Precision, accuracy, and reliability of hemoglobin assessment with use of capillary blood[1–4]

Saul S Morris, Marie T Ruel, Roberta J Cohen, Kathryn G Dewey, Bénédicte de la Brière, and Mohammed N Hassan

ABSTRACT

Background: Previous research has suggested that there may be significant within-subject variability, both site-to-site and over time, in hemoglobin concentrations in capillary blood.

Objective: This study examined the reliability of the portable hemoglobinometer (PHM) system with use of capillary blood and the implications of errors of the magnitude found for the classification of anemia status in individuals and population groups. The precision and accuracy of the method with use of venous blood were also tested.

Design: Three empirical data sets were used to measure reliability, precision, and accuracy of the PHM system [2 from Honduras (n = 87 and 141); 1 from Bangladesh (n = 73)]. Simulation data were used to assess the implications of errors for screening individuals for anemia and to estimate anemia prevalence.

Results: High within-subject variability (unreliability) was identified when capillary blood from the left hand was compared with that from the right hand (CV: 6.3%) and when measurements were taken on 4 consecutive days (CV: 7.0%). Reliability was only 69% and 50%, respectively. Precision and accuracy, however, were very high (concordance coefficients of 0.99 and 0.98 and CV < 1%).

Conclusions: The simulation data showed that errors of the magnitude found due to unreliability can lead to misclassification of anemia status in individuals and small biases in anemia prevalence estimates. We recommend replicate sampling to reduce the influence of unreliability in the use of the PHM system with capillary blood. *Am J Clin Nutr* 1999;69:1243–8.

KEY WORDS Nutrition assessment, hemoglobin, error, reliability, precision, accuracy, capillary blood, hemoglobinometer, Honduras, Bangladesh, anemia

INTRODUCTION

Hemoglobin assessments are used widely to screen individuals for anemia, to draw inferences about the iron status of populations, and to evaluate responses to nutritional interventions. The portable hemoglobinometer (PHM) system has been widely used for these purposes in recent years because it is portable, requires only a small sample of capillary blood, is relatively inexpensive and simple to use, does not require access to refrigeration or even electricity, and gives immediate, digitally displayed results (1). The method is based on the photometric measurement of whole (undiluted) blood after conversion of hemoglobin to hemiglobinazide by reagents present in the system's special capillary action microcuvettes. Previous evaluations of this instrument have revealed high precision (freedom of measurement error), with CVs in the range of 0.9–1.3%. The instrument also has high accuracy (1–3): correlation coefficients comparing the portable system with standard methods have found measurements to be nearly perfect (r = 0.99) in laboratory settings, although less so (0.61–0.89) when assessments were carried out in routine health facility situations (3, 4).

Less well documented is the reliability of the PHM system, meaning the extent to which a measurement in a given individual is reproducible over time (5) or at a single time across different sites. One study with only 6 subjects reported large variations in the hemoglobin concentration of capillary blood samples obtained from different fingers of the same individual at the same time (6). Within-person CVs ranged from 3% to 7%, substantially higher than those reported by Statland et al (7) for hemoglobin measured in venous blood samples and analyzed by standard methods (2.5% for hour-to-hour variation and 2.6% for day-to-day variation). In another study (8), venous blood samples drawn from the same individuals on 2 different occasions and assessed by using an electronic counter showed within-person variances of 28.16 $(g/L)^2$ in males and 23.23 $(g/L)^2$ in females, corresponding to CVs of ≈3.5%.

If capillary blood sampling with the PHM method is to continue to be popular in field research in developing countries, the reliability of the approach must be confirmed. Failure to control for even relatively small unreliability will widen the distribution of hemoglobin values and may result in biases in estimates of anemia

[1] From the Food Consumption and Nutrition Division, International Food Policy Research Institute, Washington, DC; the Department of Nutrition, University of California, Davis; and the Institute of Nutrition and Food Science, University of Dhaka, Bangladesh.

[2] The ideas and opinions presented here are the responsibility of the authors and should in no way be attributed to the International Fund for Agricultural Development.

