x-axis.

Results

Detailed characteristics of the ten subjects studied are provided in Appendix A.

Devices' bias

All the bias estimates for the two devices (WBP and OHEM) by each of the three methods (B&A, eB&A, and Ta) are given in Table 1. For brevity, graphical results are presented in the main text only for the WBP device and the reader is referred to appendix C for graphical results regarding the OHEM device.

Classical Bland-Altman method: Starting with the classical B&A method (Figure 1):

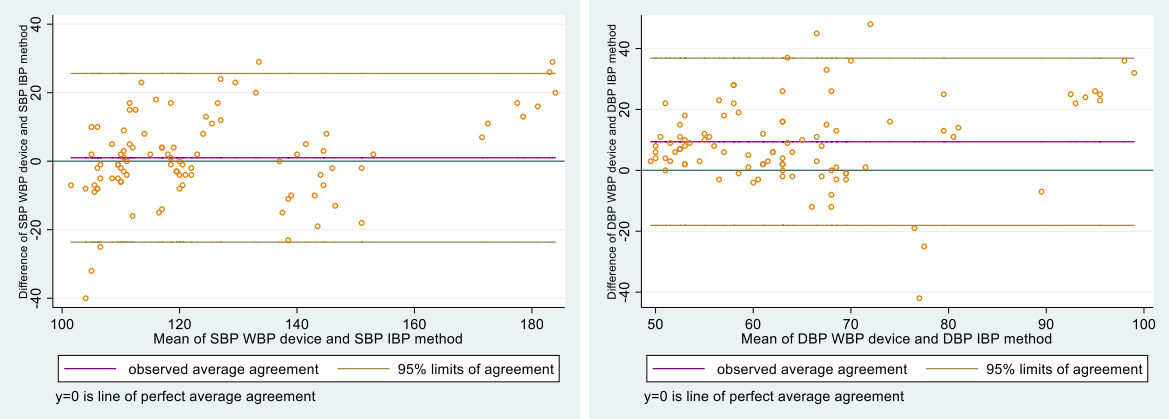

Figure 1 Classical Bland & Altman' LoA plot for SBP (left) and DBP (right) measured by the WBP device

By the standard B&A method, the estimate of the average bias (violet horizontal line) is 1 mmHg for SBP and 9.4 mmHg for DBP.

Extended Bland-Altman method: However, using the eB&A method shows that the bias (red dash-dotted regression line) increases with the level of BP for both SBP and DBP (Figure 2):

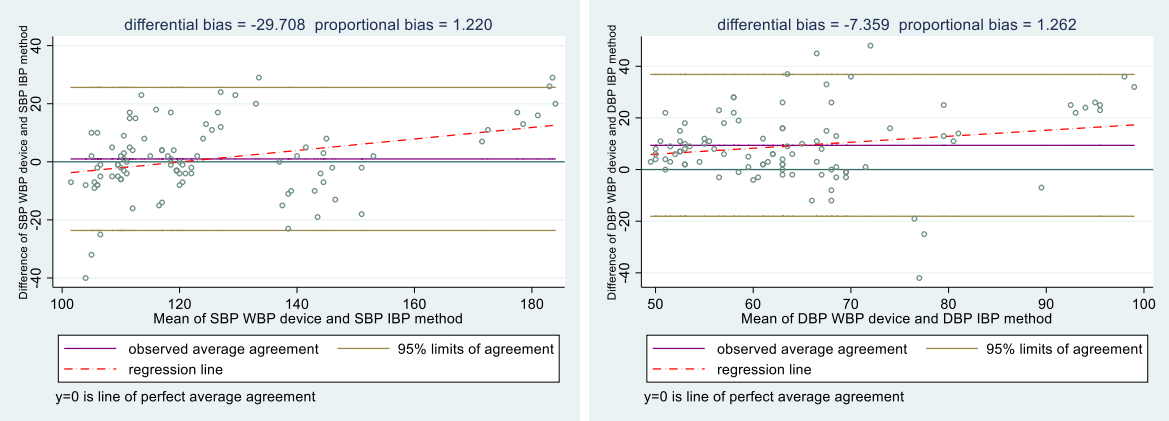

Figure 2 Extended Bland & Altman' LoA plot for SBP (left) and DBP (right) measured by the WBP device
By adding on the plots the regression line of the differences versus the means (of the two measurements), one may have the impression that the bias is increasing with the level of BP (red dash-dotted regression line). The estimates of the differential and proportional biases are provided on top of each figure.

