Data Agnosticism and Implications on Method Comparison Studies

Robert H. Thiele, MD, and Timothy L. McMurry, PhD

The evolution of method comparison studies has occurred through a series of fits and starts, the most notable of which occurred in the early 1980s as the shortcomings of the traditional, linear regression approach to method comparisons were described and a complementary, "agreement"-based methodology proposed.[1–3] However, the evolution did stop there. As clinicians began to deemphasize absolute values and focus on trend monitoring (cynics argue that the latter is easier, advocates suggest it is more useful), additional techniques were developed, the 2 most notable of which are the 4-quadrant and polar plotting techniques developed by Perrino et al.[4] and Critchley et al.[5]

In this issue of *Anesthesia & Analgesia*, Saugel et al.[6] describe, step by step, how these plots are produced and, most importantly, how they differ. By "transforming" trending data from Cartesian to polar coordinate systems, Saugel et al.[6] demonstrate, both visually and mathematically, how the selection of a particular analytical technique can impact the results. The fact that highly discordant measurements in which the average change is 0 are excluded from the quantitative estimate of agreement when using the polar technique, but not the 4-quadrant technique, is essential to proper interpretation of these tests.

Why does the polar plotting technique do this? The reason is because the polar plotting technique does not make any assumptions about which technique is better. Ironically, most published studies that use the polar plotting technique compare a new method of measurement with an accepted reference standard, and in these instances, an agnostic approach may not be appropriate.[7–9] This question of "what is truth" may sound philosophical on the surface but in reality has significant mathematical implications. For instance, if changes in cardiac output (ΔCO) are measured by a pulmonary artery catheter and a magic 8 ball, and the pulmonary artery catheter estimates that ΔCO is −2 L/min but the magic 8 ball estimates that ΔCO is 2 L/min, the polar analysis will exclude that data point because "true" ΔCO is 0 L/min. This is, in essence, information loss. The 4-quadrant technique, by contrast, will place this point in quadrant 2, which will negatively affect the estimate of concordance.

Interestingly, the Bland-Altman technique, as it was originally described, takes the same approach to the relative value of data.[1–3] In some sense, this is appropriate. As Altman pointed out, clinicians are not particularly interested in the probability that the slope of a best fit line between 2 outputs is not 0 "when the two variables are obviously associated by their very nature… What we really want to know in these studies is how well the two measures agree."[1] However, this assumes that 2 variables are actually related. A key element to the agreement strategy, which is often overlooked, is that "good" agreement does not necessarily imply any correlation if the range of data tested is small. From the viewpoint of the clinician, performance of an agreement analysis makes the a priori assumption that a correlation actually exists.

Take, for instance, 2 monitors, X and Y, which estimate cardiac output (Θ). If the slope of the best fit line relating X and Y is m, estimates of Θ can be described as

$$X_i = \Theta_i + \varepsilon_{X,i}$$

and

$$Y_i = m\,\Theta_i + \varepsilon_{Y,i}$$

where Θ represents the true value of stroke volume and ε represents normally distributed measurement errors. Now let us assume that m = 0, that is, that there is absolutely no statistical correlation between the 2 devices (i.e., they are completely independent of one another). If X and Y are "tested" over a narrow range of values (i.e., the distribution of Θ is small [Fig. 1, upper left quadrant]), it is possible for a random number generator to produce limits of agreement approaching what would be deemed clinically "acceptable" by Critchley and Critchley,[10] who in 1999 stated that "Bias and precision statistics has now replaced correlation and regression as the accepted statistical technique of comparing two techniques measuring the same physiological variable, such as cardiac output."

Clearly, many in the anesthesia community agree, because some investigators have stopped publishing correlation coefficients and even basic scatterplots altogether, relying exclusively on limits of agreement and/or trending analysis to compare methods of measurement,[7,8,11–15] yet this

From the Department of Anesthesiology, University of Virginia, Charlottesville, Virginia.

Accepted for publication April 9, 2015.

Funding: Departmental.

The authors declare no conflicts of interest.

This report was previously presented, in part, at the meeting of the International Anesthesia Research Society, 2012.

Reprints will not be available from the authors.

Address correspondence to Robert H. Thiele, MD, Department of Anesthesiology, University of Virginia, P.O. Box 800710, Charlottesville, VA 22908. Address e-mail to rht7w@virginia.edu.

DOI: 10.1213/ANE.0000000000000810

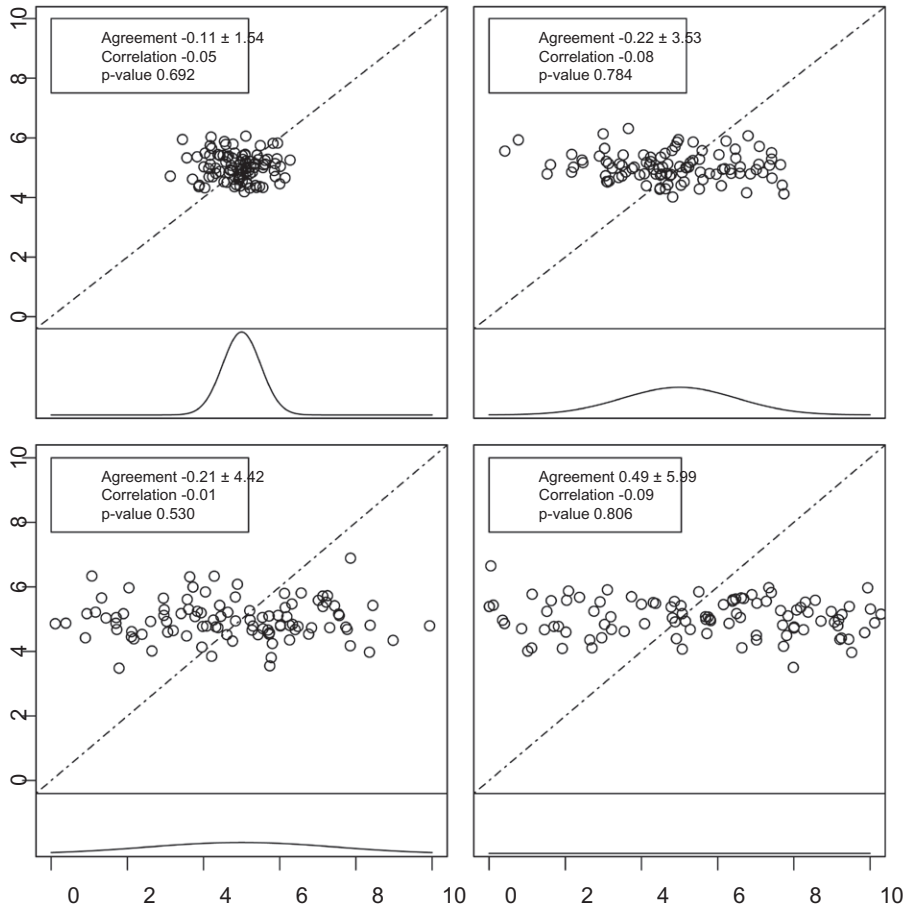

Figure 1. Limits of agreement, correlation coefficient, and limits of agreement-based *P* value for 4 cardiac output data sets of varying sample distribution (the distribution of the data is displayed on the lower aspect of each figure) and in whom there is no correlation between X and Y. This data set demonstrates that if σ_Θ is decreased, the limits of agreement will also decrease despite the fact that the intrinsic error has not changed. By contrast, the correlation coefficient is relatively insensitive to σ_Θ.

was not the intent of Bland or Altman. Indeed, in their 1983 manuscript, they stated "The first step, one which should be mandatory, is to plot the data… For the purposes of comparing the methods the line of identity is much more informative, and is essential to get a correct visual assessment of the relationship."[2] Three years later, in their oft-cited Lancet paper, they reemphasized that "The first step is to examine the data… A simple plot of the results of one method against those of the other though without a regression line is a useful start."[3]

Thus, when making method comparison studies, we suggest that the following steps are taken:

1. The authors explicitly state whether or not 1 device is being used as a reference standard.
2. The authors state, a priori, the acceptable range of clinically acceptable values for each statistical test used. For instance, the acceptable limits of agreement for blood pressure monitoring are likely to be narrower during cerebral aneurysm surgery than placement of ear tubes.
3. Data are initially presented in a scatterplot to allow visual assessment of the relationship, as suggested by Bland and Altman.[2,3] This will reduce the probability that the limits of agreement between measures, which have no meaningful relationship, are deemed clinically acceptable. Preiss and Fisher[16] described a random permutation technique, which estimates the probability that a paired data set arose from unassociated measurements, which some investigators may want to consider using.
4. The limits of agreement are calculated and plotted, as described by Bland and Altman, with allowances made for repeated measures when necessary.[2,3,17,18] For repeated measures, Bland and Altman base the estimate of the limits of agreement on both within-subject variances and the variance of the differences between-subject means.[17] Myles and Cui[18] recommend calculating the mean of repeated measures and using a random effects model to account for the reduced variation that occurs with averaging. When the residuals are not normally distributed, a transformation may be appropriate. For instance, some data may be distributed log-normally. In these instances, Dexter et al.[19] have suggested both a nonparametric (based on ranking the observations and selecting cutoff values that remain within the desired percentiles [taking into account sample size and degrees of freedom]) and a parametric approach (based on the Student *t* distribution) for the calculation of prediction limits, either of which is acceptable. Last, confidence intervals around these limits should be calculated and displayed graphically, ensuring that the limits are interpreted properly (i.e., that the sample size is sufficient to draw meaningful conclusions from the data).[3,17]

5. When trending data are analyzed, they should be performed using either the 4-quadrant (when an accepted standard is being used) or a polar plotting techniques (when there is no accepted reference standard).[4,5] If the polar plotting technique is used, particular attention must be paid to the exclusion zone, because the polar plotting technique may exclude important data points that are highly discordant.[6]

Saugel et al.[6] have pointed out a subtle but extremely important and underappreciated difference in 2 frequently used statistical techniques. They have also helped clarify a concept that has clearly confused many (the original paper describing the polar plotting technique states that "agreement is shown by the angle the vector makes with the line of identity [y x] and magnitude of change by the length of the vector (Fig. 4)," yet in the fifth published figure, and the Appendix, the average value is used, not the length of the vector).[5] It is our hope that standardizing the presentation of data will improve the inferences derived from these important studies.

