ORIGINAL LABORATORY RESEARCH REPORT

Comparison of Hemoglobin Measurements by 3 Point-of-Care Devices With Standard Laboratory Values and Reliability Regarding Decisions for Blood Transfusion

Marianne Johnson, FCA (SA), Peter C. Marwick, FCA (SA), and Johan F. Coetzee, PhD

BACKGROUND: We compared the accuracy of 3 point-of-care testing (POCT) devices with central laboratory measurements and the extent to which between-method disagreements could influence decisions to transfuse blood.

METHODS: Hemoglobin concentrations [Hb] were measured in 58 adult patients undergoing cardiothoracic surgery using 2 Ilex GEM Premier 3500 blood gas analyzers (BG_A and BG_B) and a HemoCue Hb-201$^+$ device (HemoCue). Measurements were compared with our central laboratory's Siemens Advia 2120 flow cytometry system (laboratory [Hb] [Lab[Hb]]), regarded as the gold standard. We considered that between-method [Hb] differences exceeding 10% in the [Hb] range 6–10 g/dL would likely erroneously influence erythrocyte transfusion decisions.

RESULTS: The 70 Lab[Hb] measurements ranged from 5.8 to 16.7 g/dL, of which 25 (36%) were <10.0 g/dL. Measurements by all 4 devices numbered 57. Mean POCT measurements did not differ significantly ($P > .99$). Results of the Bland–Altman analyses revealed statistically significant bias, with predominant underestimations by all 3 POCTs predominating. HemoCue upper and lower limits of agreement (LOA) were narrower, and the 95% confidence intervals (95% CIs) of the LOAs did not overlap with those of BG_A and BG_B. Similarly, a narrow mountain plot demonstrated greater precision for the HemoCue. Comparing BG_A with BG_B revealed no bias and narrow LOA. Error grid analysis within the [Hb] range 6–10 g/dL revealed that 5.3% of HemoCue measurements were beyond the permissible 10.0% error zone in contrast to 19.0% and 16.0% of the blood gas measurements. Possible inappropriate transfusion decisions based on POCT values generally erred toward unnecessary transfusions. Calculations of Cohen κ statistic indicated better chance-corrected agreement between HemoCue and Lab[Hb] regarding erythrocyte transfusions than the blood gas analyzers.

CONCLUSIONS: All 3 POCT devices underestimated the Lab[Hb] and cannot be used interchangeably with standard laboratory measurements. BG_A and BG_B can be considered to be acceptably interchangeable with each other. Whereas the HemoCue had little bias and good precision, the blood gas analyzers revealed large bias and poor precision. We conclude that the tested HemoCue provides more reliable measurements, especially within the critical 6–10 g/dL range, with reduced potential for transfusion errors. Decisions regarding erythrocyte transfusions should also be considered in the light of clinical findings. (Anesth Analg 2020;131:640–9)

KEY POINTS

- **Question:** How does hemoglobin measurement by HemoCue and blood gas analyzer compare with central laboratory measurements?
- **Finding:** The HemoCue performed with superior clinical reliability but underestimated the central laboratory values.
- **Meaning:** Decision to transfuse must not be based on hemoglobin measurements alone but must be considered together with clinical findings.

GLOSSARY

ASA = American Society of Anesthesiologists; **BG_A and BG_B** = blood gas analyzers; **CI** = confidence interval; **EDTA** = ethylenediaminetetraacetic acid; **Hb** = hemoglobin; **[Hb]** = hemoglobin concentration; **Hct** = hematocrit; **HiCN** = hemiglobincyanide; **iQM** = intelligent quality management; **Lab[Hb]** = laboratory hemoglobin concentration; **LOA** = limits of agreement; **NHLS** = National Health Laboratory Service; **POCT** = point-of-care testing devices; **RM-ANOVA** = repeated-measures analysis of variance; **STARD** = Standards for Reporting Diagnostic accuracy studies

From the Department of Anesthesiology and Critical Care, Faculty of Medicine and Health Sciences, Stellenbosch University and Tygerberg Academic Hospital, Cape Town, South Africa.

Accepted for publication October 7, 2019.

Funding: None.

DOI: 10.1213/ANE.0000000000004533

The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.anesthesia-analgesia.org).

Reprints will not be available from the authors.

Address correspondence to Marianne Johnson, FCA (SA), Department of Anesthesiology and Critical Care, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town 8000, South Africa. Address e-mail to mariannejohnson@sun.ac.za.

Point-of-care diagnostic tests (POCTs) performed at or near patient care sites generate rapid results, enabling prompt, cost-effective treatment.[1–4] In our operating rooms, POCT hemoglobinometers are used to guide decisions regarding erythrocyte transfusions while also considering the need to effectively manage a scarce, expensive resource, as well as the risks that accompany transfusions.[2,3] Depending on clinical circumstances, decisions to transfuse erythrocytes are usually made when hemoglobin concentrations ([Hb]) have decreased to values 6–10 g/dL.[5–7] Pecoraro et al[8] have emphasized the need for ongoing assessment of the accuracy of POCT. The main purpose of our investigation was to compare the performance of 3 POCT hemoglobinometers with standard laboratory measurements.

The primary outcomes were the measurement differences between the POCT hemoglobinometers and the reference values. A clinically relevant difference was considered to be a >10% difference between the reference method and POCT. The secondary outcome was whether, in the critical decision-making [Hb] range of 6–10 g/dL, reliance on the different hemoglobinometers would result in contrasting decisions to transfuse compared with the reference method.

METHODS

Ethics approval for this prospective, measurement methods study was obtained from the Health Research Ethics Committee of Stellenbosch University (protocol No. S13/10/206). The study was conducted in compliance with the submitted protocol, the International Council for Harmonization, Good Clinical Practice guidelines, and the applicable Standards for Reporting Diagnostic accuracy studies (STARD) requirement(s). Formal written informed consent was waived on the grounds that the research design involved no more than minimal risk and the tests were part of routine care for patients undergoing major surgery. All patients included in the study required arterial catheterization for clinical reasons.