Taffé method: Using the Ta method, the amount of bias (red dash-dotted regression line) also depends on the level of true BP (Figure 3):

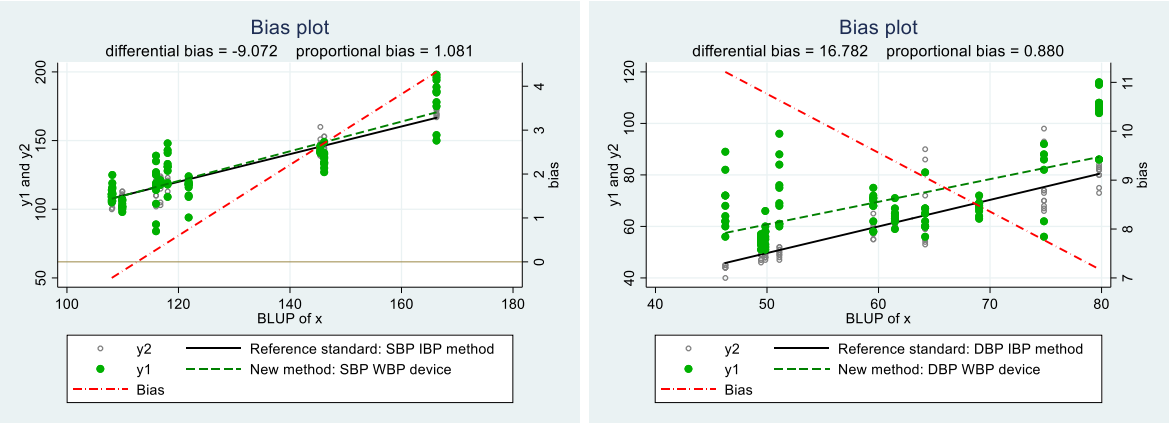

Figure 3 Bias plot for SBP (left) and DBP (right) measured by the WBP device
Estimates of the differential and proportional biases are given on top of each figure. The estimated bias (red dash-dotted regression line) is increasing with the level of true BP (i.e. BLUP of x) for SBP, whereas it is decreasing for DBP (the amount of bias in mmHG can be read from the right y-axis).

However, while for SBP (left figures 2&3) both the eB&A and Ta methods find the bias to increase with the level of true BP, for DBP (right figures 2&3) the Ta method provides a completely different picture as with the eB&A the bias increases with the level of true BP whereas it decreases with the Ta method.

To delve into the details, consider the differential and proportional biases separately:

- Differential bias: Actually, there are striking differences between the estimates of the differential bias from the eB&A and Ta methods: -29.7 versus -9.1 for SBP, and -7.4 versus

16.8 for DBP. In this example, for SBP the amount of differential bias is over estimated by the eB&A method, with respect to the Ta, whereas for DBP the direction of the differential bias is wrongly estimated by the eB&A method.

- Proportional bias: Likewise, for the proportional bias the estimates by the eB&A and Ta methods differ substantially: 1.22 versus 1.08 for SBP, and 1.26 versus 0.88 for DBP. For SBP the eB&A method overestimates the proportional bias, with respect to the Ta method, and for DBP the two methods provide proportional bias estimates of opposite directions (1 is the reference and means no bias).

We turn now to the interpretation of the Bias plot. For the sake of clarity, we start by discussing the comparison of the WBP device with the IBP method. Then, we describe in details all the components of the Bias plot.

Comparison of the measurements made by the WBP device and IBP method
The solid black and dashed green regression lines show that for true values of SBP around 110 mmHG, WBP and IBP provide similar values, whereas when the true values are around 165 mmHG the WBP device provides higher values than IBP (about 3 mmHG, as read from the right y-axis, Figure 3 left).

For DBP, the discrepancy between IBP and WBP is higher, the latter providing systematically larger BP values than the former (the discrepancy goes from 7 to 11 mmHG for values of true BP comprised between 46-80, as read from the right y-axis, Figure 3 right).

Detailed interpretation of the Bias plot
Because results are more contrasted, let us focus on the Bias plot for DBP (figure 3, right). The first regression line (solid black), y2 versus BLUP of x, has intercept 0 and slope 1, as it is the reference. The points (hollow circles, gray) scattered around the line illustrate that the reference is subject to measurement errors (the points would be all on the line without measurement errors). The second regression line (dashed green), y1 versus BLUP of x, lies above the first and illustrates that the WBP device exhibits some positive differential bias. However, the distance between the two regression lines reduces as the DBP increases, thereby illustrating that there is also a proportional bias (in the absence of proportional bias, the second regression line would be parallel to the first). Clearly, the (total) bias is not constant and varies with the level of BP.