DISCLOSURES

Name: Robert H. Thiele, MD.
Contribution: This author helped design the study and write the manuscript.
Attestation: Robert H. Thiele approved the final manuscript.
Name: Timothy L. McMurry, PhD.
Contribution: This author helped design the study and write the manuscript.
Attestation: Timothy L. McMurry approved the final manuscript.
This manuscript was handled by: Franklin Dexter, MD, PhD.

REFERENCES

1. Altman DG. Statistics and ethics in medical research: V–analysing data. Br Med J 1980;281:1473–5
2. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Statistician 1983;32:307–17
3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10
4. Perrino AC Jr, O'Connor T, Luther M. Transtracheal Doppler cardiac output monitoring: comparison to thermodilution during noncardiac surgery. Anesth Analg 1994;78:1060–6
5. Critchley LA, Lee A, Ho AM. A critical review of the ability of continuous cardiac output monitors to measure trends in cardiac output. Anesth Analg 2010;111:1180–92
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
7. Maus TM, Reber B, Banks DA, Berry A, Guerrero E, Manecke GR. Cardiac output determination from endotracheally measured impedance cardiography: clinical evaluation of endotracheal cardiac output monitor. J Cardiothorac Vasc Anesth 2011;25:770–5
8. van der Kleij SC, Koolen BB, Newhall DA, Gerritse BM, Rosseel PM, Rijpstra TA, Geisler FE, van der Meer NJ. Clinical evaluation of a new tracheal impedance cardiography method. Anaesthesia 2012;67:729–33
9. Suehiro K, Tanaka K, Mikawa M, Uchihara Y, Matsuyama T, Matsuura T, Funao T, Yamada T, Mori T, Nishikawa K. Improved performance of the fourth-generation FloTrac/Vigileo system for tracking cardiac output changes. J Cardiothorac Vasc Anesth 2015;29:656–62
10. 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
11. Møller-Sørensen H, Hansen KL, Østergaard M, Andersen LW, Møller K. Lack of agreement and trending ability of the endotracheal cardiac output monitor compared with thermodilution. Acta Anaesthesiol Scand 2012;56:433–40
12. Cooper ES, Muir WW. Continuous cardiac output monitoring via arterial pressure waveform analysis following severe hemorrhagic shock in dogs. Crit Care Med 2007;35:1724–9
13. Hadian M, Kim HK, Severyn DA, Pinsky MR. Cross-comparison of cardiac output trending accuracy of LiDCO, PiCCO, FloTrac and pulmonary artery catheters. Crit Care 2010;14:R212
14. Cecconi M, Dawson D, Casaretti R, Grounds RM, Rhodes A. A prospective study of the accuracy and precision of continuous cardiac output monitoring devices as compared to intermittent thermodilution. Minerva Anestesiol 2010;76:1010–7
15. Sokolski M, Rydlewska A, Krakowiak B, Biegus J, Zymlinski R, Banasiak W, Jankowska EA, Ponikowski P. Comparison of invasive and non-invasive measurements of haemodynamic parameters in patients with advanced heart failure. J Cardiovasc Med (Hagerstown) 2011;12:773–8
16. Preiss D, Fisher J. A measure of confidence in Bland-Altman analysis for the interchangeability of two methods of measurement. J Clin Monit Comput 2008;22:257–9
17. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60
18. Myles PS, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. Br J Anaesth 2007;99:309–11
19. Dexter F, Epstein RH, Bayman EO, Ledolter J. Estimating surgical case durations and making comparisons among facilities: identifying facilities with lower anesthesia professional fees. Anesth Analg 2013;116:1103–15

Tracking Changes in Cardiac Output: Statistical Considerations on the 4-Quadrant Plot and the Polar Plot Methodology

Bernd Saugel, MD,* Oliver Grothe, PhD,† and Julia Y. Wagner, MD*

When comparing 2 technologies for measuring hemodynamic parameters with regard to their ability to track changes, 2 graphical tools are omnipresent in the literature: the 4-quadrant plot and the polar plot recently proposed by Critchley et al. The polar plot is thought to be the more advanced statistical tool, but care should be taken when it comes to its interpretation. The polar plot excludes possibly important measurements from the data. The polar plot transforms the data nonlinearily, which may prevent it from being seen clearly. In this article, we compare the 4-quadrant and the polar plot in detail and thoroughly describe advantages and limitations of each. We also discuss pitfalls concerning the methods to prepare the researcher for the sound use of both methods. Finally, we briefly revisit the Bland-Altman plot for the use in this context. (Anesth Analg 2015;121:514–24)

Because cardiac output (CO) is an important hemodynamic parameter in caring for hemodynamically unstable patients, studies describing novel technologies for CO assessment are of high interest in the fields of perioperative and critical care medicine. In these studies, an innovative method for CO determination (i.e., studied technology) is usually compared with an established reference technology using different statistical methodologies. Methods for the assessment of the accuracy and precision of a studied technology (e.g., Bland-Altman analysis[1,2] and calculation of the percentage error[3]) have been described and discussed in detail elsewhere.[4,5]

Besides describing its absolute accuracy and precision, it is important to assess the ability of a novel technology designed to measure CO to adequately track changes in CO in comparison with the gold standard method. This means that the technologies detect changes in the same direction. Several methods for the evaluation of this trending ability have been described.[6]

Although the Bland-Altman analysis can provide insights within a trending analysis, 2 of the most frequently used graphical statistical methods in trending analysis are the 4-quadrant plot and the polar plot. We therefore restrict ourselves to the analysis of these 2 methods in this article. (For readers interested in the Bland-Altman analysis, we provide additional treatments in Appendix 1.) The 4-quadrant plot was first used for the description of trending capabilities in studies comparing one CO measurement technology with another by Perrino et al.[7,8] The polar plot was proposed by Critchley et al.[6] in 2010 as a new alternative method. In their review article published in 2010 and in another article published later,[9] Critchley et al. described the derivation of the polar plot from the 4-quadrant plot. In their important articles, Critchley et al. demonstrated the importance, and also the complexity, of CO-trending analysis compared with precision analysis and in doing so drew attention to the problem of quantifying the ability of a technology to track CO changes. Since its introduction, numerous studies have used the polar plot analysis to describe the ability of a CO-monitoring device to follow changes in the true CO measured with a gold standard technique.

We argue that a profound understanding of the statistical methods used is a prerequisite for the correct assessment of the trending ability of a CO measurement technology. Whereas the 4-quadrant plot provides a relatively intuitive picture of the analyzed data at hand, the more sophisticated polar plot demands a higher level of insight into its construction to adequately interpret the characteristics of the analyzed data. Therefore, the primary scope of the present article is to describe the computation of the 4-quadrant plot and the polar plot in detail and to derive the relation between these statistical methods. Furthermore, we describe the basic properties of both plots and, in particular, cite possibly dangerous pitfalls when analyzing polar plots. We briefly review the problem of measuring CO and assessing the trending ability of measurement technologies. We then discuss the 4-quadrant plot and the polar plot as proposed by Critchley et al. in detail. Finally, we summarize the advantages and disadvantages of both methods.

MEASURING CO—THE PROBLEM OF TRACKING CHANGES

Pulmonary artery thermodilution,[10,11] single-indicator transpulmonary thermodilution,[12–14] and lithium indicator dilution[15–17] are thought to represent clinical gold standard methods for CO determination and are therefore used as reference technologies in method comparison studies.

Various novel, less invasive, and noninvasive technologies have been described in recent years including pulse

From the *Department of Anesthesiology, Center of Anesthesiology and Intensive Care Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; and †Department of Economic and Social Statistics, University of Cologne, Cologne, Germany and Institute of Operations Research, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Accepted for publication January 15, 2015.

Funding: None.

The authors declare no conflicts of interest.

Address correspondence to Julia Y. Wagner, MD, Department of Anesthesiology, Center of Anesthesiology and Intensive Care Medicine, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. Address e-mail to jwagner.science@gmail.com.

DOI: 10.1213/ANE.0000000000000725

contour analysis (both calibrated and autocalibrated),[18–25] esophageal Doppler,[25,26] thoracic electrical bioimpedance and bioreactance,[27–32] and technologies based on the vascular unloading technique,[33–36] pulse wave transit time, and radial artery applanation tonometry.[37]

When applying different CO measurement technologies, one has to keep in mind that CO can be measured intermittently (e.g., intermittent pulmonary artery thermodilution), continuously (e.g., pulse contour analysis providing a real-time beat-to-beat report), or semicontinuously (e.g., bioreactance-derived CO readings averaged over 60 seconds).

CO is a hemo*dynamic* variable that changes over time and is modified by a variety of factors closely related to oxygen supply and consumption, such as cardiac preload, cardiac afterload, and cardiac contractility. When performing validation studies for CO-monitoring technologies, it has therefore to be kept in mind that both the studied technology and the reference technology are aiming to hit a moving target.