The POCT devices were 2 Ilex GEM Premier 3500 blood gas analyzers (BG_A and BG_B; Instrumentation Laboratory, Lexington, MA) and a HemoCue Hb 201$^+$ System (Mallinckrodt Medical, Hennef, Germany). Our reference hemoglobin (Hb) measurements were performed by the on-site National Health Laboratory Service (NHLS), using a Siemens Advia 2120i system (Siemens Healthcare Diagnostics, Eschborn, Germany), which uses the standard hemiglobincyanide (HiCN) test method. The HemoCue device uses a dry chemistry principle, shown to be accurate and reliable within 1 g/dL of reference values over a wide range of [Hb].[9–18] The blood gas analyzers measure the hematocrit (Hct) using conductivity technology, from which the [Hb] is calculated: $([Hb] = 0.31 \times Hct)$. Several studies suggest that, whereas measurements by most blood gas analyzers are within 1 g/dL of reference values, the 1-g/dL permissible difference is often exceeded.[4,17,19–23] This is more commonly an overestimation but also occasionally an underestimation. Meticulous technique in acquiring blood samples is important. Seguin et al[16] recommend that capillary blood from finger prick samples should not be analyzed using a HemoCue in critically ill patients, because tissue edema greatly confounds results.

To prepare the sampling syringes, 0.5 mL of 5000 µ/mL heparin (Heparin Sodium; Fresenius Kabi, Canada) was drawn up into a 2-mL syringe to coat the walls of the syringe and the excess heparin expressed. Although sampling into dry heparinized cuvettes results in fewer measurement errors,[23] we used wet heparinized syringes according to our clinical practice, because the sampling cuvettes are not readily available in our hospital.

Arterial blood samples (3 mL) were drawn from 58 patients ≥18 years of age undergoing cardiothoracic surgery between February and May 2014 at the Tygerberg Hospital, Cape Town, South Africa. A maximum of 2 blood samples per patient were taken. Blood was drawn preoperatively, intraoperatively, or directly postoperatively. Three times the volume (9 mL) of the sample line was initially withdrawn from the arterial line to prevent hemodilution or contamination of the blood sample. This blood was injected back into the patient via a peripheral or central line using a surgically clean technique, after the sample for analysis had been expressed. Each sample was split as follows: 1 mL for the 2 blood gas analyzers, 1 drop for the HemoCue microcuvette, and 2 mL placed in an ethylenediaminetetraacetic acid (EDTA) Vacutainer tube (BD Diagnostics, Franklin Lakes, NJ) to be sent to the central laboratory. The disposable HemoCue microcuvettes were stored at room temperature (20°C–23°C), and expiry dates were checked before use. Measurements were performed within 3 minutes of breaking the seal as specified by the manufacturer. Blood from the wet heparinized syringes was used for the 2 Ilex GEM Premier 3500 blood gas analyzers, and these samples were analyzed immediately.

The EDTA blood samples were transported at room temperature (±25°C) to the on-site NHLS, where they were analyzed using the Siemens Advia 2120i system. These measurements were considered our reference values (laboratory [Hb] [Lab[Hb]]).[24] The mean time from collection to analysis was 6.2 hours (59% within 6 hours). No specimens were analyzed after 24 hours because it has been shown that stability of hematologic analyses using the Siemens Advia 2012i is not affected by EDTA room temperature storage for up to 24 hours.[25]

To ensure consistent results, only the attending anesthesiologists and qualified clinical technologists obtained the blood specimens and operated the POCT devices. The central laboratory and clinical technologists were blinded with respect to each other's results. POCT results were available to the attending anesthesiologist for clinical decision-making.

The HemoCue performs an automatic quality control, whereby on being switched on, the performance of the optical components is verified and a warning is given if the self-test fails. The test is repeated every 2 hours if the analyzer remains switched on. The intelligent quality management (iQM) of the Ilex GEM Premier 3500 runs after every blood gas analysis, as well as every 12 hours. Via internal process control solutions and calibration validation solutions, iQM provides continuous quality management and assessment of functionality with real-time error detection and correction and assures accurate results.[26] It is assumed that all of the iQM cartridges used are similar. The Siemens equipment was calibrated and tested as per the relevant laboratory standard operating procedures with control and calibrator products supplied by Siemens. The reference micro-Hct was measured using the National Committee for Clinical Laboratory Standards H-7A–approved standard method with tripotassium EDTA as the anticoagulant.[27]

Statistical Analysis

Data were analyzed using MedCalc Statistical Software (version 17.9.6; MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org; 2017). The [Hb] data were tested for normal distributions using the Shapiro–Wilk test. Preliminary comparisons among the 4 instruments were done using repeated-measures analysis of variance (RM-ANOVA), followed by pairwise comparisons with Bonferroni correction.

Bias and precision were assessed by comparing the measurements of the 3 POCT devices with Lab[Hb], using the techniques of Bland and Altman.[9,28] In the context of this study, bias is defined as the degree and direction by which an instrument's measurements differ from the reference method. Precision refers to the closeness of the instrument's measurements to each other.

Lab[Hb] was regarded as a gold standard. In this context, bias refers to the systematic difference from the gold standard and precision refers to the variability of the measurements. We chose to compare differences between POCT measurements with Lab[Hb] rather than with the mean values of the 2 measurements (mean of POCT and Lab[Hb]).[29] Limits of agreement (LOA) expressed as percentage differences were compared graphically. To complement the Bland–Altman analysis, folded empirical cumulative distribution plots, also known as mountain plots, were constructed.[30,31] This procedure plots the ranks (in percentiles) of the percentage differences (between the POCT and the laboratory results) on the ordinate against the percentage differences on the abscissa. The ranks are then folded around the 50th percentile rank.

Error grid analyses for paired Hb values were performed using the graphical technique described by Clarke et al[32] and adapted by Morey et al[6] for Hb. The reference method is plotted on the abscissa versus the POCT measures on the ordinate. Zones are defined that demarcate acceptable and unacceptable errors. Of particular interest is the [Hb] range 6–10 g/dL, which involves critical decisions concerning blood transfusions and within which range only a 10% error is generally regarded as permissible.[6]

We investigated a hypothetical scenario wherein, after taking clinical factors into account, a decision to transfuse erythrocytes would depend on POCT [Hb] measurements for transfusion thresholds at <10, <9, <8, <7, and <6 g/dL. We calculated the Cohen κ statistic for agreements to transfuse beyond chance, between each POCT device and Lab[Hb], as recommended by Morey et al.[6]

Proportions were compared using the Freeman–Halton[33] extension of the 2-tailed Fisher exact test for a 2 × 3 contingency table. Proportions compared included the following: POCT readings that were within 10% of Lab[Hb]; those that lay within the critical isthmus of the error zone diagram (zone B); and proportions of inappropriate hypothetical erythrocyte transfusions for each POCT.