On the right y-axis, the amount of bias (labeled “bias”) can be read. This is done by selecting a value on the x-axis (for the true DBP), and reading from the dash-dotted line the corresponding value of the bias on the right y-axis. For example, for a DBP of 50 (x-axis) the bias (right y-axis) is about 10.5 mmHg and for a DBP of 80 about 7 mmHg.

When there are few individuals, like here, it is possible to distinguish the repeated measurements of each individual, as they are aligned vertically for each value of the BLUP of x (each value representing the best possible prediction for a specific individual). One can also distinguish the measurements made by each of the two measurement methods (as long as they do not overlap too much), as the symbol used for the reference (hollow circle) is different from that for the new method (point).

We now focus on assessment and comparison of the precision of the two measurement methods.

Precision of the WBP device versus the IBP method
In the classical LoA plot the width between the two limits are quite large [-23.6; 25.6] for SBP and [-18.1; 36.8] for DBP (Figure1). However, these limits are of limited usefulness as they allow one to assess the variability of the differences, but not of each device separately.

Therefore, to get an estimate of the precision of the WBP device and the IBP method, the standard deviation of each measurement method has been computed using the Ta method (Figure 4):

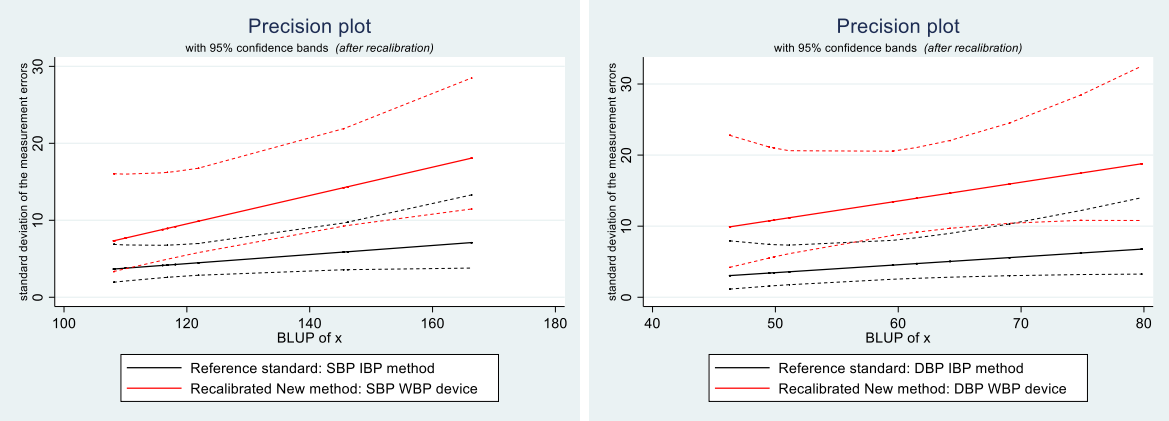

Figure 4 Precision plot for SBP (left) and DBP (right) measured by the WBP device
The x-axis represents the true blood pressure and the y-axis the standard deviation of the measurement errors. The two regression lines, on each of the two plots, clearly illustrate the higher the true blood pressure the higher the measurement errors (i.e. the standard deviation of the measurement error is increasing with the true BP values). It also shows that the precision of the IBP method tends to be better than that of the WBP device, particularly around values of 60-70 mmHG for DBP where the confidence bands do not overlap.

Clearly, as expected, the precision of the WBP oscillometric device and of the IBP method depends on the level of BP values: the standard deviation of the measurement errors increases with increasing blood pressures.

Of clinical importance, the precision of the IBP method tends to be better than that of the WBP device. To allow for a formal comparison, we have added simultaneous confidence bands around the standard deviation lines [7]. For SBP the confidence bands overlap, whereas for DBP they do not between 58 and 68 mmHG, thereby clearly illustrating the better precision of the IBP method over the WBP device. Again, the lack of statistically significant differences over the whole range of BP values is essentially attributable to the very small sample size (only 10 patients), and with many more patients one would expect the simultaneous bands not to overlap at all.

Discussion
Recently in the statistical literature [2], it has been shown that the Bland and Altman LoA method may be misleading and provide estimates of the differential and proportional biases of the wrong sign. As a result, the (total) bias is also biasedly estimated. In addition, the LoA method does not allow one to assess the precision of each measurement method separately. Therefore, a new statistical methodology to assess bias and precision has been proposed.