Considering the dynamic nature of CO, evaluating the ability of a technology for CO assessment to trend changes, in addition to assessing its accuracy and precision, is essential for the sound interpretation of the measurement performance of a novel CO-monitoring device. However, adequately describing the ability of a CO-monitoring method to timely track decreases and increases in CO is statistically complex. Several statistical approaches have been described previously.[6]

A direction of change analysis can be performed by calculating the concordance rate, that is, the ratio (percentage) of CO measurements assessed by the studied technology and the reference technology that change correctly in the same direction (decrease or increase) to the sum of all changes. However, although this direction of change analysis provides information whether the studied technology qualitatively follows CO changes assessed by the reference technology, it does not provide information on the magnitude of the changes in CO or the degree of agreement between the studied technology and the reference technology.[6]

Therefore, alternative and more sophisticated methods for trend analysis in clinical studies have been described. In recent years, the most widely used methods to illustrate the trending ability of CO-monitoring devices are the 4-quadrant plot and the polar plot.

FOUR-QUADRANT PLOT ANALYSIS

For the computation of a 4-quadrant plot, ΔCO values (i.e., differences between consecutively obtained CO values) for both the studied technology and the reference technology are calculated and plotted in a scatter plot. Figure 1 shows an example for a 4-quadrant plot with 9 artificial data points. The values on the horizontal axis (usually called the x-axis) refer to ΔCO values of the reference technology, whereas the vertical axis (the y-axis) refers to the ΔCO values of the studied technology. From visual inspection of the resulting scatter plot, one can see the distribution of data points lying within 1 of the 4 quadrants. When both the studied technology and the reference technology indicate an increase in CO, the respective data point will appear in the upper right quadrant of the 4-quadrant plot. Similarly, the lower left quadrant contains data points resulting from concordant CO measurements indicating a decrease in CO. Therefore, the upper right and the lower left quadrants of the 4-quadrant plot represent concordant measurements of the studied technology and the reference technology with regard to direction of changes. In Figure 1, these quadrants are therefore marked by green areas. From the coordinates of 1 data point within the quadrant, the magnitude of change in CO measured by the studied technology and the reference technology can directly be read off. This is an appealing property of the 4-quadrant plot. Data point 8, for example, means that the reference device detected a CO change by 0.5 L/min, whereas the studied device showed a change by 2 L/min. Although these measurements are concordant, in the sense that both devices indicated a positive change in the CO, the numerical values are not equal. Points with equal numerical values are located on the 45° diagonal within the quadrant (the dotted line in the green quadrants in Fig. 1). In data point 5, for example, both devices detect a CO change by 1 L/min.

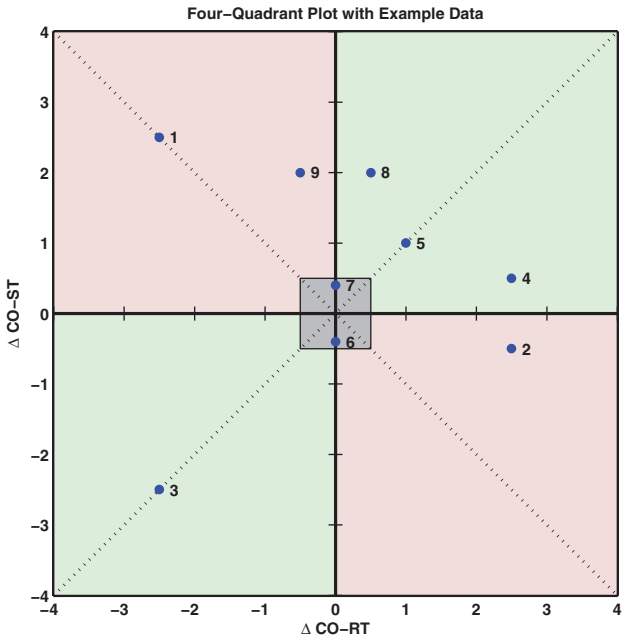

Figure 1. Four-quadrant plot showing 9 artificial Δcardiac output (ΔCO) values for both the studied technology (ST) and the reference technology (RT). The values on the horizontal axis refer to ΔCO values of the RT (ΔCO-RT), whereas the vertical axis refers to the ΔCO values of the ST (ΔCO-ST). When both the ST and the RT indicate an increase in CO, the respective data point will appear in the upper right quadrant of the 4-quadrant plot. In contrast, the lower left quadrant contains data points resulting from concordant CO measurements, indicating a decrease in CO (concordant measurements = green areas). From the coordinates of 1 data point within the quadrant, the magnitude of change in CO measured by the ST and the RT can be read off directly. Points with equal numerical values are located on the 45° diagonal within the quadrant (the dotted line in the green quadrants). When measurements of ΔCO-RT and ΔCO-ST disagree with regard to the direction of change, the respective data points will appear in the upper left or lower right quadrant of the plot (red areas). The higher the number of data points in the green quadrants compared with those in the red quadrants, the higher the concordance between the measurement devices.

When measurements of ΔCO obtained with the devices disagree with regard to the direction of change (i.e., the studied technology indicates an increase in CO while the

reference technology indicates a decrease in CO or vice versa), the respective data points will appear in the upper left or lower right quadrant of the plot. These quadrants are therefore marked by red areas in Figure 1. Again, values of the points on the horizontal axis refer to changes indicated by the reference device, and the values on the vertical axis refer to changes indicated by the studied device. Data point 9 reflects a detected change by the reference device of −0.3 and of 2 L/min by the studied device. Situations in which both devices show changes of the same absolute values but in opposite directions are reflected by points on the decreasing 45° diagonal within the red quadrants (the dotted line in the red quadrants in Fig. 1). In data point 1, for example, the reference device measured a change of −2.5 L/min, whereas the studied device showed a positive change by 2.5 L/min.

The higher the number of data points in the green quadrants compared with the number of data points in the red quadrants, the higher is the concordance between the measurement devices. The simplest way to further quantify the level of concordance is to calculate the proportion of data points in the quadrants representing direction of change agreement (green quadrants) in all data points. However, many other concordance measures have been proposed. (We refer to Nelsen[38] for an overview.)

Because no clinically applicable CO measurement system is perfectly accurate and precise, very small changes in CO readings may be attributed to noise and are supposed to not contribute sufficiently to or even disturb trending analysis. Therefore, it was suggested that an exclusion zone be defined at the center of the 4-quadrant plot to remove measurements driven by noise and increase the signal-to-noise ratio. Points of this zone that are also considered to represent clinically insignificant changes are excluded from further analysis. In Figure 1, an exclusion zone for absolute changes below 0.5 L/min is marked by the gray area. Data points 6 and 7 fall in this area and should therefore not be used in the assessment of the trending ability of CO-monitoring technologies. In this example, both points would indicate nonconcordant measurements. However, because of their small absolute values, it is not clear whether they represent real changes in the measurements or are mainly driven by noise.

In summary, the 4-quadrant plot is an intuitive tool to illustrate the trending ability of measurement devices that allows for fast visual assessment of the characteristics of the studied technology and the reference technology. It is important to note that not only the quadrant of a point is important. From the x and y coordinates of a data point, we also obtain information about the magnitude and direction of CO changes of both technologies.

For example, Figure 2 shows 4-quadrant plots for 4 different situations. Clearly, Figure 2A shows a situation with low trending ability. The measurements in this example are, in fact, completely independent of each other. Figure 2B shows a better trending ability, whereas Figure 2D shows a quite good trending ability with only very few discordant measurements. Figure 2C shows a large number of discordant measurements. Here, the studied device tends to indicate changes in the opposite direction of the reference device.

A limitation of the 4-quadrant plot and concordance analysis is the lack of clearly defined cutoff values for the definition of good, acceptable, and poor agreement. Many

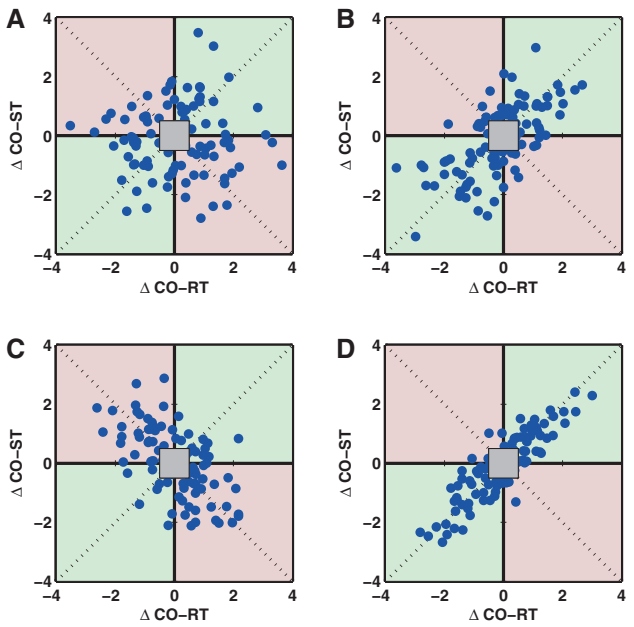

Figure 2. Four-quadrant plots for 4 different situations. Artificial Δcardiac output (ΔCO) values for both the studied technology (ST) and the reference technology (RT) are shown. The values on the horizontal axis refer to ΔCO values of the RT (ΔCO-RT), whereas the vertical axis refers to the ΔCO values of the ST (ΔCO-ST). A, Shows a situation with low trending ability. The measurements in this example are in fact completely independent of each other. B, Shows a better trending ability, whereas (D) shows a quite good trending ability with very few discordant measurements. C, Shows a large number of discordant measurements. Here, the studied device tends to indicate changes in the opposite direction of the reference device.

such values have been suggested previously, but there are no generally accepted thresholds to describe the trending ability of CO measurement technologies. Also, because the results of the 4-quadrant plot analysis depend on the time interval between consecutive measurements, the plot can be influenced by choosing different time intervals for the analysis.

