

Pierre Squara
Maurizio Cecconi
Andrew Rhodes
Mervyn Singer
Jean-Daniel Chiche

Tracking changes in cardiac output: methodological considerations for the validation of monitoring devices

Received: 10 April 2009
Accepted: 18 June 2009
Published online: 11 July 2009
© Springer-Verlag 2009

J.-D. Chiche (✉)
Réanimation Médicale, Hôpital Cochin,
Assistance Publique-Hôpitaux de Paris,
Université Paris Descartes, 27,
Rue du Faubourg Saint-Jacques,
75014 Paris, France
e-mail: jean-daniel.chiche@cch.aphp.fr

P. Squara
Réanimation CERIC, Clinique Ambroise
Paré, 27, Boulevard Victor Hugo,
92200 Neuilly, France

M. Cecconi
Department of Anesthesia and Intensive
Care, Azienda Ospedaliero Universitaria
Udine, Piazzale Santa Maria della
Misericordia, 33100 Udine, Italy

A. Rhodes
Department of Intensive Care Medicine,
St George's Hospital,
London SW17 0QT, UK

M. Singer
Department of Intensive Care Medicine,
University College London,
Cruciform Building, Gower St,
London WC1E 6BT, UK

Abstract *Background:* Until now, tools for continuous cardiac output (CO) monitoring have been validated as if they were tools for snapshot measurements. Most authors have compared variations in cardiac output between two time-points and used Bland–Altman representations to describe the agreement between these variations. The impacts of time and of repetitive measurements over time are not taken into consideration. *Purpose:* This special article proposes a conceptual framework for the validation of CO monitoring devices. Four quality criteria are suggested and studied: (1) accuracy (small bias), (2) precision (small random error of measurements), (3) short response

time and (4) accurate amplitude response. Because a tolerance is obviously admitted for each of these four criteria, we propose to add as a fifth criterion the ability to detect significant CO directional changes. Other important issues in designing studies to validate CO monitoring tools are reviewed: choice of patient population to be studied, choice of the reference method, data acquisition method, data acceptability checking, data segmentation and final evaluation of reliability. *Conclusion:* Application of this framework underlines the importance of precision and time response for clinical acceptability of monitoring tools.

Keywords Cardiovascular monitoring · Cardiac output · Methodology · Haemodynamics

Introduction

Measurement of cardiac output (CO) provides important information that can be used for diagnosis and therapeutic optimization of patients with haemodynamic instability. Over the past 20 years, efforts have focused on providing less invasive, continuous CO monitoring devices. With every new device comes the need for validation and evaluation of its usefulness in the clinical setting.

Consideration of time as an independent variable creates differences between measurement and monitoring. Because there are intrinsic differences in information that can be obtained from measurement devices versus monitoring devices, the need for specific validation methods for monitoring tools has been suggested [1]. This special article proposes a conceptual framework for validation of CO monitoring devices that may have some relevance for other types of monitoring.

Quality criteria for monitoring devices' acceptability

To validate a new CO monitoring technology, most authors have compared the CO measurements obtained with a new studied technology (ST) and a reference technology (RT) at two time-points [2]. The Bland–Altman representation, created to compare two technologies when none can be considered as a gold standard, is widely used to describe the agreement between ST and RT [3]. This approach is limited to the estimation of the bias and of the inter-patient variability of the bias [4]. Precision and time response are not taken into consideration. In addition, RT must be considered a priori superior to ST. So, alternative strategies need to be developed [1].

Physicians trust in a given value if the device is accurate and they trust in a change in this value if the device is precise. Thus, an acceptable device for CO monitoring must fulfill the two traditional quality criteria required for CO measurements: high accuracy and high precision. The term “precision” is sometimes improperly used for inter-patient variability of the bias (or “precision” of the bias) illustrated by the Bland–Altman plot. Precision is also sometimes improperly used to describe the global variability of measurement that may include: (1) the true precision of the monitoring system, (2) the physiological intra-patient variability of the measured variable, (3) artefacts, (4) the inter-patient variability when data from different patients are pooled together and (5) the inter-device variability when using different machines. In this paper, “precision” is restricted to its strict metrological definition, i.e., the variability of values due to random errors of measurement. Therefore, precision is an intrinsic system property, independent from bias, and does not necessarily require a reference method to be analyzed (Figs. 1, 2).