To illustrate these points and the usefulness of this new statistical methodology, we have used a small data set on BP measurements and applied both the LoA method (standard and extended) and the new Ta method. Our results have confirmed that the Bland and Altman LoA method may be quite misleading, since for DBP it has provided differential and proportional bias estimates of the wrong sign, and for SBP it has strongly over-estimated these parameters. Of note, confidence intervals are wide and overlapping between methods due to extreme scarcity of the data, and do not allow a formal comparison in this dataset.

Regarding the assessment of precision, we have illustrated that with the Bland and Altman methodology one can assess only the precision of the differences but not of each device separately, whereas this is the case with the Ta method. Actually, it has been shown that bias and precision of the two oscillometric devices and of the reference method were not uniform throughout the range of measurements. Using the conventional Bland and Altman methodology would not have highlighted this important fact. This may have clinical consequences, as bias may be acceptable for a certain range of BP but not outside it. Likewise, precision of the instrument may be acceptable only within a specific range.

The main differences between the standard B&A and Ta methods, is that the latter imposes less constraints and allows bias and precision to vary with the true latent blood pressure. Intuitively, this is sound as often when studying a process small values exhibit small variability and large values larger variability. Actually, the conditions for the standard B&A method to be valid and provide unbiased estimates of the differential and proportional biases are very restrictive as they imply that the ratio of the two variances of the measurement error has to be strictly equal to the proportional bias, a very special condition which is unlikely to hold in practice [2]. Therefore, the standard B&A method almost always provides biased estimates and should be abandoned.

The Ta method requires several measurements by one of the two instruments, whereas the classical B&A method can be applied with only one measurement per device. However, as mentioned above, unless the special condition holds the B&A method will provide a biased estimate of the bias. Actually, it has been shown that without repeated measurements it is not possible to separate the differential from the proportional bias [8]. Neither is it possible to estimate the precision of each instrument separately. Therefore, for validating a new measurement method it is mandatory to take several measurements per individual at least with one of the two instruments.

In studies [2] and [7], it has been shown by simulations that the Ta method performed very well to estimate bias and precision with sample sizes of 100 individuals and 10 to 15 repeated measurements from the reference standard, and as few as 1 measurement from the new method. However, when the focus is limited to the estimation of the differential and proportional biases, additional (limited) simulations (results not reported) illustrated that under certain circumstances (e.g. differential bias of -4 and proportional bias of 1.2, see [2] page 5), with as few as 20 individuals and 3 to 5 repeated measurements the estimates of the differential and proportional biases were already in the right order of magnitude (average estimated differential bias = -5.15 and proportional bias = 1.25). Nevertheless, for assessing precision more repeated measurements are required, typically at least 8 to 12 by one of the two instruments. It is difficult to give more precise recommendations regarding the minimum number of individuals and repeated measurements required to get precise estimates, as it depends on the true levels of differential and proportional biases, and on the heteroscedasticity.

The results of this study are of clinical importance. Indeed, BP measuring is the most common procedure of medical physical exams. It is commonly carried out with mercury sphygmomanometers. However, nowadays these devices tend to be supplanted by automatic oscillometric devices because of environmental concerns [9]. Before their use in medical practice, these oscillometric devices must pass a validation process. Currently, two validation protocols are commonly used: the Association for Advancement of Medical Instrumentation (AAMI) and the European Society for Hypertension (ESH) [10,11]. To be validated by the AAMI protocol the mean difference between the test device and the mercury standard must not be larger than 5 mmHg (bias criterion), and the standard deviation of the differences not larger than 8 mmHg (precision criterion) [10]. For the ESH protocol the percentages of readings falling within 5, 10, and 15 mmHg of the mercury standard, must be equal to or greater than specified cut-off values [11]. Therefore, it is clear that the AAMI and ESH validation protocols follow similar rules of assessment as the original Bland and Altman methodology and, consequently, suffer from the same limitations (i.e. a constant bias is assumed, as well as homogeneity of the variance of the differences, thereby not allowing to distinguish differential bias from proportional, neither to assess precision of each measurement method separately), and should be replace by the new methodology.

Given that the Ta method has been shown (by simulations [2]) to provide unbiased estimates of the differential and proportional biases, it is recommended to use this method instead of the Bland & Altman's, at least with BP data. It has been made available in the Stata and R packages [12,13].