As described above, very small ΔCO values should not be included in the trending analysis, and, thus, a central exclusion zone should be applied. Authors normally use a ΔCO exclusion zone of 0.5 L/min or 10%. However, the exclusion zone should be adapted considering the range of ΔCO values observed in the study population and the time interval between CO readings used for the calculation of ΔCO.

Further, it should be remembered that, in addition to small ΔCO values, very large ΔCO values also might limit the validity of trending analysis. Whether zones excluding very high ΔCO values should be used in the 4-quadrant plot analysis is still a matter of debate.[6]

POLAR PLOT ANALYSIS

Basically, the polar plot by Critchley et al. is methodologically derived from a 4-quadrant plot and is supposed to be a more advanced statistical method for the description of the trending ability of a CO monitor. In the following section, we provide a detailed and critical analysis of this statistical approach by using worked examples. While explaining the individual steps of derivation of the polar plot from the

4-quadrant plot, we simultaneously point out some critical aspects of the polar plot methodology that have not yet been previously described.

In general, the polar plot is based on polar coordinates. This means that every point is addressed by (a) an angle and (b) a radius instead of horizontal and vertical coordinates (x, y). The angle (a) is the angle between the horizontal axis and the line from the point of interest to the central point (0, 0). The radius (b) is the distance of the point of interest to the central point (0, 0). In contrast, in the usual Cartesian coordinate system, points are addressed by their coordinates (x, y) on the horizontal and vertical axes. It is important to notice that both ways of addressing the points are mathematically equivalent and do not affect the position of the points. For example, for data point 5 in Figure 1, the angle between the horizontal axis and the line between (1, 1) and (0, 0) is 45°, and its distance to the point (0, 0) may be calculated by using the formula of Pythagoras to $\sqrt{1^2+1^2}=\sqrt{2}$. Thus, this point can either be described by the coordinates (x = 1, y = 1), where x and y refer to the horizontal and vertical coordinates, respectively, or by the tuple (angle = 45°, radius = $\sqrt{2}$) of angle and radius. Given any of the two, one would find the same point in the plot.

The innovation of the polar plot by Critchley et al. is not only to use the angle and radius to address the points (which would not change the points), but also to transform the points: (a) the angle in the polar plot by Critchley et al. coincides with the angle between the (x, y) to (0,0) line with the 45° diagonal (instead of the horizontal axis) in the 4-quadrant plot, (b) the radius in the polar plot by Critchley et al. is calculated as $\frac{|x+y|}{2}$. The variables x and y refer to the horizontal and vertical coordinates in the 4-quadrant plot again. Figure 3 illustrates the transformation of data points from the 4-quadrant plot to the polar plot by Critchley et al. The left graph is a 4-quadrant plot with the same 9 data points as shown in Figure 1. We did not color the 4 quadrants of the plot as in Figure 1 but added some markings that will be explained later. The right graph of Figure 3 is a polar plot as proposed by Critchley et al.[6] The numbers around the graph denote the values of angles measured from the horizontal axis. The dotted circles mark the radius measured from the center of the plot with values r = 1, 2, and 3 in the plot. Every data point of the 4-quadrant plot is also shown in the polar plot. For example, data point 5 with coordinate (x = 1, y = 1) in the 4-quadrant plot has an angle of 0 with the 45° diagonal line (it happens to lie on this line) and it holds $\frac{|1+1|}{2}=1$. The point is therefore drawn at angle 0 and radius r = 1 in the polar plot. To better illustrate how points transform between the plots, in addition to the blue data points, the 4-quadrant plot shows colored data points that lie in a circle. Transformed to the polar plot by Critchley et al., these points mark 2 circles. The yellow and dark blue points (which lie in the concordance quadrants in the 4-quadrant plot) are transformed to points near the horizontal axis in the polar plot. The red and light blue data points (which lie in the discordance quadrants in the 4-quadrant plot) are transformed to the center of the polar plot. Critchley et al. define an exclusion zone in the polar plot that is marked by the gray circle in the middle of the plot. Points within this area are removed from the analysis. The intention of the exclusion zone is to increase the signal-to-noise ratio. The gray area marked in the 4-quadrant plot on the left side denotes the set of points that would be transformed to the exclusion zone of the polar plot. It is very important to notice that this area not only contains points referring to small changes in the original measurements, but also, the contrary may be true. Points like data point 1, which show a clear discordant behavior of the devices, are mapped to the exclusion zone and are therefore removed from the analysis (also data points 6 and 7 are transformed to this zone). Thus, the exclusion zone of the polar plot excludes

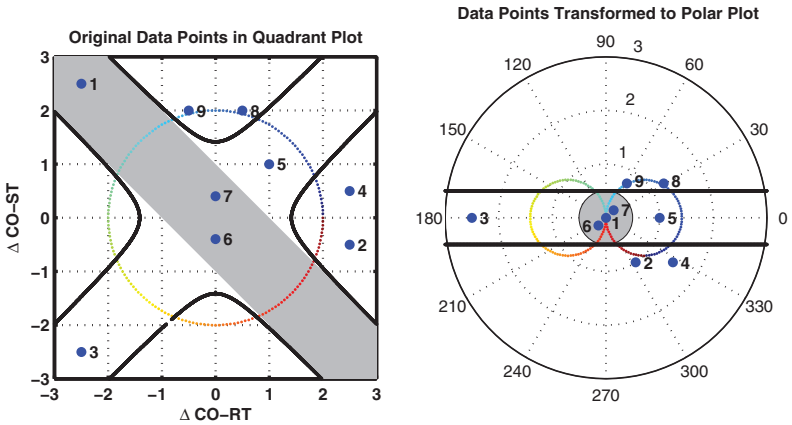

Figure 3. Illustration of the transformation of artificial Δcardiac output (ΔCO) values from the 4-quadrant plot to the polar plot. Left: 4-quadrant plot with 9 data points. The values on the horizontal axis refer to ΔCO values of the reference technology (RT) (ΔCO-RT), whereas the vertical axis refers to the ΔCO values of the studied technology (ST) (ΔCO-ST). Right: polar plot as proposed by Critchley et al.[6] Every data point in the 4-quadrant plot is also shown in the polar plot indicated by the same number. Additionally, the 4-quadrant plot shows colored data points that lie in a circle and correspond to the points with the same color in the 2 circles within the polar plot. The exlusion zone defined by Critchley et al. is marked by the gray circle in the polar plot. The gray area marked in the 4-quadrant plot on the left side denotes the set of points that would be transformed to the exclusion zone of the polar plot. The bold horizontal lines in the polar plot mark an area of points with a high degree of trending capacity. These lines are transformed to the 4-quadrant plot. The areas outside (i.e., above and under) the horizontal lines in the polar plot correspond to the areas in the North, East, South, and West of the bold lines in the 4-quadrant plot, respectively. Note that point 1 (representing a strongly discordant measurement) falls in the exclusion zone of the polar plot.

all measurements from the data where either both changes are small (=the noisy measurements) or the changes are of similar absolute value but contrary in direction (=the most discordant measurements).

Critchley et al. define points under and above certain horizontal lines in the polar plot as reflecting a low degree of trending capacity and the points between these lines as reflecting a high degree of trending capacity. Such horizontal lines are shown in the polar plot in Figure 3. These lines are transformed to the 4-quadrant plot. The areas outside (i.e., above and under) the horizontal lines in the polar plot correspond to the areas in the North, East, South, and West of the black lines in the 4-quadrant plot, respectively. Data points 8 and 9, and 2 and 4 lie within these areas. From the 4-quadrant plot, we can understand what these areas stand for better than we can understand what they stand for from the polar plot: they refer to the cases when one of the devices shows a relatively large change, whereas the other device shows a rather small change. It is again important to notice that the cases of opposite or nearly opposite changes do not fall into these regions but, rather, into the exclusion zone of the polar plot and are therefore removed from the analysis instead of classified as data points reflecting low trending ability.

Critchley et al. propose further refinements of the polar plot methodology. First, they propose not to use horizontal lines as separators between points standing for high and low trending ability, but, rather, lines with angles of +30° and −30° to the horizontal axis. These are shown in Figure 4 together with transformations of these lines into the 4-quadrant plot. It can be seen that these straight lines also correspond to straight lines now and that the angle between them is also 2 × 30° = 60°. Points that are classified to indicate low trending ability are those lying in the North or the South of the polar plot (corresponding to North-West and to South-East in the 4-quadrant plot). Note, however, that the points in the gray area of the 4-quadrant plot fall again into the exclusion zone of the polar plot and are therefore removed from the analysis. Similar to the bias and limits of agreement that are used in the Bland-Altman method,[39] Critchley et al. propose the angular bias and the radial limits of agreement to assess trending ability.[9] The angular bias is the mean of all polar angles from a set of data points and the radial limits of agreement is described as the radial sector that contains 95% of the data points.

To set guidelines for good trending ability, clinical data from different CO measurement method comparison studies were analyzed. Based on these analyses, Critchley et al. defined an angular bias <±5° and radial limits of agreement <±30° for good trending ability. The idea of radial limits may lead to the misunderstanding that all points with the same angle in a polar plot reflect the same trending ability. Note, however, that this is not the case and that points with the same angle in the polar plot but different radii may indeed represent very different levels of trending abilities.