In addition to accuracy and precision criteria, the validation of a monitoring device must take into account the changes in the monitored variable over time. Indeed, good time response and accurate response amplitude are

vitaly important qualities. As we are forced to accept a certain margin of error (tolerance) for accuracy, precision, and time response, the cumulated effect of these errors may compromise the clinical utility of the monitoring device. An estimation of the sensitivity and specificity in the detection of significant directional changes is therefore a fifth important criterion in the evaluation of the reliability of a device. This ability to track a disease's progression and therapeutic impact is more dependent on precision and time response than on device accuracy.

Principles of evaluation of monitoring devices

The following steps should be considered: (1) choice of patient population to be studied, (2) choice of reference method, (3) data acquisition method, (4) data acceptability (5) data segmentation and (6) analysis of quality criteria.

Choice of patient population to be studied

In order to generalize the results of a validation study, patients must represent the population where the device is likely to be used. For intensive care, new monitoring devices should be tested in patients with septic and non-septic conditions, spontaneous and mechanical ventilation, treated with vasopressive drugs or vasodilators. In addition, the population studied must have a high likelihood of changes in CO during the study period. This can best be performed by building into the protocol an intervention that should, on its own, change CO, such as an intravenous fluid challenge, a change in PEEP levels or the infusion of drugs likely to induce a change in CO.

Choice of the reference method

Ideally, validation of a ST requires an independent validation of each of the five quality criteria listed above against a specific gold standard that would give instantaneous exact values. However, in practice, we are often limited to assessing the clinical acceptability of an ST by comparison with a widely used RT based on these five quality criteria.

A CO gold standard is necessary only to assess accuracy and amplitude of response of CO monitoring systems. Although enunciated in 1870, the Fick principle [5] is still considered the gold standard for CO measurement in the physiology laboratory [6]. In fact, besides very specific and limited situations such as the use of an extracorporeal circuit or an internal Doppler flow probe [7, 8], a gold standard for CO accuracy is most often

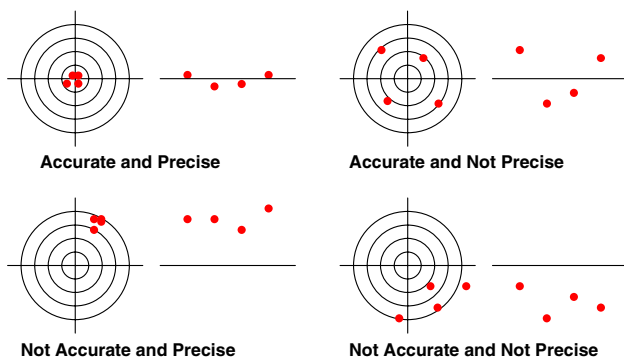


Fig. 1 The device is accurate when the average is close to the target center. The device is precise when all measurements are close together (adapted with permission [42])

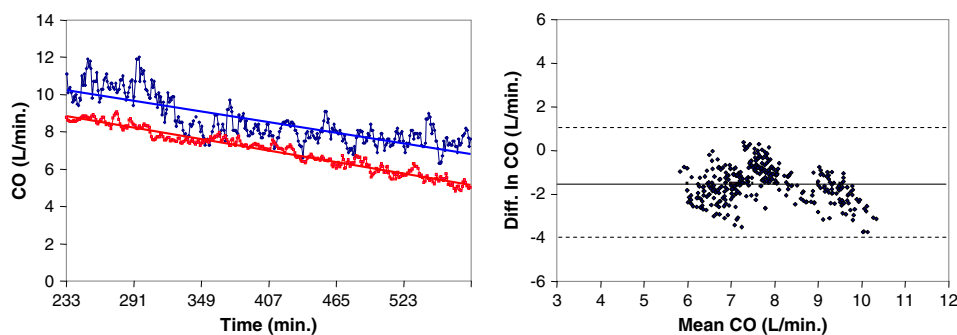


Fig. 2 Left panel CO trend line. Red line ST curve, Black line RT curve. Right panel Bland–Altman representation. Since the slopes are quite parallel (ST 53%, RT 40%) the bias can be seen from the distance between the two slopes in the Y axis (left panel) or from the averaged difference between the values (-1.53 L/min in the right panel). The precision, namely the variability around the slope is smaller for ST (7%) than for RT (27%; left panel). Due to the

