How to “validate” newly developed cardiac output monitoring devices

J. J. Vos¹ · T. W. L. Scheeren¹

Received: 25 August 2015 / Accepted: 25 August 2015 / Published online: 13 October 2015
© Springer Science+Business Media Dordrecht 2015

In the past decade, technological advances have catalyzed the development of (advanced) hemodynamic monitoring devices from invasive towards less invasive methods. Almost every month a new device appears on the market claiming to measure or estimate hemodynamic variables in a minimal or non-invasive fashion. These devices allow clinicians to broaden the scope beyond traditional pressure-based hemodynamic monitoring and by now, flow(-related) variables such as cardiac output (CO) can be assessed this way in almost all patients in the anesthetic or critical care setting. Hopefully, these technological advancements prove advantageous in terms of patient outcome in the (near) future. However, an accurate and precise estimation of the assumed-to-be-measured hemodynamic variable is an absolute prerequisite before such new devices can be clinically implemented, or even before outcome-related studies can be performed. Additionally, given the (hemodynamic) heterogeneity of various patient populations (e.g. patients with septic shock versus patients with cardiogenic shock), both accuracy and precision should be investigated in the relevant patient population(s). Therefore, researchers all over the world are stimulated to perform comparison studies in which new devices, software versions or sensor revisions are compared with their clinical reference methods or “gold standards”, and journals are overflowed with manuscripts on such evaluation studies.

The basis for statistical reporting for such studies was set almost 30 years ago by Bland and Altman in their landmark paper [1], in which they introduced their famous “Bland–Altman plot”. In this (scatter)plot, agreement between two measurement methods is assessed by plotting the mean of the measurements against the difference of the measurements and it allows the calculation of the limits of agreement (LOA; $1.96 \times \text{SD of the bias}$) as a measure of reproducibility. As such, the Bland–Altman plot appreciates two important aspects: it assesses method agreement based on the “closeness” of individual data points and it does not require the definition of a gold standard, i.e. it assumes no fixed “true” value. Since its introduction, the Bland–Altman plot (initially developed for comparing CO measurements) has become the minimal standard for statistical reporting of method comparison studies in general [2].

While the Bland–Altman plot remains highly popular, it bears some important limitations that are especially relevant for CO method agreement studies. At first, the plot only provides an estimation of agreement in the light of a linear relationship between the two measurement methods, and it does not take the magnitude of these observations into consideration, while this is highly important as a “closer” agreement is required for a CO value of 1.5 L/min compared to a CO of 3.5 L/min. To overcome this issue, Critchley and Critchley [3] introduced percentage errors (calculated as the LOA divided by the mean of the measurements) to compensate agreement for the magnitude of measurement. In CO agreement studies, the calculation of percentage errors is well adopted and there is general consensus for accepting new CO measurement methods if these devices meet the so-called Critchley criteria, which means that the percentage error is below 30 %.

As all clinicians will be aware of, CO is not a static variable and changes continuously secondary to complex

✉ T. W. L. Scheeren
t.w.l.scheeren@umcg.nl

J. J. Vos
j.j.vos@umcg.nl

¹ Department of Anesthesiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, PO Box 30 001, 9700 RB Groningen, The Netherlands

physiologic regulatory mechanisms. As such, the use of the original Bland–Altman plot alone in CO method agreement studies is further reduced by the fact that it requires independency of all data points, which is obviously not the case in most CO method agreement studies as multiple measures per subject are obtained. In such a case, the LOA can be estimated too narrow and to correct for this issue, two techniques—the “Myles adaptation” [4] and the adaptation by Bland and Altman themselves [5]—are available. While these adaptations improve the analysis of method agreement and while “polar plots” and “4-quadrant plots” have recently emerged as new statistical tools for evaluating trending ability [6], one important aspect concerning data variability remains untouched. In this issue of the Journal of Clinical Monitoring and Computing, Hapfelmeier et al. [7], add an important chapter to the statistics of CO method agreement studies as they dismantle the factors that affect the agreement and redefine the statistical term “precision”. The authors elegantly explain that every method—both the reference and the investigated method—has a precision of method (POM), defined as the “certainty of a method about its estimates of a true value”. The precision of agreement (POA), i.e. the standard deviation of the observed bias, is composed of the POM of both methods but this issue is infrequently handled in most method comparison studies. In fact, most CO agreement studies are designed to provide only one measurement value per “true” value, also if in time, measurements are repeated. As such, these studies seem to provide “repeated measures”, but in fact they do not—according to the authors—as there is only one measurement value per single true value and hence, they do not allow calculation of true precision as the POM cannot be calculated. Therefore, an estimation of “precision” is provided that can be statistically incorrect, according to the authors, and they suggest that per individual true value multiple data points should be obtained for allowing the calculation of the POM of the reference technique. The same group previously emphasized this issue, because the Critchley criteria are based on an *assumed* precision of the reference technique of 20 %, a value that is likely to differ between measurement methods and during various clinical situations [8].