A further refinement of the polar plot is to turn all data points that lie on the left-hand side of the graph by 180° (half-circle polar plot). This is shown in Figure 5, where the circle of colored points in the 4-quadrant plot is transformed to 2 circles lying on the right-hand side of the polar plot (for convenience, we use 2 half-circles with different radii to avoid overlapping of the circles in the polar plot). Again, data points between the 30° lines in the East of the polar plot are classified to denote high trending ability, whereas data points outside this area are classified to reflect low trending ability. Again, all points of the gray area in the 4-quadrant plot are transformed to the exclusion zone of the polar plot and are therefore removed from further analysis.

FOUR-QUADRANT PLOT OR POLAR PLOT ANALYSIS—WHICH METHOD SHOULD BE USED?

The aim of this article is to contribute to a better understanding of the 4-quadrant plot and polar plot as statistical methods referring to tracking changes in CO. For a better overview, the advantages and drawbacks of the 2 techniques are summarized in Table 1. Therefore, the awareness of the major critical aspects of the newer polar plot method

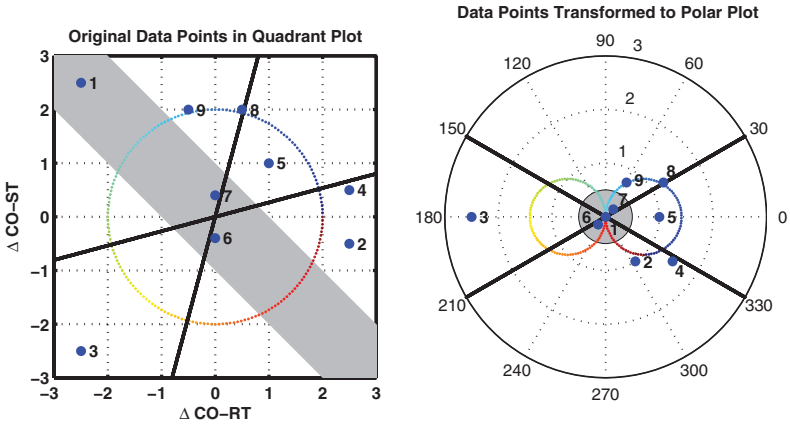

Figure 4. Illustration of the transformation of artificial Δcardiac output (ΔCO) values from the 4-quadrant plot (left) to the polar plot (right). Every data point in the 4-quadrant plot is also shown in the polar plot indicated by the same number or color. The exlusion zone defined by Critchley et al. is marked by the gray circle in the polar plot. The gray area marked in the 4-quadrant plot on the left side denotes the set of points which would be transformed to the exclusion zone of the polar plot. A further refinement of the polar plot methodology is to use lines with angles of +30° and −30° to the horizontal axis to mark an area of points with a high degree of trending capacity. These straight lines correspond to straight lines in the 4-quadrant plot. The angle between them is also 2 × 30° = 60°. Points that are classified to reflect low trending ability are those lying in the North or South of the polar plot (corresponding to North-West and to South-East in the 4-quadrant plot). RT = reference technology; ST = studied technology.

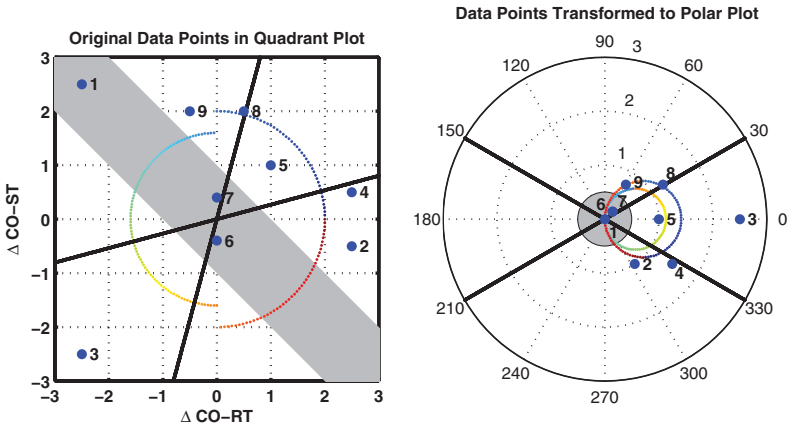

Figure 5. Illustration of the transformation of artificial Δcardiac output (ΔCO) values from the 4-quadrant plot (left) to the polar plot (right). Every data point in the 4-quadrant plot is also shown in the polar plot indicated by the same number or color. The exlusion zone defined by Critchley et al. is marked by the gray circle in the polar plot. The gray area marked in the 4-quadrant plot on the left side denotes the set of points which would be transformed to the exclusion zone of the polar plot. A further refinement of the polar plot is to turn all data points that lie on the left-hand side of the graph by 180° (half-circle polar plot). The 2 half-circles of colored points in the 4-quadrant plot are transformed to 2 circles lying on the right-hand side of the polar plot. Data points between the 30° lines in the East of the polar plot are classified as reflecting high trending ability, whereas those outside this area are classified as indicating a low trending ability. RT = reference technology; ST = studied technology.

is crucial. To sum up these aspects, the polar plot transforms the original data points in a nonlinear way, which makes identification of the actual situations leading to single data points difficult. In addition, the nonlinear transformation leads to the exclusion of parts of the data by mapping them into the so-called exclusion zone: the exclusion zone of the polar plot excludes all measurements from the data not only where both changes are small (=the noisy measurements), but also where the changes are of similar absolute value but contrary in direction, which means the most discordant measurements that correspond to the low trending ability of the 2 technologies.

To further illustrate, 2 simultaneously measured ΔCO data points with the same absolute value (or a close absolute value) but changing in opposite directions (e.g., device 1: +1 L/min and device 2: −1 L/min) are set to 0 (or fall in the exclusion zone) and are therefore ignored in the polar plot analysis, although they might be of high relevance for evaluating the trending ability of a CO monitor. This becomes obvious because there are no data points on or close to the 90° or 270° line, respectively, in the polar plot. (To further illustrate these properties and limitations, a number of studies are available to the interested reader that evaluate the trending ability of 2 CO measurement technologies and

Table 1. Comparison Between the 4-Quadrant Plot and the Polar Plot Methodology

	Four-quadrant plot	Polar plot
Interpretability	• Intuitive picture of the analyzed data • The magnitude and direction of changes of both CO measurement technologies can directly be read off and compared • Determination of the level of concordance is possible by calculating the proportion of data points in the quadrants representing the direction of change agreement	• Demands a high level of insight into its construction to adequately interpret the characteristics of the analyzed data • With the refinement to turn all data points on the left-hand side of the polar plot by 180°, interpretability becomes more difficult
Exclusion zone	• Increases signal-to-noise ratio • Measurements where both changes are small are excluded (noisy measurements)	• Increases signal-to-noise-ratio • In addition to measurements where both changes are small (noisy measurements), also measurements where the changes are of similar absolute value but contrary in direction are excluded (=the most discordant measurements)
Time-series analysis	• Only changes between consecutively measured data points are considered	• Only changes between consecutively measured data points are considered
Measures that can be derived from the plot	• Concordance • Angular bias (see Appendix 2) • Radial limits of agreement (see Appendix 2)	• Angular bias • Radial limits of agreement
Cutoff values	• Up to now, lack of clearly defined cutoff values for the definition of good, acceptable, and poor agreement in the existing literature • The existing cutoff values[9] for the angular bias (<±5°) and radial limits of agreement (<±30°) can be used	• Cutoff values for the angular bias (<±5°) and radial limits of agreement (<±30°) have been suggested before[9]

CO = cardiac output.

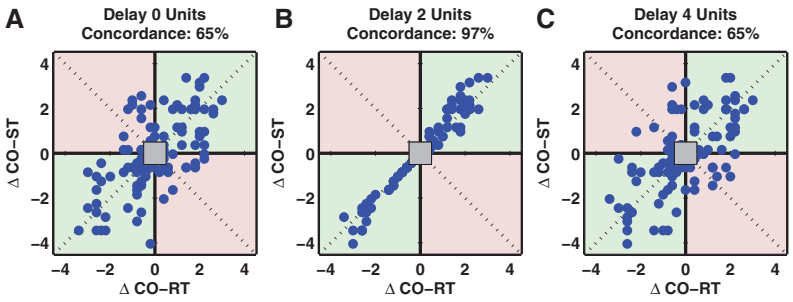

Figure 6. Illustration of the approach to correct for a time delay between 2 cardiac output (CO) measurement technologies using artificial ΔCO values. A, Shows a 4-quadrant plot for the original data. B and C, Show analyses in which the ST is delayed by 2 and 4 time units, respectively. Clearly, in (B), the concordance rate and, therefore trending capacity, is highest. RT = reference technology; ST = studied technology.

apply both the 4-quadrant plot and the polar plot methodology.[40]) Furthermore, adequate interpretation of the trending ability of a CO monitor by using a polar plot is highly demanding. With the refinement to turn all data points on the left-hand side of the polar plot by 180°, interpretability becomes even more difficult. With regard to the relatively simple applicability and interpretability of the 4-quadrant plot, the question arises as to whether the complexity of the polar plot is justifiable.

EXTENSIONS AND PERSPECTIVES

In this section, we address the challenge of trending analysis in patient groups with inhomogeneous levels of CO values.