slope and the large variability of the RT, the percentage of error (2SD of the difference/RT mean value) as given by a Bland–Altman plot is large, ± 2.54 L or $\pm 30\%$; this indicates a large variability of the bias but tells us nothing about the real ST precision because bias and precision in both techniques are mixed together

lacking. A widely used technology known to provide acceptable accuracy can then be taken as RT. In clinical practice, the bolus thermodilution technique has been widely accepted as a reference method [9–13]. However, it requires manual intervention, therefore leading to intra- and inter-investigator variability [11]. In any case the comparison between an ST and any RT must be performed using the same time scale. As an example, when using bolus thermodilution as RT, a 3–5 min averaging of the ST must be compared to the thermodilution obtained by averaging at least three concordant measurements, which takes a minimum of 3–5 min. Since the standard error of the mean (SEM) due to random errors of measurements depends on the number of measurements averaged, the precision of the RT has to be considered [14]. Thus, the study design should include the determination of the coefficient of variation (2SD/mean) and the coefficient of error (2SEM/mean) for the RT to adequately assess and report the validity of the ST.

Since no CO monitoring device has fulfilled an independent ideal validation of all five quality criteria, there is a lack of consensual monitoring RTs. The clinical global acceptability of an ST must be validated against a comparable system, i.e., real-time, automatic and continuous, with validation data that are completely understood. A CO monitoring device should ideally provide reliable data to track long-term changes and also to provide snapshot information. Although some devices may analyze stroke volume beat-by-beat, a “snapshot measurement” is in reality a signal averaging obtained during a specific period of time. A 1-min averaged value seems adequate to properly assess changes in CO during critical situations or therapeutic tests. When considering a clinically acceptable RT for CO monitoring, continuous cardiac output monitoring using a pulmonary artery catheter (PAC-CCO) allows measurements of CO with both intermittent thermodilution boluses and two different time-averaging

intervals for continuous monitoring. With a poor time response [15] and low precision [16–18], the PAC-CCO method is by no means perfect. Despite these limitations, the PAC-CCO, however, remains the most commonly used RT [1, 19–21] as the averaged bias has been deemed to be acceptable in a wide variety of clinical situations [15, 22–27]. There is an increasing body of evidence that less invasive methods allowing continuous CO monitoring with better time response and better precision will fulfill most of our criteria in the near future [14, 28–32]. However, few of these have as yet been submitted to a thorough analysis as recommended here [1], allowing researchers to reliably assess their limitations for each of the five listed quality criteria. This is of critical importance since any technology can be used as the RT only when researchers are experienced in the technique and well versed in its limitations.

Data acquisition method

A completely automatic, continuous data-collection technique should be used for both RT and ST to avoid errors inherent to collection of large numbers of values and to limit any inter-observer variability. Time sampling may be an issue when comparing two devices using different technologies. From the original signal to the CO value, seen on the screen or stored in a data base, there is some form of complex signal processing specific to each technology. It is unusual in clinical practice for the clinician to modify an instrument’s signal and it is not in the scope of this article to study signal processing. In summary, all manufacturers must compromise between time response and variability. Smaller sampling time is necessarily associated with faster time response but also with larger variability. The final choice is based on intrinsic properties of each technology. As example, for PAC-CCO, the ‘STAT’ value of the

PAC-CCO provides a 1-min running average of random micro-thermodilutions. The value of CO seen on the main screen of the monitor is an average of at least five STAT measurements. As mathematically predicted, averaging more data decreases the variability and the time response but not the accuracy [33]. Ideally, when comparing two devices, the smallest possible time sampling allowed by the two manufacturers should be used for greatest clinical applicability

Data acceptability

Before analyzing the data collected as described above, it is important to validate the data. A list of limitations must be prospectively determined based on the technologies of both ST and RT. For example, periods of time when the patients are agitated, when one or both systems become disconnected, or where there is clear evidence of a situation leading to unrealistic results, must be deleted. This is a critical step of the validation that can be altered by user input. Then, data must be evaluated by an independent assessor who is blinded to the choice of monitoring technology (RT or ST). Situations where unrealistic results are found can be studied separately a posteriori to identify specific limitations of one device.