Sample Size Calculation. A sample size calculation to detect a 10% difference between a POCT device and Lab[Hb] at the 10 g/dL level was done using statistical software (NCSS PASS [NCSS, LLC, Kaysville, UT], release April 2007, NCSS Statistical Software, Kaysville, UT; www.NCSS.com). A sample size of 66 achieves 90% power to detect a difference of 1.0 g/dL between the null hypothesis mean of 10.0 g/dL and the alternative hypothesis mean of 9.0 g/dL. We assumed an SD of 2.5 g/dL and a significance level (α) of .05 using a 2-sided 1-sample *t* test. We chose a difference between 10 and 9 g/dL because we planned to perform error grid analysis in accordance with the recommendations of Morey et al.[6] Their reasoning is that a [Hb] of 10 g/dL is generally regarded as the upper limit for considering erythrocyte transfusion, and a ±10% error would represent a deviation of 1 g/dL.[34] Furthermore, 1 unit of packed cells constitutes the smallest transfusion-associated risk and would generally raise an average patient's [Hb] by 1 g/dL.[6] It was decided to study 70 blood samples to allow for possible missing values.

RESULTS

A total of 35 men and 23 women participated, with a mean age of 50.8 years (range, 24–82 years). Four patients had previously received packed red blood cells (1 patient 1 unit, 2 patients 3 units, and 1 patient 4 units). None of the patients presented with active bleeding. (See Supplemental Digital Content, Table S1, http://links.lww.com/AA/C963, for surgical procedures performed.)

Seventy Lab[Hb] (reference) values were obtained that ranged from 5.8 to 16.7 g/dL (mean, 11.02 g/dL; SD, 2.3; 95% confidence interval [95% CI], 10.53–11.63). Measurements from all 4 instruments numbered 57. Data were normally distributed. No significant differences were detected between the 3 POCT devices' mean measurements. All 3 POCT measurements revealed negative bias, mean values being approximately 6% less than mean Lab[Hb] (RM-ANOVA $P < .001$; Table 1). The proportions of the POCT measurements that were within 10% of Lab[Hb] are presented in Table 2.

Results of the Bland–Altman analyses confirmed statistically significant bias, with underestimations of Lab[Hb] by all 3 POCT devices predominating (Figure 1; Table 1). The observed mean differences were similar for all 3; however, HemoCue differences were less dispersed than those of the 2 blood gas devices as revealed by comparison of their LOA. Figure 2 compares the 95% CI of the 3 LOAs expressed as percentage differences. The 95% CI of the HemoCue LOA did not overlap with those of the 2 blood gas devices. Mountain plots (Figure 3) demonstrated that the median percent differences (bias) were similar. However, the much narrower HemoCue plot indicated greater measurement precision. In addition, there were fewer disparities from Lab[Hb] beyond the clinically acceptable 10% limit (shaded area in Figure 3).

Error grid analysis (Figure 4) revealed that, altogether 3 (5.3%) of 57, 13 (18.6%) of 70, and 11 (15.9%) of 69 values were located in the yellow zone B for the HemoCue, BG_A, and BG_B POCT devices, respectively ($P = .60$). No values appeared in the red zone C. (See legend to Figure 4 for an explanation and interpretation of zones A, B and C.) Within the narrow, critical isthmuses where Lab[Hb] <10, the proportions of POCT measurements that lay in the yellow zone B were 3 (14.3%) of 21, 9 (39.1%) of 23, and 11 (47.8%) of 23 for the HemoCue, BG_A, and BG_B POCT devices, respectively ($P = .049$; however, no pairwise differences were detected after multiple comparisons).

Hypothetically, possible inappropriate blood transfusion decisions could have occurred with all 3 POCT devices, should their measurements have been used for decision-making guidance. These were BG_A 10 (23.3%) of 43, BG_B 11 (26.8%) of 41, and HemoCue 9 (26.5%) of 34 ($P = .93$). Because all 3 POCT devices tended to underestimate true [Hb] values, possible inappropriate transfusion decisions based on POCT values generally erred toward unnecessary transfusions. However, in

Table 1. Results of the Bland–Altman Analysis: Comparison of the 3 Point-of-Care Devices With Standard Laboratory Measurements

	BG_A Versus Lab[Hb]	BG_B Versus Lab[Hb]	HemoCue Versus Lab[Hb]	BG_A Versus BG_B[a]
Sample size	70	69	57	71
Mean difference (g/dL)	0.69	0.66	0.67	−0.05
SD	1.00	0.99	0.54	0.45
95% CI of mean difference	0.44–0.93	0.42–0.90	0.52–0.81	−0.16 to 0.05
P value	<.0001	<.0001	<.0001	.31
Lower LOA	−1.29	−1.27	−0.40	−0.95
95% CI	−1.70 to −0.87	−1.68 to −0.87	−0.65 to −0.15	−1.13 to −0.76
Upper LOA	2.66	2.59	1.74	0.84
95% CI	2.25–3.07	2.19–3.00	1.49–1.99	0.65–1.02
Slope	−0.098	−0.013	−0.002	−0.01
P value	.055	.040	.95	.62
95% CI	−0.199 to 0.002	−0.201 to −0.005	−0.064 to 0.068	−0.05 to 0.03
Results of the Bland–Altman Analysis Expressed as Percentages				
Mean difference (%)	6.4%	6.2%	6.1%	−0.5%
SD	10.2%	10.0%	4.8%	4.5%
95% CI of mean difference	4.0%–8.8%	3.8%–8.6%	4.8%–7.4%	−1.6% to 0.6%
P value	<.0001	<.0001	<.0001	.31
Lower LOA	−13.6%	−13.4%	−3.3%	−9.5%
95% CI	−17.7% to −9.4%	−17.5% to −9.3%	−5.6% to −1.1%	−11.4% to −7.6%
Upper LOA	26.4%	25.8%	15.6%	8.4%
95% CI	22.2%–30.6%	21.6%–29.9%	13.4%–17.8%	6.6%–10.3%

Lab[Hb] = Siemens Advia 2120, ie, reference value. BG_A and BG_B = the 2 blood gas analyzers, ie, Ilex GEM Premier 3500 blood gas analyzers. HemoCue = HemoCue Hb 201+ System. Difference = mean difference (g/dL). Slope = proportional bias, that is, slope of regression line, difference versus reference value.