An important consideration when evaluating the trending ability of the studied CO measurement technology is the fact that the levels of CO values within the studied patient population might lie in a broad range (e.g., between 2 and 15 L/min). If the range of CO values of the different patients within the studied patient group is small, it is perfectly appropriate to use absolute values of CO changes to compare the reference and the studied CO measurement technologies. However, if the range of CO values within the studied patient population is broad, it might be more appropriate to use relative CO changes in the comparison analysis. For example, an absolute decrease in CO of 2 L/min has more importance for a patient with CO values around 3 L/min than for a patient with CO values around 12 L/min. Therefore, if the range of CO values of the different patients within the studied patient group is broad, the CO changes are more comparable if one uses relative CO changes.

Such proportional analyses have also been applied on different topics dealing with inhomogeneous patient data.[41–43]

Next, we would like to address 2 possible extensions of the 4-quadrant plot method. First, we show that measures usually derived from the polar plot (the angular bias and the radial limits of agreement) may as well be derived from the 4-quadrant plot. Second, we apply the 4-quadrant plot method for the case of delayed trending ability, that is, when one of the CO measurement devices reacts faster to CO changes than the other.

The angular bias (<±5°) and radial limits of agreement (<±30°) are useful measures to rate and compare the trending ability of different CO measurement technologies and different studies. However, we do not necessarily need the

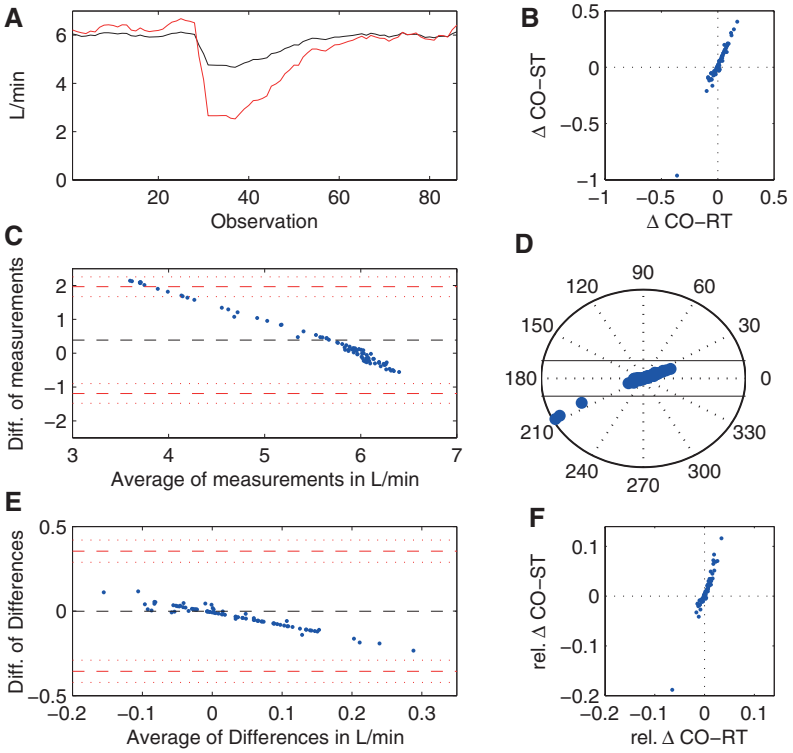

Figure 7. Illustration to exemplarily describe the Bland-Altman analysis in the context of trending analysis, using artificial cardiac output (CO) values. A, Shows CO measurements, both RT and ST derived, done every 5 minutes. Four-quadrant plots based on absolute (B) and relative (F) numbers are shown. The data points in the 4-quadrant plots indicate an almost perfect linear relation of the RT and ST. The corresponding polar plot is given in (D). However, the directions of the changes cannot be seen from the Bland-Altman plots (C and E). The black dashed horizontal line in (C) and (E) shows the mean of the differences (=bias) between the 2 methods, the red dashed horizontal lines show the upper and lower 95% limits of agreement (1.96 × SD), and the dotted horizontal lines show the confidence bands for the limits of agreement. RT = reference technology; ST = studied technology.

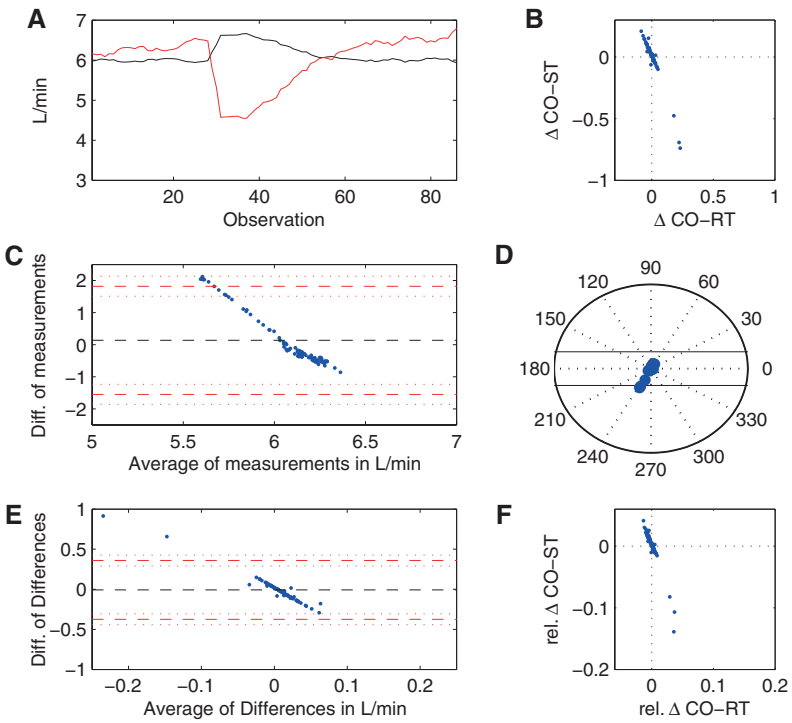

Figure 8. Illustration to exemplarily describe the Bland-Altman analysis in the context of trending analysis, using artificial cardiac output (CO) values. A, Shows CO measurements, both RT and ST derived, done every 5 minutes. Four-quadrant plots based on absolute (B) and relative (F) numbers are shown. The data points in the 4-quadrant plots indicate a poor trending ability. The corresponding polar plot is given in (D). C and E, Show Bland-Altman plots, demonstrating that distinguishing the current scenario from the one shown in Figure 7 is not possible. The black dashed horizontal line in (C) and (E) shows the mean of the differences (=bias) between the 2 methods, the red dashed horizontal lines show the upper and lower 95% limits of agreement (1.96 × SD), and the dotted horizontal lines show the confidence bands for the limits of agreement. RT = reference technology; ST = studied technology.

complexity and the drawbacks of the polar plot methodology to calculate the 2 measures. Angular bias is the mean of the angles in the polar plot. Because the angles in the polar plot correspond basically to the angles of the (0, 0; x, y) line to the 45° line in the 4-quadrant plot, this number can also be calculated from the points of the 4-quadrant plot. The radial limit is the symmetric angle around the 45° line in which 95% of the data points fall and can also be calculated from the points of the 4-quadrant plot. A detailed description of the calculation of the 2 measures is provided in Appendix 2.

The second extension of the 4-quadrant plot methodology considers the following scenario. Both the reference and the studied technologies detect the same changes in CO. However, the studied technology detects these changes with a time delay. If the time delay is small enough, it could be neglected with regard to clinical relevance. However, this clinically irrelevant time delay might have a serious impact on the concordance analysis and thus the measurement performance of the studied technology might be misjudged. Similar to cross-correlation analysis with time delay that is used, say, by engineers in signal analysis, the 4-quadrant plot methodology can be adapted to this case. Therefore, we shift the time series of measurements derived from the studied technology by a certain time lag and again perform our analyses. That is, we calculate the corresponding concordance rate and create the 4-quadrant plot for the shifted studied technology's time series. Figure 6 illustrates this for artificial data. Figure 6A shows a 4-quadrant plot for the original data, that is, simultaneously recorded data pairs of CO values. Figure 6, B and C, show analyses in which the studied device is delayed by 2 and 4 time units, respectively. Clearly, in the second case (Fig. 6B), trending capacity is highest. The concordance coefficients are 0.65, 0.97, and 0.65.

Therefore, we suggest extending the trending analysis in CO measurement by accounting for a potential time delay. That is, we suggest defining a clinically acceptable range of time delays before the beginning of the clinical study comparing the trending ability of 2 CO measurement devices. When analyzing the obtained CO data, separate analyses must be performed for each time lag within the previously defined range of time delays. The final evaluation of the measurement performance of the studied technology should then be based on the time lag with the highest agreement.

CONCLUSIONS

In comparison with the 4-quadrant plot analysis, there is an absence of definite advantages combined with a complex interpretability of the polar plot method. Therefore, the polar plot analysis can currently not be recommended as the superior method for the statistical evaluation of trending ability of a CO monitor compared with the 4-quadrant plot. Furthermore, accounting for a potential time delay between 2 CO measurement technologies is an important aspect in the field of trending analysis that needs to be addressed.

DISCLOSURES

Name: Bernd Saugel, MD.
Contribution: This author was responsible for the intellectual content of the article and manuscript preparation.
Attestation: Bernd Saugel approved the final manuscript.
Name: Oliver Grothe, PhD.
Contribution: This author was responsible for the intellectual content of the article and manuscript preparation.
Attestation: Oliver Grothe approved the final manuscript.
Name: Julia Y. Wagner, MD.
Contribution: This author was responsible for the intellectual content of the article and manuscript preparation.
Attestation: Julia Y. Wagner approved the final manuscript.
This manuscript was handled by: Franklin Dexter, MD, PhD.