Data segmentation

When enough data points are available, a cross correlation may show if the trend waveforms of two devices are comparable or not. Additionally, it is sometimes necessary to divide the monitoring trend line into periods of unchanged, increasing or decreasing value. Studying time-periods mixing different slopes in the trends leads to invalidation of the estimation of precision because actual changes in CO contribute to the variability of CO measurements. Moreover, differences in time responsiveness between ST and RT may spuriously increase the bias. We may also imagine that the quality criteria of both ST and RT are not equivalent when the CO is increasing or decreasing, especially when sophisticated smoothing algorithms are used. Thus, accuracy and precision are best analyzed during periods of stable CO.

Easy database segmentation can be performed using the slope of the trend line of the monitored variable. Inflection points between two consecutive slopes may also be automatically determined using the minimum sum of residuals for the two segments proposed by John-Alder [34]. Figure 3 shows the example of a trend of a CO monitoring device as ST and a simultaneously recorded trace of a PAC-CCO used as RT. In this example, the changes in the two variables tracked each other fairly closely during the transitions, though the ST tended to change more quickly than the PAC-CCO.

Analysis of quality criteria

Essential qualities of any new monitoring device include a small bias, a high precision, a short time response and accurate amplitude response. However, for medical decision making, we may have different tolerance thresholds for these criteria. A 30% bias in CO does not necessarily change the therapeutic target [35] whereas a low precision and/or a low responsiveness may hide the occurrence of significant CO changes and may lead to unacceptable delays in therapeutic changes. Optimal assessment of each of these criteria relies on the use of specific statistical methods and mandates a prospective definition of tolerance level.

Bias

The bias between an ST and an RT is the averaged difference in values during a specific period of time. The inter-patient biases can be displayed using a Bland–Altman representation. It can be averaged and reported in L/min, or in % (bias/mean value). The inter-patient variability of the bias can be reported using 2SD of the bias (limits of agreement = $\pm 2SD$) [3], the coefficient of variation (2SD of the bias/mean value) [35] or using the frequency distribution of the relative error (RE = absolute value of the bias/RT mean value) [1].

As differences in delay and amplitude of response between ST and RT may spuriously affect the estimated bias when the monitored variable changes rapidly (Fig. 3), the bias is optimally assessed during periods of time where the monitored variable is stable. For CO, stability over the selected time interval can be defined

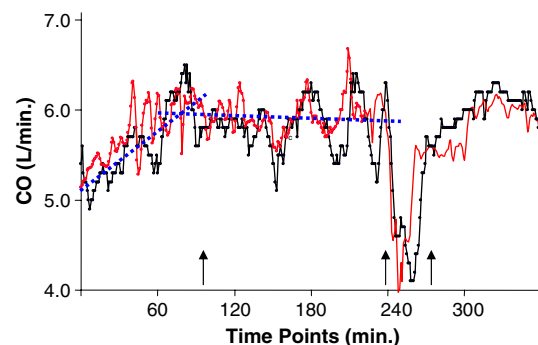


Fig. 3 CO trend line segmentation in different phases. Red line ST curve, Black line RT curve. From min 1–90 (arrow), the RT slope (dotted blue line) is increasing (+25%). From min 90–235, the slope is unchanged (–2%). Between min 235–250, a PEEP test was performed, resulting in a sudden fall in CO (negative challenge), then stopped, resulting in a sudden increase (positive challenge). From min 275–360 the slope is increasing (+20%). After min 235, the slopes are not represented in this figure for better readability. These different periods of time must be analyzed separately (reprinted from [1])