Therefore, the authors are to be congratulated on their sound elucidation of the statistical background of CO method agreement studies and the review emphasizes some very important aspects of data variability that are infrequently taken into account. Nevertheless, as also suggested previously [9], calculating the POM of the reference technique under clinical circumstances—which requires multiple CO measurements per individual time point—is clinically not very feasible. Hence, especially for CO agreement studies where the true value varies continuously due to intrinsic physiologic factors and medical-based

interventions, a straightforward statistical analysis remains complex. Nevertheless, whilst calculating the precision of the reference method is not very feasible in CO agreement studies, we suggest that all future studies follow the “guide” set by Hapfelmeier and co-workers and comply with their recommendations. As such, at least unity between studies can be increased which allows a more sound meta-comparison of all CO method agreement studies and the ongoing discussion [10–12] on whether or not new CO monitoring devices can or should be implemented clinically, can be funded on more sound and appropriate statistical analyses.

References

1. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–10.
2. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PLoS One*. 2012;7:e37908. doi:[10.1371/journal.pone.0037908](https://doi.org/10.1371/journal.pone.0037908).
3. Critchley LA, Critchley JA. A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. *J Clin Monit Comput*. 1999;15:85–91.
4. Myles PS, Cui J. Using the Bland–Altman method to measure agreement with repeated measures. *Br J Anaesth*. 2007;99:309–11. doi:[10.1093/bja/aem214](https://doi.org/10.1093/bja/aem214).
5. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17:571–82. doi:[10.1080/10543400701329422](https://doi.org/10.1080/10543400701329422).
6. Saugel B, Grothe O, Wagner JY. Tracking changes in cardiac output: statistical considerations on the 4-quadrant plot and the polar plot methodology. *Anesth Analg*. 2015;121:514–24. doi:[10.1213/ANE.0000000000000725](https://doi.org/10.1213/ANE.0000000000000725).
7. Hapfelmeier A, Cecconi M, Saugel B. Cardiac output method comparison studies: the relation of the precision of agreement and the precision of method. *J Clin Monit Comput*. 2015. doi:[10.1007/s10877-015-9711-x](https://doi.org/10.1007/s10877-015-9711-x).
8. Cecconi M, Rhodes A, Poloniecki J, Della Rocca G, Grounds RM. Bench-to-bedside review: the importance of the precision of the reference technique in method comparison studies—with specific reference to the measurement of cardiac output. *Crit Care*. 2009;13:201. doi:[10.1186/cc7129](https://doi.org/10.1186/cc7129).
9. Critchley LA. Bias and precision statistics: should we still adhere to the 30 % benchmark for cardiac output monitor validation studies? *Anesthesiology*. 2011;114:1245–6. doi:[10.1097/ALN.0b013e318215e1e5](https://doi.org/10.1097/ALN.0b013e318215e1e5).
10. Saugel B, Cecconi M, Wagner JY, Reuter DA. Noninvasive continuous cardiac output monitoring in perioperative and intensive care medicine. *Br J Anaesth*. 2015;114:562–75. doi:[10.1093/bja/aeu447](https://doi.org/10.1093/bja/aeu447).
11. Saugel B, Reuter DA. Are we ready for the age of non-invasive haemodynamic monitoring? *Br J Anaesth*. 2014;113:340–3. doi:[10.1093/bja/aeu145](https://doi.org/10.1093/bja/aeu145).
12. Wagner JY, Saugel B. When should we adopt continuous non-invasive hemodynamic monitoring technologies into clinical routine? *J Clin Monit Comput*. 2015;29:1–3. doi:[10.1007/s10877-014-9619-x](https://doi.org/10.1007/s10877-014-9619-x).