Abbreviations: CI, confidence interval; Lab[Hb], laboratory hemoglobin concentration; LOA, limits of agreement, ie, mean difference ± 1.96 SD.

[a]Bland–Altman analysis: difference versus mean of BG_A and BG_B; B-A graph available for perusal in Supplemental Digital Content, http://links.lww.com/AA/C963.

3 blood samples, blood gas [Hb] measurements were >10 g/dL, but the corresponding Lab[Hb] values were <10 g/dL, thus implicating wrong decisions "not" to transfuse at that particular threshold. Calculations of Cohen κ statistics regarding chance-corrected agreements concerning hypothetical transfusion decisions are presented in Supplemental Digital Content, Table S2, http://links.lww.com/AA/C963. Lab[Hb] <8 g/dL was sparse, resulting in very wide confidence intervals for κ; therefore, calculations are presented only for thresholds <10 and <9.

HemoCue measurements incurred wrong decisions to transfuse solely due to underestimations. These occurred within a HemoCue range 7.9–9.7 g/dL corresponding to a Lab[Hb] range 8.5–11.3 g/dL for both transfusion thresholds.

Table 2. Proportions of POCT Measurements That Were Within ±10% of Standard Laboratory Values

Device	±10%[a]	Lab[Hb] > POCT	Lab[Hb] < POCT
BG_A	0.67	0.51	0.16
95% CI	0.56–0.77	0.40–0.63	0.09–0.26
BG_B	0.67	0.48	0.19
95% CI	0.55–0.77	0.37–0.59	0.11–0.30
HemoCue	0.81	0.68	0.12
95% CI	0.69–0.89	0.56–0.79	0.06–0.23

BG_A and BG_B = blood gas machine values, and Lab[Hb] = standard laboratory hemoglobin values. POCT > Lab[Hb] = proportions of POCT measurements that were <10% greater than Lab[Hb]. POCT < Lab[Hb] = proportions of POCT measurements that were <10% less than Lab[Hb].

Abbreviations: CI, confidence interval; Lab[Hb], laboratory hemoglobin concentration; POCT, point-of-care testing.

[a]*P* = .15 (Fisher exact test).

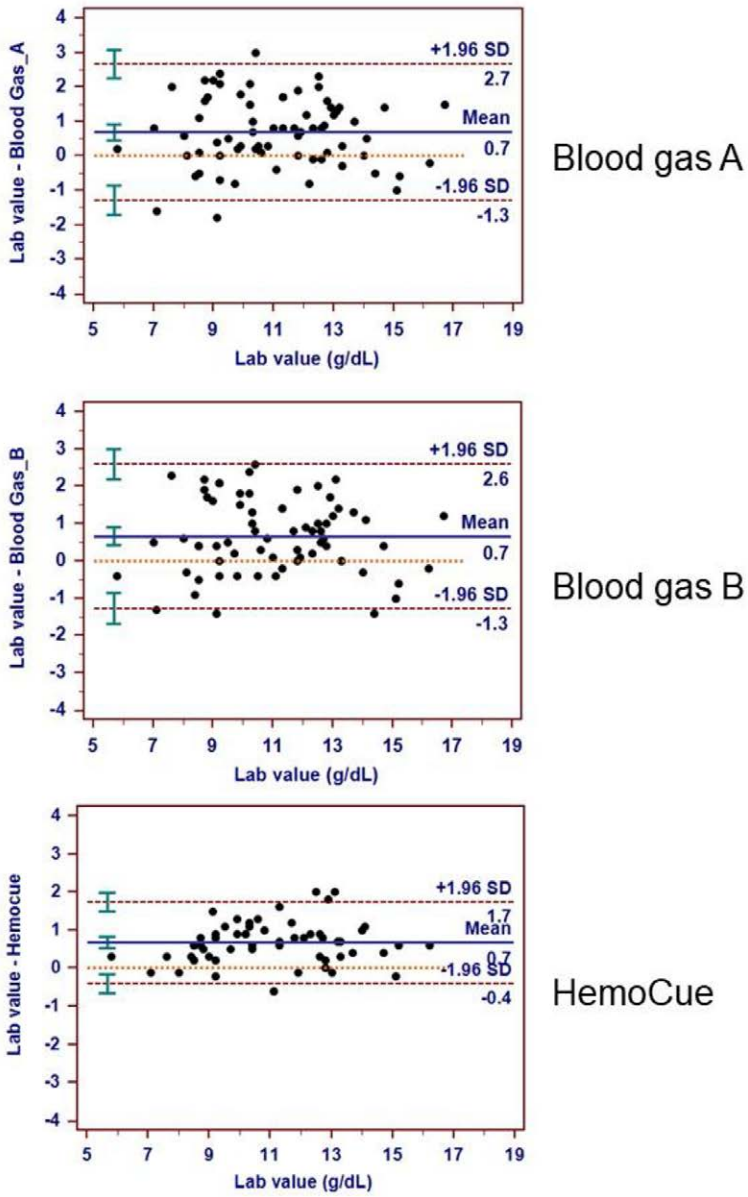

Figure 1. Bland–Altman graphs comparing measurements from the 3 point-of-care devices with standard laboratory values. Lab value: Hemoglobin concentrations measured by the central laboratory. Lab value - blood gas: difference between laboratory value and the point-of-care device. Orange-dotted line: line of perfect agreement (zero difference). Solid blue line: mean value of the differences. The error bar indicates the 95% confidence interval of the mean. Outer-dotted lines: lower and upper limits of agreement (1.96 SDs). The error bars indicate the 95% confidence intervals of the limits of agreement.