when the RT trend line slope is within $\pm 10\%$ with a precision (defined below) lower than 20%. In this situation, the Standard error of the RT mean value is given by the formula $SEM = SD/\sqrt{n}$. Even if $SD = 20\%$ for the RT, SEM is only 2% when 100 points are averaged during a period of stable CO and RT; averaged value can then be considered as close to the real CO value. The bias can be calculated first for the smallest sampling of time points (i.e., 1 min), and then averaged for each individual patient's period of stability, resulting in an inter-patient equivalence independent of the duration of analysis. Additionally, this provides the reader with two extreme averaging possibilities: the smallest and the longest. Any intermediary time averaging will provide intermediary results.

Tolerance in bias

Measurements of bias reported for the PAC-CCO range between 5 and 15% [9–11]. This is widely considered to be acceptable as the consequences of bias on medical decision making are limited [36] unless clinicians titrate therapy to a specific goal [37] rather than to an individualized target [38]. When comparing different modalities of CO monitoring, an inter-patient coefficient of variation ($2SD/\text{mean}$) lower than 30% has been suggested as an acceptable limit [35]. Others have used more restrictive criteria of acceptability (relative error as compared to the RT $<20\%$) [1]. In addition, the regression line between the ST and RT must be close to the identity line. Even though the global bias is satisfactory, overestimation of low values and underestimation of high values argue for poor detection of significant CO changes [1].

Precision

Precision is high when variability due to random error of measurements is low. We have seen that precision is an intrinsic quality of the system. The limits of agreement of the bias ($2SD$ inter-patients), the coefficient of variation ($2SD/\text{mean}$) and the coefficient of repeatability ($2SD$ intra-patient) given by a Bland-Altman representation are not suitable for estimating the real precision of the ST, unless the RT gives exact values (which is contradictory, as the Bland-Altman representation was created to compare two modalities of measurement where neither is a gold standard). Precision of each modality of measurement can be estimated by the variability ($2SD$) around its own linear trend line slope in order to focus on real random error of measurements (Fig. 2). This is independent from bias, and high or low precision can be found either with good or poor accuracy. As for the analysis of the bias, precision is optimally evaluated during stable phases to minimize the effect of physiological changes on

variability. When the RT slope is flat, the variability around the linear trend is equivalent to the coefficient of variation ($2SD/\text{mean}$). When the slope is not flat, precision is inferior to the coefficient of variation.

Tolerance in precision

Tolerance in precision is also specific for each monitored variable. For example, the in vitro precision of PAC-CCO (SD/mean) measurements have been estimated to be between 9.2 and 11.6% using different devices [11]. In clinical conditions, the reported precision can be as high as 20% [16, 36]. There must, therefore, be an a priori description of tolerance limits for precision according to the researchers' views as to how the monitor would be used. It is worth stressing the implications of precision on the ability of a device to detect change. A measurement with a precision of 20% would only allow the users to detect a change of 27.8% in the underlying signal with 95% certainty [14]. This may be too high to be clinically useful. This would suggest, therefore, that a measurement precision level $<10\%$ is desirable. In monitoring devices, a 10% precision is also desirable but it may be improved by the number of values. Again, since $SEM = 2SD/\sqrt{n}$, a precision of 20% in four successive values is equivalent to a 10% precision in a snapshot measurement.

The response time and amplitude

Response time and amplitude of the ST can be compared to that of the RT at times when acute therapeutic challenges are made. As an example, a sudden drop in CO can be expected when a positive end-expiratory pressure (PEEP) of 20 cmH₂O is applied to the patient. In contrast, in cases where high PEEP is lowered and/or after initiation of an inotrope infusion or passive leg raising, a rapid increase in CO is often seen. To measure the delay and amplitude of changes, the inflexion points in the ST and RT curves after the start and end of the challenge must be determined. The time response of each device is the delay between the start of the challenge and the upper inflexion point. The amplitude of response of each device is the difference of value between the upper and the lower inflexion points. The inflexion point can also be determined automatically using the John-Alder method [34]. However, rapid changes in CO based on few data are influenced by the precision of each monitoring modality. A more independent estimation of the amplitude response may therefore be obtained by comparing the two slopes of CO variation when CO consistently changes during a significant period of time. When the CO is stable or steadily changing, and when an appropriate number of data are considered to reach small SEM (see bias), the RT slope can then be considered as close to the real CO slope.