Conclusion

Despite being the most widely used statistical method to assess bias and precision of measurement devices in the medical field (e.g. oscillometric devices), the Bland & Altman LoA method has been shown to suffer several important limitations and provide biased estimates. These defects can be overcome by using a recently published statistical methodology [2]. The computation of the Bias and Precision plots provide a more detailed and precise evaluation of the accuracy (i.e. bias) and precision (i.e. standard deviation of the measurement errors) of a new measurement method than the LoA method. Consequently, it is recommended to adopt this new statistical method to assess bias and precision instead of the conventional Bland & Altman LoA method.

Author's contributions: PT is first author. PT, PH, and MH designed the study. PT carried out the statistical analysis and wrote the first draft. All the authors read and approved the final version of the article.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements: We would like to thank Pr Gérard Waeber (Department of Medicine, Lausanne University Hospital), Pr Lucas Liaudet (Intensive Care Unit, Lausanne University Hospital) and Dr Grégoire Wuerzner (Division of Nephrology and Hypertension, Lausanne University Hospital) for providing the data used for the illustration.

Data sharing: No additional data available.

References

[1] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-310.

[2] Taffé P. Effective plots to assess bias and precision in method comparison studies. *Stat* Methods Med Res 2018;27:1650-1660.

[3] Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Statistician 1983;32,307-317.

[4] Jotterand Chaparro C, Taffé P, Moullet C, Depeyre JL, Longchamp D, Perez MH, Cotting J. Performance of Predictive Equations Specifically Developed to Estimate Resting Energy Expenditure in Ventilated Critically Ill Children. J Pediatr 2017;184:220-226.e5.

[5] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-160.

[6] Halfon M, Wuerzner G, Marques-Vidal P, et al. Use of oscillometric devices in atrial fibrillation: a comparison of three devices and invasive blood pressure measurement. Blood Press 2018;27:48-55.

[7] Taffé P. Assessing bias, precision, and agreement in method comparison studies. Stat Methods Med Res 2019. DOI: 10.1177/0962280219844535.

[8] Dunn G. Statistical evaluation of measurement errors: design and analysis of reliability studies. 2nd ed. London: Arnold, 2004.

[9] Watson T, Lip GY. Blood pressure measurement in atrial fibrillation: goodbye mercury? J Hum Hypertens 2006;20:638-640.

[10] O'Brien E, Pickering T, Asmar R, et al. Working Group on Blood Pressure Monitoring of the European Society of Hypertension International Protocol for validation of blood pressure measuring devices in adults. Blood Press Monit 2002;7:3-17.

[11] Stergiou GS, Karpettas N, Atkins N, O'Brien E. European Society of Hypertension International Protocol for the validation of blood pressure monitors: a critical review of its application and rationale for revision. Blood Press Monit 2010;15:39-48.

[12] Taffé P, Peng M, Stagg V, Williamson T. biasplot: A package to effective plots to assess bias and precision in method comparison studies. Stata J 2017;17:208-221.

[13] Taffé P, Peng M, Stagg V, Williamson T. MethodCompare: An R package to assess bias and precision in method comparison studies. Stat Methods Med Res 2018. DOI: 10.1177/0962280218759693.

Tables

Table 1 Bias computed for each oscillometric device according the three methods

	Average bias in mmHg (LoA)	Differential bias in mmHg (95% CI)		Proportional bias (95% CI)	
	B&A	**eB&A**	**Ta**	**eB&A**	**Ta**
SBP					
WBP device	1.0	-29.7	-9.1	1.22	1.08
	(-23.6; 25.6)	(-52.0; -7.4)	(-52.5; 34.4)	(1.08; 1.36)	(0.72; 1.45)
OHEM device	-8.4	-26.6	-15.3	1.13	1.05
	(-29.2; 12.5)	(-47.7; -5.5)	(-40.6; 10.1)	(0.99; 1.27)	(0.89; 1.21)
DBP					
WBP device	9.4	-7.4	16.8	1.26	0.88
	(-18.1; 36.8)	(-28.7; 13.9)	(-34.6; 68.2)	(0.98; 1.55)	(-0.01; 1.77)
OHEM device	4.3	2.7	7.4	1.03	0.95
	(-7.0; 15.6)	(-4.6; 10.1)	(-12.5; 27.4)	(0.91; 1.15)	(0.58; 1.31)

WBP: WatchHome BP device. OHEM: OmronHEM907 device; B&A: Bland Altman method; eB&A: extended Bland Altman method; Ta: Taffé method; SBP: systolic blood pressure; DBP: diastolic blood pressure

Appendices

Appendix A

Table 1: Characteristics of the studied population (n=10)