DISCUSSION

Overall, mean POCT measurements were statistically significantly smaller than reference measurements; however, the mean differences were minor (Table 1). None of the 3 POCT devices demonstrated noteworthy proportional bias (Table 1). Although the slope of the regression between the [Hb] differences and Lab[Hb] for BG_A was statistically significant (−0.01 dL/dL; $P = .040$), it was clinically unimportant. Mean differences between Lab[Hb] and the 3 POCT devices were similar; however, the HemoCue differences were less widely spread, as revealed by Figures 1–3. The mountain plot (Figure 3) indicates that 88% of the HemoCue differences were within 10% of Lab[Hb]. Thus, whereas the HemoCue had little bias and good precision, the blood gas analyzers revealed larger bias and poor precision. Interestingly, other studies report an overestimation of the reference [Hb] by blood gas analyzers, whereas our blood gas analyzers more frequently underestimated the reference value.[4,21,35] We currently do not have a scientific explanation for this phenomenon.

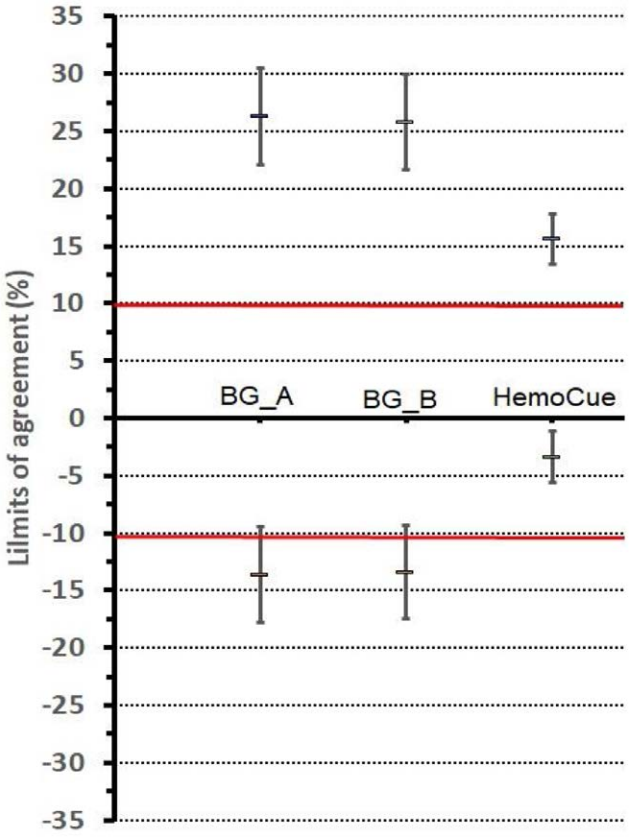

Figure 2. Limits of agreement for the 3 point-of-care (POCT) devices expressed as percent differences from central laboratory values (laboratory value − POCT)/laboratory value. Error bars indicate 95% confidence intervals of the limits of agreement where limits of agreement are defined as mean difference from the reference measurement ± 1.96 SDs. Red horizontal lines indicate the predefined acceptable 10% error. For a device to be considered interchangeable with the gold standard, the 95% confidence intervals should liewithin the area bounded by the 2 red lines. BG_A indicates blood gas analyzer A; BG_B = blood gas analyzer B; POCT, point-of-care testing.

The 95% CIs of the LOAs in the Bland–Altman analysis reveal that none of the POCT devices can be considered to be interchangeable with the reference method. For 2 methods of measurement to be considered interchangeable, the lower bound of the 95% CI of the lower LOA should be greater than the predefined acceptable negative error (lower red line in Figure 2), and the upper bound of the 95% CI of the upper LOA should be less than the acceptable positive error (upper red line in Figure 2). The 2 blood gas analyzers proved to be acceptably interchangeable (Table 1; Figures 2–3), although the extreme upper and lower bounds of the 95% CI of the LOAs slightly exceeded 1 g/dL and 10% (Table 1). (A Bland–Altman graph of the 2 blood gas analyzers can be viewed in Supplemental Digital Content, Figure S1, http://links.lww.com/AA/C963.)

For our Bland–Altman analysis, we chose to plot measurement differences (Y-X) against X, where X is the reference (gold standard) method as recommended by Krouwer[29] instead of against the averages of the 2 methods, that is, ([X + Y]/2). He showed by means of repetitive simulations that the imprecision of the reference method (X) results in different estimates of the correlation coefficients between (Y − X) versus X and between (Y − X) versus (X + Y)/2. When the X measurements have small imprecision, the correlation coefficients are much smaller when plotting (Y − X) against X than when plotting (Y − X) against (X + Y)/2. The converse happens when the reference measure (X) exhibits significant imprecision, thus confirming the recommendation by Bland and Altman to use (X + Y)/2 in the latter instance. Nevertheless, we also conducted a sensitivity analysis by performing Bland–Altman analyses using (Y − X) versus (X + Y)/2. These graphs appeared to be virtually identical, and they did not make any difference to our conclusions (available for perusal in Supplemental Digital Content, Figures S2–S3, http://links.lww.com/AA/C963).

The American Society of Anesthesiologists (ASA) Practice Guidelines for Perioperative Blood Transfusion (ASA-Guidelines)[34] recognize that the information defining exactly when a perioperative blood transfusion is necessary is not obtainable from the literature, because clinical considerations also influence decisions to transfuse erythrocytes. The ASA-Guidelines recommend that erythrocyte transfusions are mostly unneeded when [Hb] >10 g/dL but should be administered when [Hb] <6 g/dL, for example, in a young healthy patient, especially when bleeding is acute. When [Hb] is 6–10 g/dL, the ASA-Guidelines advise that clinical factors should be