Another method is to compare the ST with a measurement reference method in at two points in time and to compare the differences obtained with the two technologies.

Tolerance in time and amplitude response

Tolerance in amplitude response should be logically equivalent to tolerance in bias, < 30% as compared to an RT. However, amplitude response must be obtained within an acceptable delay, defined as the tolerance in time response. The best compromise between time and amplitude response depends on the clinical question. In many instances we can be satisfied with a reliable semi-continuous or discontinuous technique and do not really need beat-to-beat stroke volume derivation. In contrast, as when optimizing a compromised status, testing nitric oxygen inhalation or PEEP of PLR or inotropes, a very fast indication of the directional changes is of major importance. Again, the best compromise depends on intrinsic characteristics of the signal, information which is not available to clinicians. The PAC-CCO response time is close to 10 min [15, 39, 40]. This is too long for many hemodynamic challenges used in daily practice, such as passive leg raising [41]. For an optimal RT, it would clearly be desirable to have a time response significantly shorter than 10 min, preferably quantifiable in seconds. A tolerance in time response close to 1 min with a tolerance in CO amplitude greater than 30% seems to be acceptable in most clinical situations.

Evaluation of reliability for indicating directional change

All measurements of each patient can be classified in one of three time-based categories—unchanged, increasing or decreasing trend—according to the data segmentation method described above. Unacceptable discordances in directional changes can be defined as either an excessive difference between the ST and the RT slopes, or a negative or intraclass correlation value. Sensitivity for

directional change is then calculated as $TP/(TP + FN)$ and specificity as $TN/(FP + TN)$ where TP = change in both systems, TN = no change in both systems, FP = change in ST only, and FN = change in RT only.

Tolerance in the reliability for indicating directional changes

A reliable monitoring device must detect a significant change with sensitivity and specificity very close to one.

Conclusion

In a way, comparing monitoring devices to measurement devices is like comparing cinematography to still photography. In monitoring systems, a considerable increase in the available number of data points and the time-dependent additional properties such as time response and ability to identify significant changes allows a larger tolerance in bias. Ideally, monitoring capabilities include snapshot data that are often used by clinicians for medical decision making. This leads us to give critical importance to precision and time response. Regarding the evaluation of novel CO monitoring devices, solutions to the limited availability of gold standards for validation of accuracy and amplitude response can only come from comparison with discrete reference technologies such as bolus thermol dilution in stable situations following the framework proposed, or by a stepwise improvement based on clinical experience.

Conflicts of interest statement P.S. is a consultant for Cheetah Medical and has been a consultant for Edwards. M.C. has no conflict of interest to disclose. M.S. is a consultant for Deltex & Edwards. A.R. has been a consultant for LIDCO, Edwards & Cheetah Medical. J.D.C. has been a consultant for Edwards & Cheetah Medical, and is member of the Medical Advisory Board of GE Healthcare & Massimo.

References

1. Squara P, Denjean D, Estagnasie P, Brusset A, Dib JC, Dubois C (2007) Noninvasive cardiac output monitoring (NICOM): a clinical validation. *Intensive Care Med* 33:1191–1194
2. de Wilde RB, Schreuder JJ, van den Berg PC, Jansen JR (2007) An evaluation of cardiac output by five arterial pulse contour techniques during cardiac surgery. *Anaesthesia* 62:760–768
3. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
4. Cecconi M, Grounds M, Rhodes A (2007) Methodologies for assessing agreement between two methods of clinical measurement: are we as good as we think we are? *Curr Opin Crit Care* 13:294–296
5. Fick A (1870) Ueber die Messung des Blutquantums in den Herzventrikeln. Würzburg
6. Stickland MK, Welsh RC, Haykowsky MJ, Petersen SR, Anderson WD, Taylor DA, Bouffard M, Jones RL (2006) Effect of acute increases in pulmonary vascular pressures on exercise pulmonary gas exchange. *J Appl Physiol* 100:1910–1917