REFERENCES

1. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007;17:571–82
2. Bland JM, Altman DG. Agreed statistics: measurement method comparison. Anesthesiology 2012;116:182–5
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. 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
5. Squara P, Cecconi M, Rhodes A, Singer M, Chiche JD. Tracking changes in cardiac output: methodological considerations for the validation of monitoring devices. Intensive Care Med 2009;35:1801–8
6. Critchley LA, Lee A, Ho AM. A critical review of the ability of continuous cardiac output monitors to measure trends in cardiac output. Anesth Analg 2010;111:1180–92
7. Perrino AC Jr, O'Connor T, Luther M. Transtracheal Doppler cardiac output monitoring: comparison to thermodilution during noncardiac surgery. Anesth Analg 1994;78:1060–6
8. Perrino AC Jr, Harris SN, Luther MA. Intraoperative determination of cardiac output using multiplane transesophageal echocardiography: a comparison to thermodilution. Anesthesiology 1998;89:350–7
9. Critchley LA, Yang XX, Lee A. Assessment of trending ability of cardiac output monitors by polar plot methodology. J Cardiothorac Vasc Anesth 2011;25:536–46
10. Ganz W, Donoso R, Marcus HS, Forrester JS, Swan HJ. A new technique for measurement of cardiac output by thermodilution in man. Am J Cardiol 1971;27:392–6
11. Swan HJ, Ganz W, Forrester J, Marcus H, Diamond G, Chonette D. Catheterization of the heart in man with use of a flow-directed balloon-tipped catheter. N Engl J Med 1970;283:447–51
12. Sakka SG, Reinhart K, Meier-Hellmann A. Comparison of pulmonary artery and arterial thermodilution cardiac output in critically ill patients. Intensive Care Med 1999;25:843–6
13. Sakka SG, Reinhart K, Wegscheider K, Meier-Hellmann A. Is the placement of a pulmonary artery catheter still justified solely for the measurement of cardiac output? J Cardiothorac Vasc Anesth 2000;14:119–24
14. Marx G, Schuerholz T, Sümpelmann R, Simon T, Leuwer M. Comparison of cardiac output measurements by arterial transcardiopulmonary and pulmonary arterial thermodilution with direct Fick in septic shock. Eur J Anaesthesiol 2005;22:129–34
15. Linton RA, Band DM, Haire KM. A new method of measuring cardiac output in man using lithium dilution. Br J Anaesth 1993;71:262–6
16. Linton R, Band D, O'Brien T, Jonas M, Leach R. Lithium dilution cardiac output measurement: a comparison with thermodilution. Crit Care Med 1997;25:1796–800
17. Cecconi M, Dawson D, Grounds RM, Rhodes A. Lithium dilution cardiac output measurement in the critically ill patient: determination of precision of the technique. Intensive Care Med 2009;35:498–504
18. Gödje O, Höke K, Goetz AE, Felbinger TW, Reuter DA, Reichart B, Friedl R, Hannekum A, Pfeiffer UJ. Reliability of a new algorithm for continuous cardiac output determination by pulse-contour analysis during hemodynamic instability. Crit Care Med 2002;30:52–8
19. Felbinger TW, Reuter DA, Eltzschig HK, Bayerlein J, Goetz AE. Cardiac index measurements during rapid preload changes: a comparison of pulmonary artery thermodilution with arterial pulse contour analysis. J Clin Anesth 2005;17:241–8
20. Chakravarthy M, Patil TA, Jayaprakash K, Kalligudd P, Prabhakumar D, Jawali V. Comparison of simultaneous estimation of cardiac output by four techniques in patients undergoing off-pump coronary artery bypass surgery—a prospective observational study. Ann Card Anaesth 2007;10:121–6
21. Hamzaoui O, Monnet X, Richard C, Osman D, Chemla D, Teboul JL. Effects of changes in vascular tone on the agreement between pulse contour and transpulmonary thermodilution cardiac output measurements within an up to 6-hour calibration-free period. Crit Care Med 2008;36:434–40
22. De Backer D, Marx G, Tan A, Junker C, Van Nuffelen M, Hüter L, Ching W, Michard F, Vincent JL. Arterial pressure-based cardiac output monitoring: a multicenter validation of the third-generation software in septic patients. Intensive Care Med 2011;37:233–40
23. Pittman J, Bar-Yosef S, SumPing J, Sherwood M, Mark J. Continuous cardiac output monitoring with pulse contour analysis: a comparison with lithium indicator dilution cardiac output measurement. Crit Care Med 2005;33:2015–21
24. Button D, Weibel L, Reuthebuch O, Genoni M, Zollinger A, Hofer CK. Clinical evaluation of the FloTrac/Vigileo system and two established continuous cardiac output monitoring devices in patients undergoing cardiac surgery. Br J Anaesth 2007;99:329–36
25. Pugsley J, Lerner AB. Cardiac output monitoring: is there a gold standard and how do the newer technologies compare? Semin Cardiothorac Vasc Anesth 2010;14:274–82
26. Bein B, Worthmann F, Tonner PH, Paris A, Steinfath M, Hedderich J, Scholz J. Comparison of esophageal Doppler, pulse contour analysis, and real-time pulmonary artery thermodilution for the continuous measurement of cardiac output. J Cardiothorac Vasc Anesth 2004;18:185–9
27. Spiess BD, Patel MA, Soltow LO, Wright IH. Comparison of bioimpedance versus thermodilution cardiac output during cardiac surgery: evaluation of a second-generation bioimpedance device. J Cardiothorac Vasc Anesth 2001;15:567–73
28. Sageman WS, Riffenburgh RH, Spiess BD. Equivalence of bioimpedance and thermodilution in measuring cardiac index after cardiac surgery. J Cardiothorac Vasc Anesth 2002;16:8–14
29. Gujjar AR, Muralidhar K, Banakal S, Gupta R, Sathyaprabha TN, Jairaj PS. Non-invasive cardiac output by transthoracic electrical bioimpedance in post-cardiac surgery patients: comparison with thermodilution method. J Clin Monit Comput 2008;22:175–80
30. Chakravarthy M, Rajeev S, Jawali V. Cardiac index value measurement by invasive, semi-invasive and non invasive techniques: a prospective study in postoperative off pump coronary artery bypass surgery patients. J Clin Monit Comput 2009;23:175–80
31. Squara P, Denjean D, Estagnasie P, Brusset A, Dib JC, Dubois C. Noninvasive cardiac output monitoring (NICOM): a clinical validation. Intensive Care Med 2007;33:1191–4
32. Marik PE. Noninvasive cardiac output monitors: a state-of the-art review. J Cardiothorac Vasc Anesth 2013;27:121–34
33. Broch O, Renner J, Gruenewald M, Meybohm P, Schöttler J, Caliebe A, Steinfath M, Malbrain M, Bein B. A comparison of the Nexfin® and transcardiopulmonary thermodilution to estimate cardiac output during coronary artery surgery. Anaesthesia 2012;67:377–83
34. Westerhof N, Elzinga G, Sipkema P. An artificial arterial system for pumping hearts. J Appl Physiol 1971;31:776–81
35. Westerhof N, Lankhaar JW, Westerhof BE. The arterial Windkessel. Med Biol Eng Comput 2009;47:131–41
36. Truijen J, van Lieshout JJ, Wesselink WA, Westerhof BE. Noninvasive continuous hemodynamic monitoring. J Clin Monit Comput 2012;26:267–78
37. Saugel B, Meidert AS, Langwieser N, Wagner JY, Fassio F, Hapfelmeier A, Prechtl LM, Huber W, Schmid RM, Gödje O. An autocalibrating algorithm for non-invasive cardiac output determination based on the analysis of an arterial pressure waveform recorded with radial artery applanation tonometry: a proof of concept pilot analysis. J Clin Monit Comput 2014;28:357–62
38. Nelsen R. Concordance and copulas: a survey. In: Cuadras C, Fortiana J, Rodriguez-Lallena J, eds. Distributions with Given Marginals and Statistical Modelling. Dordrecht, The Netherlands: Springer, 2002:169–77
39. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10
40. Bubenek-Turconi SI, Craciun M, Miclea I, Perel A. Noninvasive continuous cardiac output by the Nexfin before and after preload-modifying maneuvers: a comparison with intermittent thermodilution cardiac output. Anesth Analg 2013;117:366–72
41. Ledolter J, Dexter F. Analysis of interventions influencing or reducing patient waiting while stratifying by surgical procedure. Anesth Analg 2011;112:950–7

42. Ledolter J, Dexter F, Epstein RH. Analysis of variance of communication latencies in anesthesia: comparing means of multiple log-normal distributions. Anesth Analg 2011;113:888–96
43. Wachtel RE, Dexter F, Epstein RH, Ledolter J. Meta-analysis of desflurane and propofol average times and variability in times to extubation and following commands. Can J Anaesth 2011;58:714–24
44. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995;346:1085–7
45. Myles PS, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. Br J Anaesth 2007;99:309–11
46. Hamilton C, Lewis S. The importance of using the correct bounds on the Bland-Altman limits of agreement when multiple measurements are recorded per patient. J Clin Monit Comput 2010;24:173–5
47. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60

APPENDIX 1
Bland-Altman Analysis in the Context of Trending Analysis

The Bland-Altman analysis is the leading method to assess the accuracy and precision between a reference and a studied technology in cardiac output (CO) measurement comparison studies. In its simplest form, the Bland-Altman plot quantifies how much the reference (e.g., the gold standard) and studied technology may deviate from each other. To this end, it provides boundaries such that 95% of the nonsystematic differences between measurements from reference and studied technology lie within these boundaries (the exact value of 95% holds only if the differences are normally distributed random variables, otherwise it holds only for approximately 95% of the data). If the calculated boundaries of possible differences between the methods are too large to be clinically negligible, the researcher decides that the 2 methods are not interchangeable. However, if the boundaries refer to differences between the methods that are clinically negligible, the Bland-Altman analysis rates them as equally good and thus interchangeable.