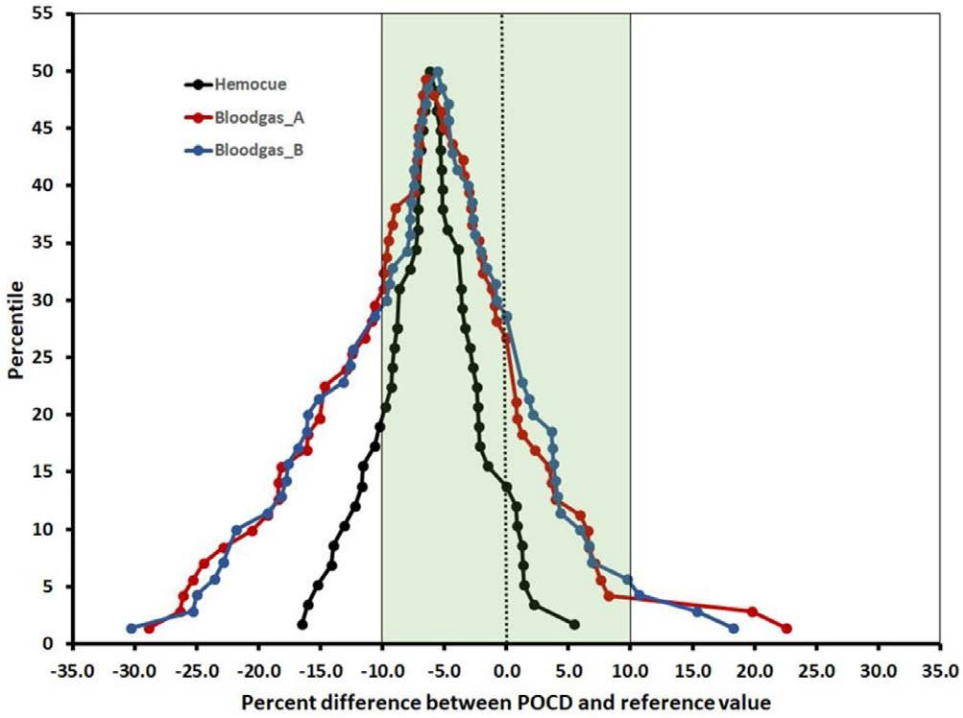

Figure 3. Mountain plot comparing hemoglobin concentration percent differences for the 3 point-of-care hemoglobinometers with central laboratory measurements. The narrower HemoCue plot indicates greater precision. Furthermore, there are fewer values beyond acceptable 10% errors (shaded area). POCD indicates point-of-care device.

considered, for example, actual or anticipated continued bleeding, volume status, signs of inadequate perfusion of vital organs, poor cardiopulmonary reserve, and high oxygen consumption. These clinical issues are the reason for requiring reliably accurate measurements in the region 6–10 g/dL and thus the necessity for the narrow isthmus in the green Zone A in the error grid diagrams of Morey et al[6] (Figure 4).

Cohen κ statistic is a measure of reliability of the degree of dichotomous agreement between 2 raters, corrected for pure chance (also termed "chance-corrected agreement").[36] Other than error grid analysis, Morey et al[6] recommended calculating κ regarding decisions to transfuse erythrocytes over the [Hb] range 6–10 g/dL. Our results (Supplemental Digital Content, Table S2, http://links.lww.com/AA/C963) indicate different κ statistics for 2 thresholds. Regarding threshold <10, both blood gas analyzers revealed poor agreement, whereas any agreement regarding HemoCue was no different from pure chance. Regarding threshold <9, the blood gas analyzers showed poor agreement, whereas the HemoCue revealed moderate-to-good agreement.

A weakness of this study is the paucity of [Hb] in the range 6–8 g/dL, because only 4 Lab[Hb] were <8 g/dL. Thus, no meaningful conclusions could be drawn concerning thresholds <8 g/dL. As with previously published studies concerning POCT hemoglobinometers,[37,38] most of our measurements were >10 g/dL (64% of Lab[Hb]), which Morey et al[7] have pointed out are of little interest in the context of perioperative erythrocyte transfusions and which weigh Bland–Altman analysis in favor of higher concentrations. However, considering the results of the error grid analysis within the narrow critical isthmus where [Hb] <10 g/dL, in conjunction with a favorable κ statistic, there is probably sufficient evidence to suggest that the HemoCue is the more reliable instrument, provided that arterial samples and not capillary samples are analyzed. Nevertheless, it will require additional studies of sufficient magnitude within the range 6–10 g/dL to confirm that possibility.

An additional potential weakness is that there were 57 HemoCue-Lab[Hb] pairs instead of the intended 70. During the study, only 1 HemoCue device was available for the entire operating room suite, and on occasion the HemoCue device was in use elsewhere and could not be traced before the arterial samples had been dispatched to the central laboratory. We performed a retrospective sample size determination with regard to the Bland–Altman analysis for the HemoCue device measurements to be considered interchangeable with those of Lab[Hb]. (MedCalc statistical software version 18.11.6 uses the method described by Lu et al[39] to calculate the sample sizes). The obtained mean difference between the HemoCue and the Lab[Hb] measurements was 0.67 g/dL, with an SD of 0.54 g/dL. To demonstrate