7. Botero M, Kirby D, Lobato EB, Staples ED, Gravenstein N (2004) Measurement of cardiac output before and after cardiopulmonary bypass: Comparison among aortic transit-time ultrasound, thermodilution, and noninvasive partial CO₂ rebreathing. *J Cardiothorac Vasc Anesth* 18:563–572
8. Keren H, Burkhoff D, Squara P (2007) Evaluation of a noninvasive continuous cardiac output monitoring system based on thoracic bioelectance. *Am J Physiol Heart Circ Physiol* 293:H583–H589
9. Stetz CW, Miller RG, Kelly GE, Raffin TA (1982) Reliability of the thermodilution method in the determination of cardiac output in clinical practice. *Am Rev Respir Dis* 126:1001–1004
10. Hillis LD, Firth BG, Winniford MD (1985) Analysis of factors affecting the variability of Fick versus indicator dilution measurements of cardiac output. *Am J Cardiol* 56:764–768
11. Rubini A, Del Monte D, Catena V, Attar I, Cesaro M, Soranzo D, Rattazzi G, Alati GL (1995) Cardiac output measurement by the thermodilution method: an in vitro test of accuracy of three commercially available automatic cardiac output computers. *Intensive Care Med* 21:154–158
12. De Backer D, Moraine JJ, Berre J, Kahn RJ, Vincent JL (1994) Effects of dobutamine on oxygen consumption in septic patients. Direct versus indirect determinations. *Am J Respir Crit Care Med* 150:95–100
13. Lu Z, Mukkamala R (2006) Continuous cardiac output monitoring in humans by invasive and noninvasive peripheral blood pressure waveform analysis. *J Appl Physiol* 101:598–608
14. Cecconi M, Dawson D, Grounds R, Rhodes A (2009) Lithium dilution cardiac output measurement in the critically ill patient: determination of precision of the technique. *Intensive Care Med* 35:498–504
15. Haller M, Zollner C, Briegel J, Forst H (1995) Evaluation of a new continuous thermodilution cardiac output monitor in critically ill patients: a prospective criterion standard study. *Crit Care Med* 23:860–866
16. Le Tulzo Y, Belghith M, Seguin P, Dall'Ava J, Monchi M, Thomas R, Dhainaut JF (1996) Reproducibility of thermodilution cardiac output determination in critically ill patients: comparison between bolus and continuous method. *J Clin Monit* 12:379–385
17. Zollner C, Goetz AE, Weis M, Morstedt K, Pichler B, Lamm P, Kilger E, Haller M (2001) Continuous cardiac output measurements do not agree with conventional bolus thermodilution cardiac output determination. *Can J Anaesth* 48:1143–1147
18. Bendjelid K, Schutz N, Suter PM, Romand JA (2006) Continuous cardiac output monitoring after cardiopulmonary bypass: a comparison with bolus thermodilution measurement. *Intensive Care Med* 32:919–922
19. Maxwell RA, Gibson JB, Slade JB, Fabian TC, Proctor KG (2001) Noninvasive cardiac output by partial CO₂ rebreathing after severe chest trauma. *J Trauma* 51:849–853
20. Della Rocca G, Costa MG, Coccia C, Pompei L, Di Marco P, Vilardi V, Pietropaoli P (2003) Cardiac output monitoring: aortic transpulmonary thermodilution and pulse contour analysis agree with standard thermodilution methods in patients undergoing lung transplantation. *Can J Anaesth* 50:707–711
21. Kotake Y, Moriyama K, Innami Y, Shimizu H, Ueda T, Morisaki H, Takeda J (2003) Performance of noninvasive partial CO₂ rebreathing cardiac output and continuous thermodilution cardiac output in patients undergoing aortic reconstruction surgery. *Anesthesiology* 99:283–288
22. Boldt J, Menges T, Wollbruck M, Hammermann H, Hempelmann G (1994) Is continuous cardiac output measurement using thermodilution reliable in the critically ill patient? *Crit Care Med* 22:1913–1918