Women (%)	4 (40.0)
Age in years mean (SD)	78.9 (9.1)
Arm circumference in CM mean (SD)	28.7 ± 2.3
Mean systolic blood pressure in mmHg (SD)	
Omron HEM907™	118 ± 24
Microlife WatchBPHome™	126 ± 24
Invasive measure	125 ± 20
Mean systolic blood pressure in mmHg (SD)	
Omron HEM907™	63 ± 14
Microlife WatchBPHome™	70 ± 15
Invasive measure	60 ± 13
Mean heart frequency	
measured by Invasive measure	98 ± 12
Diabetes (%)	3 (30.0)
Hypertension (%)	8 (80.0)
Kidney failure (%)	6 (66.7)

Appendix B

To illustrate the concepts of differential and proportional biases, consider the following relation between the true (latent) SBP, i.e. true_SBP, and the measured SBP:

$$SBP = a + b*true\_SBP + error$$

Then, "a" is called "differential bias" and "b" proportional bias". Therefore, in the absence of differential bias "a" = 0 and in the absence of proportional bias "b" = 1. In this case, one has SBP = true_SBP + error, where "error" represents measurement errors.

The (total) bias is thus given by:

$$bias = SBP - true\_SBP = a + (b - 1)*true\_SBP + error$$

Therefore, in the presence of a proportional bias ("b" $\neq$ 1) the bias is not constant and depends on the level of the true latent trait (here true_SBP).

To compute the bias one needs an estimate of "true_SBP" for each individual. Bland & Altman use the average of the two measurements to estimate that quantity,[1,4] whereas Taffé[2] uses an empirical Bayes approach. The great advantage of the empirical Bayes method is that it uses the data from all the individuals to get the best possible linear prediction of the true value of SBP for each individual (called "BLUP of x", i.e. Best Linear Unbiased Prediction of the latent trait x).

Notice that when the measurement method is the reference (the IBP method in our case), one has:

$$IBP = true\_SBP + error$$

That is, there is no differential ("a" = 0) nor proportional bias ("b" = 1), just measurement errors. Clearly, very often, even the gold standard is subject to measurement errors, despite having no bias.

Finally, to remove the differential and proportional biases, one may recalibrate the SBP measurements by computing:

$$SBP^* = (SBP - a)/b$$

where SBP* represents the recalibrated (i.e. de-biased) SBP measurements.

Appendix C

Bias of the OHEM device

We have successively used the three statistical methods (B&A, eB&A, and Ta) to analyze the SBP and DBP data measured by the OHEM device.

Starting with the B&A method:

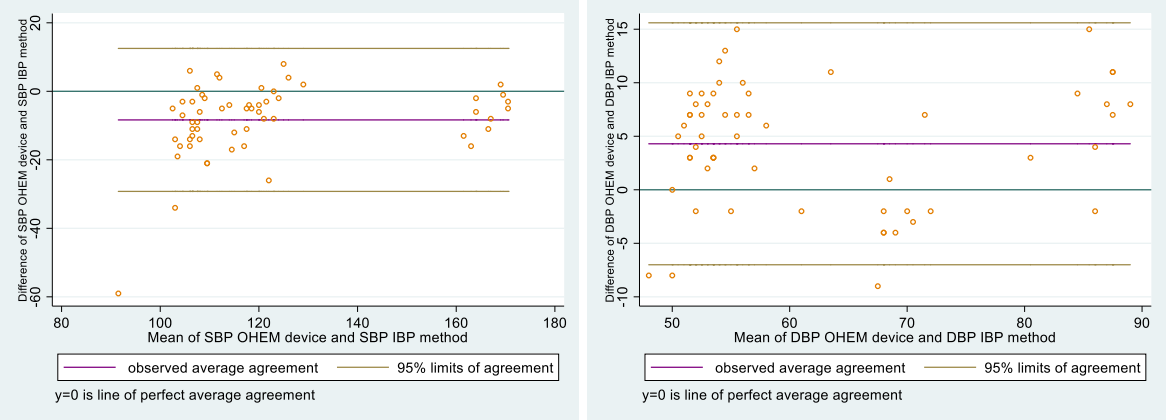

Figure 1bis Classical Bland & Altman' LoA plot for SBP (left) and DBP (right) measured by the OHEM device
The LoA plots reveals an average bias of -8.4 mmHg for SBP and 4.3 mmHg for DBP.