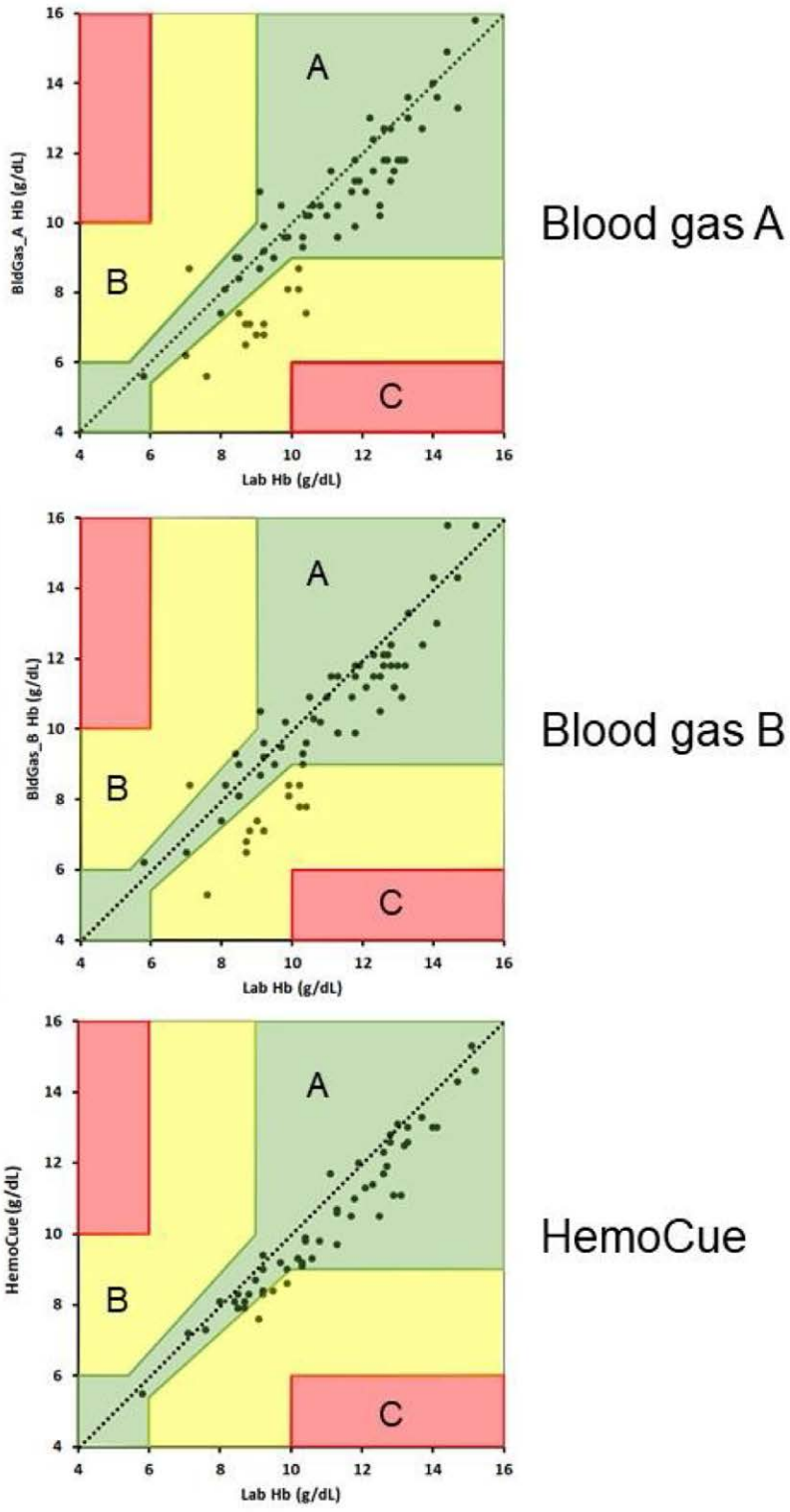

Figure 4. Hemoglobin error grid analysis. Central laboratory measurements are plotted along the abscissa and point-of-care hemoglobinometer results on the ordinate. The dotted line represents the line of identity. Zone A (green): the uppermost and lowermost areas of this zone represent areas in which errors are of little clinical importance because they do not influence decisions regarding blood transfusions. The band between hemoglobin concentrations 6–10 g/dL form a critical range, because errors >10% can influence decisions to transfuse in either direction. This range is based on the practice guidelines for transfusion published by the ASA[36] whereby red cell transfusions are advised for hemoglobin concentrations <6 g/dL and deemed unnecessary when >10 g/dL. Zone B (yellow): errors in this zone can result in wrong decisions regarding red cell transfusions. A reliable hemoglobinometer should result in <5% of errors situated in zone B. Zone C (red): errors in this zone represent potential major clinical mistakes. Errors in the upper red zone overestimate true hemoglobin concentrations resulting in possible failures to transfuse severely anemic patients. On the other hand, errors in the lower red zone grossly underestimate true values with possibilities of unnecessary transfusions and their attendant complications. ASA indicates American Society of Anesthesiologists.

equivalence, the maximum allowable mean difference (Δ) would need to be ≥1.73 g/dL (0.67 + [0.54 × 1.96]). Our predefined Δ was 1.0 g/dL; therefore, considering the results from our 57 blood sample pairs as a pilot study, it would not be feasible to attempt to demonstrate equivalence. Indeed, if one were to regard 1.73 g/dL as Δ, this would require >100,000 pairs to demonstrate equivalence, assuming a type II error of 0.2.

An additional consideration is that there were insufficient data to demonstrate statistically significant differences regarding the proportions of POCT measurements that were within ±10% of Lab[Hb] (P = .15). A retrospective sample size calculation (MedCalc

statistical software version 18.11) estimated that, for 1 degree of freedom and a 0.2 type II error, 167 pairs per POCT device would be required to demonstrate a significant difference using a 1-sided Fisher exact test. In addition, regarding HemoCue measurements, considering that we also demonstrated a significant mean difference from Lab[Hb] of 0.67 g/dL with an acceptably narrow 95% CI (0.52–0.81 g/dL), we decided not to gather more data.

Our results are in agreement with the study by Patel et al,[17] who compared 8 POCT hemoglobinometers with a standard HiCN method. They concluded that the HemoCue was the second most accurate device, with small underestimating bias, and that the GEM 3000 blood gas analyzer was the most inaccurate, with large underestimating bias.

We conclude that, compared with the tested blood gas analyzers, the HemoCue revealed superior performance, having small bias and good precision. It can be used with greater confidence in the clinical decision-making process regarding decisions to transfuse erythrocytes, while bearing in mind the small negative offset. Additional research involving larger datasets in the [Hb] range 6–10 g/dL is required to confirm superior clinical reliability. As recommended by the ASA-Guidelines,[34] any decision to transfuse must not be based on [Hb] measurements alone but must be considered together with clinical issues. ■

ACKNOWLEDGMENTS

The corresponding author thanks Professor Johan Coetzee, without whose ongoing support and input this article would never have seen the light.

DISCLOSURES

Name: Marianne Johnson, FCA (SA).
Contribution: This author helped gather the data and write the manuscript.
Name: Peter C. Marwick, FCA (SA).
Contribution: This author helped gather the data and write the manuscript.
Name: Johan F. Coetzee, PhD.
Contribution: This author helped perform the statistical analyses and write the manuscript.
This manuscript was handled by: Maxime Cannesson, MD, PhD.