23. Burchell SA, Yu M, Takiguchi SA, Ohta RM, Myers SA (1997) Evaluation of a continuous cardiac output and mixed venous oxygen saturation catheter in critically ill surgical patients. *Crit Care Med* 25:388–391
24. Luchette FA, Porembka D, Davis K Jr, Branson RD, James L, Hurst JM, Johannigman JA, Campbell RS (2000) Effects of body temperature on accuracy of continuous cardiac output measurements. *J Invest Surg* 13:147–152
25. Mihm FG, Gettinger A, Hanson CW 3rd, Gilbert HC, Stover EP, Vender JS, Beerle B, Haddow G (1998) A multicenter evaluation of a new continuous cardiac output pulmonary artery catheter system. *Crit Care Med* 26:1346–1350
26. Medin DL, Brown DT, Wesley R, Cunnion RE, Ognibene FP (1998) Validation of continuous thermodilution cardiac output in critically ill patients with analysis of systematic errors. *J Crit Care* 13:184–189
27. Sun Q, Rogiers P, Pauwels D, Vincent JL (2002) Comparison of continuous thermodilution and bolus cardiac output measurements in septic shock. *Intensive Care Med* 28:1276–1280
28. Su NY, Huang CJ, Tsai P, Hsu YW, Hung YC, Cheng CR (2002) Cardiac output measurement during cardiac surgery: esophageal Doppler versus pulmonary artery catheter. *Acta Anaesthesiol Sin* 40:127–133
29. Dark PM, Singer M (2004) The validity of trans-esophageal Doppler ultrasonography as a measure of cardiac output in critically ill adults. *Intensive Care Med* 30:2060–2066
30. Chew M, Poelaert J (2003) Accuracy and repeatability of pediatric cardiac output measurement using Doppler: 20-year review of the literature. *Intensive Care Med* 1889–1894
31. de Waal EE, Kalkman CJ, Rex S, Buhre WF (2007) Validation of a new arterial pulse contour-based cardiac output device. *Crit Care Med* 35:1904–1909
32. Friesecke S, Heinrich A, Abel P, Felix S (2009) Comparison of pulmonary artery and aortic transpulmonary thermodilution for monitoring of cardiac output in patients with severe heart failure: validation of a novel method. *Crit Care Med* 37:119–123
33. Singh A, Juneja R, Mehta Y, Trehan N (2002) Comparison of continuous, static, and intermittent cardiac output measurements in patients undergoing minimally invasive direct coronary artery bypass surgery. *J Cardiothorac Vasc Anesth* 16:186–190
34. John-Alder H, Bennet A (1981) Thermal dependence of endurance and locomotory energetics in a lizard. *Am J Physiol* 241:R342–R349
35. Critchley LA, Critchley JA (1999) A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. *J Clin Monit Comput* 15:85–91
36. Squara P (2004) Matching total body oxygen consumption and delivery: a crucial objective? *Intensive Care Med* 30:2170–2179
37. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Tomlanovich M (2001) Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 345:1368–1377

38. Squara P, Fourquet E, Jacquet L, Broccard A, Uhlig T, Rhodes A, Bakker J, Perret C (2003) A Computer Program for Interpreting Pulmonary Artery Catheterization Data: Results of the European HEMODYN Resident Study. *Intensive Care Med* 29:735–741
39. Poli de Figueiredo LF, Malbouisson LM, Varicoda EY, Carmona MJ, Auler JO Jr, Rocha e Silva M (1999) Thermal filament continuous thermodilution cardiac output delayed response limits its value during acute hemodynamic instability. *J Trauma* 47:288–293
40. Siegel LC, Hennessy MM, Pearl RG (1996) Delayed time response of the continuous cardiac output pulmonary artery catheter. *Anesth Analg* 83:1173–1177
41. Monnet X, Teboul J (2008) Passive leg raising. *Intensive Care Med* 34:659–663
42. Cecconi M, Rhodes A, Poloniecki J, Della Rocca G, Grounds R (2009) 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* 13:201