Using the eB&A method allows one to let the bias depend on the level of true BP (estimated by the mean of the two measurements for each individual) and estimate the proportional and differential biases from the regression line coefficients:

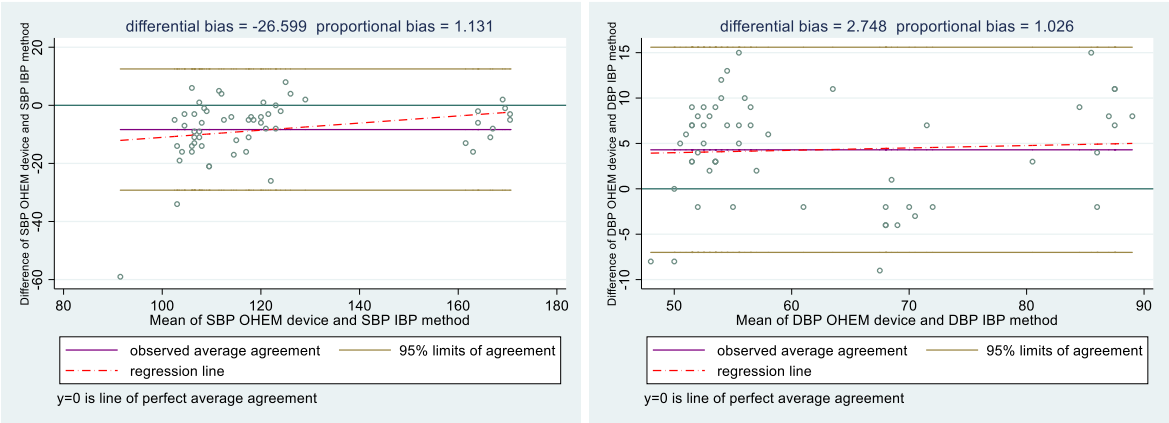

Figure 2bis Extended Bland & Altman' LoA plot for SBP (left) and DBP (right) measured by the OHEM device
By adding on the plots the regression line of the differences versus the means (of the two measurements), one may have the impression that the bias is increasing with the level of BP (red dash-dotted regression line). The estimates of the differential and proportional biases are provided on top of each figure.

Note that the bias is computed using a mathematical formula, and depends on the differential and proportional biases (see Appendix B).

Finally, using the Ta method, the amount of bias also depends on the level of true BP (estimated by the BLUP) and estimates of the differential and proportional biases are given on top of the Bias plot:

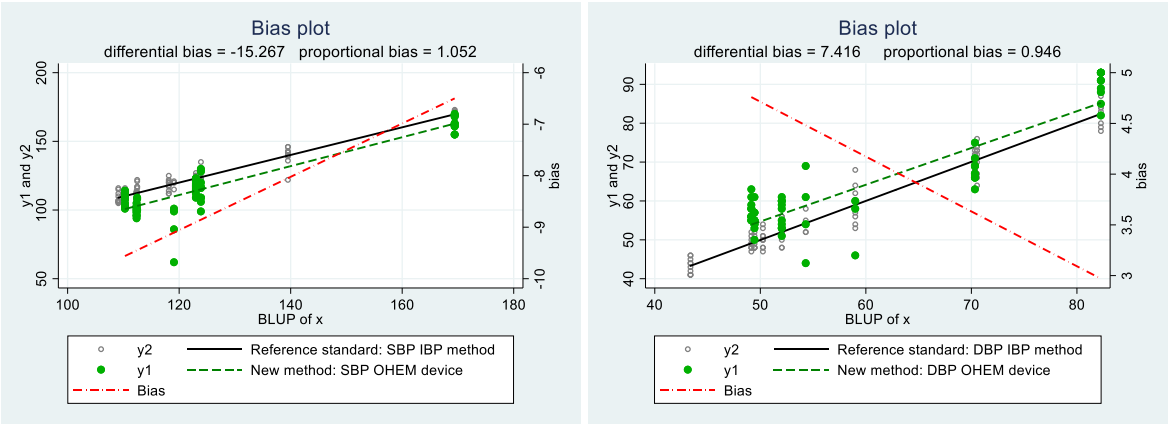

Figure 3bis Bias plot for SBP (left) and DBP (right) measured by the OHEM device
Estimates of the differential and proportional biases are given on top of of each figure. The estimated bias (red dash-dotted regression line) is increasing with the level of true BP (i.e. BLUP of x) for SBP, whereas it is decreasing for DBP (the amount of bias in mmHG can be read from the right y-axis).

Focusing, on the estimation of the differential bias, there are striking differences between the estimates from the eB&A and Ta methods: -26.6 versus -15.3 for SBP, and 2.7 versus 7.4 for DBP. In this example, for SBP the amount of differential bias is over estimated by the eB&A method with respect to the Ta, whereas for DBP it is under-estimated.