REFERENCES

1. Price CP. Point of care testing. *BMJ*. 2001;322:1285–1288.
2. Clevenger B, Richards T. Pre-operative anaemia. *Anaesthesia* 2015;70(suppl 1):20–28, e6-8.
3. Mirski MA, Frank SM, Kor DJ, Vincent JL, Holmes DR Jr. Restrictive and liberal red cell transfusion strategies in adult patients: reconciling clinical data with best practice. *Crit Care*. 2015;19:202.
4. Ray JG, Post JR, Hamielec C. Use of a rapid arterial blood gas analyzer to estimate blood hemoglobin concentration among critically ill adults. *Crit Care*. 2002;6:72–75.
5. Practice guidelines for perioperative blood management: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Blood Management. *Anesthesiology*. 2015;122:241–275.
6. Morey TE, Gravenstein N, Rice MJ. Let's think clinically instead of mathematically about device accuracy. *Anesth Analg*. 2011;113:89–91.
7. Morey TE, Gravenstein N, Rice MJ. Assessing point-of-care hemoglobin measurement: be careful we don't bias with bias. *Anesth Analg*. 2011;113:1289–1291.
8. Pecoraro V, Germagnoli L, Banfi G. Point-of-care testing: where is the evidence? A systematic survey. *Clin Chem Lab Med*. 2014;52:313–324.
9. Bond MM, Richards-Kortum RR. Drop-to-drop variation in the cellular components of fingerprick blood: implications for point-of-care diagnostic development. *Am J Clin Pathol*. 2015;144:885–894.
10. Giraud B, Frasca D, Debaene B, Mimoz O. Comparison of haemoglobin measurement methods in the operating theatre. *Br J Anaesth*. 2013;111:946–954.
11. Hiscock R, Kumar D, Simmons SW. Systematic review and meta-analysis of method comparison studies of Masimo pulse co-oximeters (Radical-7™ or Pronto-7™) and HemoCue® absorption spectrometers (B-Hemoglobin or 201+) with laboratory haemoglobin estimation. *Anaesth Intensive Care*. 2015;43:341–350.
12. Kim MJ, Park Q, Kim MH, Shin JW, Kim HO. Comparison of the accuracy of noninvasive hemoglobin sensor (NBM-200) and portable hemoglobinometer (HemoCue) with an automated hematology analyzer (LH500) in blood donor screening. *Ann Lab Med*. 2013;33:261–267.
13. Miller RD, Ward TA, Shiboski SC, Cohen NH. A comparison of three methods of hemoglobin monitoring in patients undergoing spine surgery. *Anesth Analg*. 2011;112:858–863.
14. Mimoz O, Frasca D, Médard A, Soubiron L, Debaene B, Dahyot-Fizelier C. Reliability of the HemoCue® hemoglobinometer in critically ill patients: a prospective observational study. *Minerva Anestesiol*. 2011;77:979–985.
15. Schapkaitz E, Mahlangu J. Point-of-care estimation of haemoglobin concentration in all age groups in clinical practice. *S Afr Fam Pract*. 2011;53:271–274.
16. Seguin P, Kleiber A, Chanavaz C, Morcet J, Mallédant Y. Determination of capillary hemoglobin levels using the HemoCue system in intensive care patients. *J Crit Care*. 2011;26:423–427.
17. Patel KP, Hay GW, Cheteri MK, Holt DW. Hemoglobin test result variability and cost analysis of eight different analyzers during open heart surgery. *J Extra Corpor Technol*. 2007;39:10–17.
18. Sanchis-Gomar F, Cortell-Ballester J, Pareja-Galeano H, Banfi G, Lippi G. Hemoglobin point-of-care testing: the HemoCue system. *J Lab Autom*. 2013;18:198–205.
19. Khanna P, Rajagopalan V, Singh G, Prabhakar H. A comparison of non-invasive versus invasive methods of haemoglobin estimation in patients undergoing intracranial surgery. *SAJA* 2014;20:160–163.
20. Rivas Chirino L, da Silva Viana J, Tavares CA, Palmeiro A, Oliveira FJ. Use of a blood gas analyzer to measure blood hemoglobin during liver transplantation: a study of 935 paired samples. *Transplant Proc*. 2006;38:810–811.
21. Myers GJ, Browne J. Point of care hematocrit and hemoglobin in cardiac surgery: a review. *Perfusion*. 2007;22:179–183.
22. Bénéteau-Burnat B, Bocque MC, Lorin A, Martin C, Vaubourdolle M. Evaluation of the blood gas analyzer Gem PREMIER 3000. *Clin Chem Lab Med*. 2004;42:96–101.
23. Gehring H, Hornberger C, Dibbelt L, et al. Accuracy of point-of-care-testing (POCT) for determining hemoglobin concentrations. *Acta Anaesthesiol Scand*. 2002;46:980–986.

24. Berkow L. Factors affecting hemoglobin measurement. *J Clin Monit Comput.* 2013;27:499–508.
25. Gramz J, Koerte P, Stein D. Managing the challenges in point-of-care testing: an ecosystem approach. *Point of Care* 2013;12:76–79.
26. Zwart A, van Assendelft OW, Bull BS, England JM, Lewis SM, Zijlstra WG. Recommendations for reference method for haemoglobinometry in human blood (ICSH standard 1995) and specifications for international haemiglobinocyanide standard (4th edition). *J Clin Pathol.* 1996;49:271–274.
27. Oddoze C, Lombard E, Portugal H. Stability study of 81 analytes in human whole blood, in serum and in plasma. *Clin Biochem.* 2012;45:464–469.
28. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135–160.
29. Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. *Stat Med.* 2008;27:778–780.
30. Krouwer JS, Monti KL. A simple, graphical method to evaluate laboratory assays. *Eur J Clin Chem Clin Biochem.* 1995;33:525–527.
31. Monti KL. Folded empirical distribution function curves-mountain plots. *Am Stat*. 1995;49:342–345.
32. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care.* 1987;10:622–628.
33. Freeman GH, Halton JH. Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika.* 1951;38:141–149.
34. American Society of Anesthesiologists Task Force on Perioperative Blood Transfusion and Adjuvant Therapies. Practice guidelines for perioperative blood transfusion and adjuvant therapies: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Blood Transfusion and Adjuvant Therapies. *Anesthesiology* 2006;105:198–208.
35. Maslow A, Bert A, Singh A, Sweeney J. Point-of-care hemoglobin/hematocrit testing: comparison of methodology and technology. *J Cardiothorac Vasc Anesth.* 2016;30:352–362.
36. Rigby AS. Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient. *Disabil Rehabil.* 2000;22:339–344.
37. Berkow L, Rotolo S, Mirski E. Continuous noninvasive hemoglobin monitoring during complex spine surgery. *Anesth Analg.* 2011;113:1396–1402.
38. Macknet MR, Allard M, Applegate RL II, Rook J. The accuracy of noninvasive and continuous total hemoglobin measurement by pulse CO-Oximetry in human subjects undergoing hemodilution. *Anesth Analg.* 2010;111:1424–1426.
39. Lu MJ, Zhong WH, Liu YX, Miao HZ, Li YC, Ji MH. Sample size for assessing agreement between two methods of measurement by Bland–Altman method. *Int J Biostat* 2016;12.