To address this mathematically, let $g_j^i, s_j^i, i \in \{1,\ldots,N^p\}, j \in \{1,\ldots,N_i\}$, be the values of the gold standard method (g) and the studied technology (s), respectively, where N^p denotes the number of patients and N_i denotes the number of measurements per patient i. The Bland-Altman plot displays $\frac{g_j^i + s_j^i}{2}$ on the x-axis and $g_j^i - s_j^i$ on the y-axis for $i \in \{1,\ldots,N^p\}, j \in \{1,\ldots,N_i\}$. It is common in the scientific literature to use $\frac{g_j^i + s_j^i}{2}$ for the x-axis values in the Bland-Altman plot rather than only the gold standard g_j^i. In their work published in 1995, Bland and Altman illustrate with an example why plotting the difference between g_j^i and s_j^i against the standard method instead of the average of g_j^i and s_j^i is misleading.[44]

One then considers the mean and SD of the differences $g_j^i - s_j^i$ on the y-axis for $i \in \{1,\ldots,N^p\}, j \in \{1,\ldots,N_i\}$.

Now it follows that if the differences are normally distributed random variables (which is a reasonable assumption for measurement errors) and independent, 95% of the differences lie between mean ($\bar{D}$) minus 1.96 multiplied by SD and mean ($\bar{D}$) plus 1.96 multiplied by SD, which are the so-called limits of agreement (see Bland and Altman 1986).[39] Here, the mean refers to the systematic difference (if we know the mean, we could simply correct for this difference by subtracting it from the measurements of the studied technology) and 2 × 1.96 = 3.92 multiplied by the SD to the boundaries for nonsystematic differences, respectively. These numbers hold if the data are independent and the true SD is known. Having to estimate the SD from the data, however, one should provide confidence bands for the estimates, that is, for the estimated limits of agreement. To this end, Bland and Altman estimate the variance of the limits of agreement by $\mathrm{Var}\left(\bar{D} \pm 1.96 \times s_d\right) = \left(\frac{1}{n} + \frac{1.96^2}{2\times(n-1)}\right) \times s_d^2$, where s_d^2 is the estimated variance of the differences above. Confidence bands for the limits of agreement will be t multiplied by $\sqrt{Var(\bar{D} \pm 1.96 \times s_d)}$ to each side, where t is the 95% quantile of the t distribution with n – 1 degrees of freedom. Note that this formula holds only for independent data.

Often the data are not independent and identically distributed, for example, when >1 measurement in 1 individual is performed (see Bland and Altman 2012 for an overview).[2] Extensions and improvements of the Bland-Altman analysis therefore deal in particular with the calculation of the SD and the limits above in such cases.[45,46] Usually, one distinguishes between repeated and nonrepeated measurements. The term repeated measurements refers to a scenario in which we perform >1 measurement in the same patient. In this scenario, we have to further distinguish between cases in which the true value is constant and cases in which the true value varies over time (e.g., when measuring a patient's CO). In the latter case, we are using the repeated measurements to monitor, the patient's CO, for instance. If the true value is constant, repeating the measurement results in a higher precision of the measurements because the measurement errors average out across the measurements. If the true value varies over time, we do not increase the precision with each measurement. Furthermore, it is important whether we perform measurements in only 1 patient or in several patients. The variability of the measurements may be different from 1 patient to another (so called fixed effect on the patient). Again, denote the measurements with the 2 methods by g (gold standard) and by s (studied technology) and note that we are interested in the variance of D = g − s. Partitioning of the variances of each method leads to $\mathrm{Var}(g) = \sigma_t^2 + \sigma_{gI}^2 + \sigma_{gw}^2$ and $\mathrm{Var}(s) = \sigma_t^2 + \sigma_{sI}^2 + \sigma_{sw}^2$, where σ_t^2 is the variance due to changes in true value over time, σ_{gI}^2 and σ_{sI}^2 are variances due to different variations from patient to patient (fixed effects), and σ_{gw}^2 and σ_{sw}^2 are the variances of the measurement errors. The estimation of the variances of g and s is based on estimating all these single terms.

For the sake of simplicity, we now focus on the case where we only monitor 1 patient (so $\sigma_{gI}^2 = \sigma_{sI}^2 = 0$). The true value, however, varies over time (so σ_t^2 is not equal to 0). Because we do not know the true value, we cannot estimate σ_t^2 without further information. But, if we look at the time series of the differences D = g − s instead, the variation of

the true value cancels out, and we find $\mathrm{Var}(g - s) = \sigma_{gw}^2 + \sigma_{sw}^2$, which is the sum of the variances of the measurement errors without further effects. Bland and Altman assume for this case (see, e.g., Bland and Altman, 1999, section 5.3[47]) that the single measurements are independent, so that we can estimate Var(g – s) by calculating the SD between the differences D = g – s of the measurements. Furthermore, we can calculate the confidence bands for the limits of agreement as stated above for the independence case.

Before applying the Bland-Altman plot for trending analysis, we should recall the essence of trending analysis. Two methodologies show a good trending if they react on changes in the same direction. The idea is that both methods could be used to monitor and stabilize a patient.

When applying Bland-Altman analysis in this context, there are generally 2 different possibilities. First, one could apply it directly to the measurements as summarized above, and second one could apply the Bland-Altman plot to $g_j^i - g_{j-1}^i$, $s_j^i - s_{j-1}^i$ for $i \in \{1,\ldots,N^p\}, j \in \{2,\ldots,N_i\}$, that is, to the changes measured by each methodology instead of the actual measurements.

In Figure 7, we visualize both ways for a simulation example. The data points are deliberately chosen to illustrate the challenges of the Bland-Altman plot in the context of trending analysis. Multiple measurements in 1 subject are shown. Figure 7A shows CO measurements done every 5 minutes while monitoring a patient who has been stable in CO for longer. Suddenly, the patient's CO decreases. The CO then increases again after some intervention. It can be seen from the graph that both CO measurement devices detect the CO change but react with different sensitivity. However, both devices are suitable for detecting a patient's decrease in CO. This is illustrated in the 4-quadrant plot (Fig. 7B). All data points lie on a straight line in the first and third quadrant of the plot, indicating an almost perfect linear relation of the reference technology and the studied technology, which, therefore, always detect the same direction of CO change. We additionally plotted a polar plot by Critchley et al. (Fig. 7D) and a 4-quadrant plot based on relative numbers (Fig. 7F) as discussed in Extensions and Perspectives. Figure 7C shows a Bland-Altman plot. Figure 7E shows a Bland-Altman plot applied to changes of the measurement. Now, the x-axis refers to the average of the detected changes of both devices at the same time, whereas the y-axis refers to the difference of the detected changes. The directions of the changes cannot be seen from the Bland-Altman plot (Fig. 7C) and the 4-quadrant plot should be consulted.

Figure 8 emphasizes this point. In the figure, the patient's decrease in CO is detected with the reference technology (gold standard method), whereas the studied technology always detects the changes in the opposite direction (Fig. 8A). The 2 Bland-Altman plots (Fig. 8, C and E) report similar mean differences and 95% limits of agreement as before (and are thus not able to distinguish the current scenario from the one before), whereas the 4-quadrant plot clearly shows the poor trending ability of the devices (Fig. 8, B and F).

Overall, Bland-Altman analysis is an appealing concept for assessing the absolute agreement and precision of technologies. However, if the technologies show a good trending ability but deviate in absolute measures, Bland-Altman analysis cannot confirm trending ability. In these cases, the methods as discussed in this article should additionally be applied.

APPENDIX 2

Calculation of the Measures Angular Bias and Radial Limits of Agreement (Usually Derived from the Polar Plot) from the 4-Quadrant Plot

Let x_i and y_i denote the ΔCO values of the reference and studied device as before. Then the angle between the (0,0)-(xi,yi)-line and the 45° axis is $\vartheta_i = \mathrm{atan2}\left(\frac{x_i}{y_i}\right) \times \frac{180}{\pi - 45}$ where atan2 is a common variation of the arctangent function to cope with the different orthants of the data.

The mean of all these thetas corresponds to the angular bias calculated from the polar plot. It may be depicted in the 4-quadrant plot by a line.

Note that the numbers calculated this way do not necessarily match the numbers as calculated from the polar plot. The reason is the exclusion zone of the polar plot, which excludes also most of the discordant pairs from the data. The drawbacks of the exclusion zone in the polar plot are described in detail in Polar Plot Analysis. However, by excluding the same data points from the analysis, both the angular bias when applying the polar plot methodology and the angular bias derived from the 4-quadrant plot methodology coincide.

The radial limit of agreement is the symmetric angle around the 45° line in which 95% of the data points fall. It may be calculated from the theta values above. Therefore let $\alpha_i, i = 1\ldots n$ denote an ordered set of absolute values of the ϑ_i calculated above beginning with the smallest value. Let $m = \lfloor 0.95 \times n \rfloor + 1$, that is, the 0.95 multiplied by the total number of points rounded to the next natural number. The angle of the radial limit is then the m-th smallest value of the absolute values of ϑ_i as calculated above. Again, by excluding the same data points from the analysis both the radial limit of agreement when applying the polar plot methodology and the radial limit of agreement derived from the 4-quadrant plot methodology coincide.