Likewise, for the proportional bias, the estimates by the eB&A and Ta methods differ substantially: 1.13 versus 1.05 for SBP, and 1.03 versus 0.95 for DBP. Again, in this example, for SBP the eB&A method overestimates the proportional bias with respect to the Ta method, and for DBP the two

methods provide proportional bias estimates of opposite directions (1 is the reference and means no bias).

Turning to the bias estimate (red dash-dotted line), one can see that the estimated bias is increasing with the level of true BP (i.e. BLUP of x) for SBP, whereas it is decreasing for DBP (the amount of bias in mmHG can be read from the right y-axis). The latter result is the opposite from what was obtained by the eB&A method. Clearly, the two methods provide completely different bias estimates in these data.

Focusing on the left y-axis, the two regression lines (solid black and dashed green) allow one to assess the difference between the reference (i.e. IBP) and the device (i.e. OHEM). One can see on the left plot that for true values of SBP (i.e. BLUP of x) around 110 mmHG the bias of the OHEM device is about -9.5 mmHG, whereas for values around 170 mmHG it is about -6.5 mmHG. For DBP (right plot), the discrepancy between IBP and OHEM is of opposite sign, the latter providing systematically larger BP values than the former (the discrepancy goes from 3 to 4.7 mmHG for values of true BP comprised between 43-82).

Precision of the OHEM device

Coming back to Figure 1bis, the width between the two limits are quite large [-29.2; 12.5] for SBP and [-7.0; 15.6] for DBP. However, these limits are of limited usefulness as they allow one to assess the variability of the differences, but not of each device separately.

Therefore, to get an estimate of precision for each device (OHEM and IBP), the standard deviations have been computed using the Ta method. As expected, the precision of the OHEM oscillometric device and of the IBP method depends on the level of BP values: the standard deviation of the measurement errors increases with increasing blood pressures. This is illustrated in the Figure below where the x-axis represents the BLUP of x (i.e. the best possible linear prediction of the true BP for each individual) and the y-axis the standard deviation of the measurement errors:

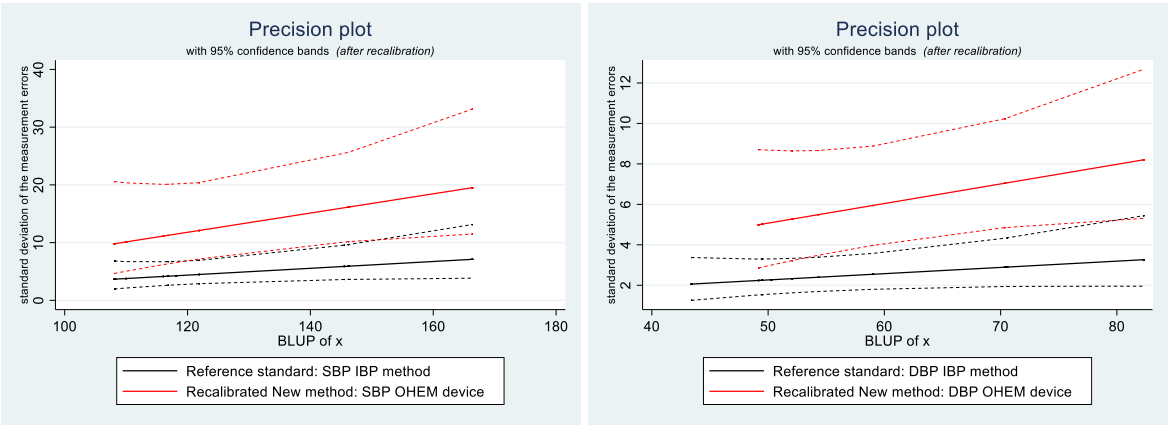

Figure 4bis Precision plot for SBP (left) and DBP (right) measured by the OHEM device.
The x-axis represents the true blood pressure and the y-axis the standard deviation of the measurement errors. The two regression lines, on each of the two plots, clearly illustrate the higher the true blood pressure the higher the measurement errors (i.e. the standard deviation of the measurement error is increasing with the true BP values). It also shows that the precision of the IBP method tends to be better than that of the OHEM device, particularly for values between 55 and 75 mmHG for DBP where the confidence bands do not overlap.

Notice that, the right plot illustrates that, as there was no measurement available with the OHEM device for the patient having the lowest IBP measurements, the standard deviation line starts at a DBP of around 49 instead of 43 mmHg.

Clearly, from Figure 4bis the standard deviation of the measurement errors increases with increasing blood pressures. Of clinical interest, the precision of the IBP method tends to be better than that of the OHEM device.