[3] Supported by the International Fund for Agricultural Development (TA grant 301-IFPRI).

[4] Reprints not available. Address correspondence to SS Morris, IFPRI-FCND, 2033 K Street NW, Washington, DC, 20006-1002. E-mail: s.morris@cgiar.org.

Received May 15, 1998.

Accepted for publication December 21, 1998.

TABLE 1
Characteristics of data sets and type of error measured[1]

Study and type of analysis	Location	Sample	Blood draws	Biochemical analyses
A, reliability (n = 87)	Honduras, western region	Women	Capillary, 1 each hand [2][2]	PHM [2]
B, reliability (n = 73)	Central Bangladesh	Females (9–49 y of age)	Capillary, consecutive days [4]	PHM [4]
C, precision and accuracy (n = 141)	San Pedro Sula, Honduras	Infants (4 mo)	Venous [1]	PHM [2], LM [2]

[1] PHM, portable hemoglobinometer. LM, laboratory method.
[2] Number of blood draws in brackets.

prevalence. Failure to control for unreliability may also bias associations with risk factors and estimates of response to interventions toward the null, thereby underestimating true effects.

This paper reports the use of data from 2 studies to measure unreliability of hemoglobin measurements with capillary blood in the PHM system and the use of data from a third study to estimate precision and accuracy of the method with use of venous blood. A simulation model was used to test the implications of errors of the magnitude found in our empirical data for screening individuals for anemia and estimating anemia prevalence.

SUBJECTS AND METHODS

Data

Data from 2 studies, 1 conducted in the western region of Honduras and 1 in Bangladesh, were used to estimate the reliability of the PHM (HemoCue; HemoCue Inc, Mission Viejo, CA) with use of capillary blood. A third data set from urban Honduras (San Pedro Sula) was used as a reference point to document the precision and accuracy of the method in a typical field setting in a developing country with use of venous blood. A summary of the characteristics of each sample is presented in **Table 1**.

Study A: western Honduras (altitude: 540–1740 m)

This study looked at food and nutrition security in western Honduras. Study subjects were women from 10 different rural communities and were either themselves members of or were closely related to male members of village-level development organizations receiving credit and technical assistance from a large international development project. In addition, the women were highly self-selected in that only a small percentage of all eligible women presented themselves to have their hemoglobin concentrations checked when invited to do so. Eighty-seven women had capillary blood sampled from the middle finger of each of their right and left hands (order unspecified). The 2 samples were taken within 5 min of each other by a single specially trained nurse. The fingertip was pierced by using the PHM system microlances, the first 2 drops of blood to appear were wiped away to stimulate blood flow, and blood was collected into a special microcuvette by capillary action. The cuvettes were analyzed immediately in the portable photometer. Ethical approval for the study was obtained from the government health administration of the local health region and verbal consent was obtained from the women coming to have their blood tested.

Study B: central Bangladesh (near sea level)

Study B was a study of the effect on micronutrient status of the adoption of 3 different packages of agricultural innovations that was conducted in the administrative regions of Saturia, Mymensingh, and Jessore. Seventy-three female subjects from 7 different villages (3 in Saturia, 2 in Mymensingh, and 2 in Jessore) were asked to contribute blood on 4 consecutive days; 1 woman was absent on the last day and therefore contributed only 3 measurements, and on 1 other occasion there was an interval of 2 d between consecutive measurements. Samples, from the left hand of right-handed women and from the right hand of left-handed women, were taken in the morning in Mymensingh and in the afternoon in Jessore and Saturia. Consecutive samples were not necessarily from the same finger. Blood sampling and analysis of hemoglobin by the PHM method was performed by specially trained doctors using the same method described for study A. Verbal consent was obtained from the women for this procedure after a careful explanation of its objectives and risks.

Study C: San Pedro Sula, Honduras (near sea level)

In this study of anemia in exclusively breast-fed infants, a sample of 141 infants aged 4 mo had 2 mL venous blood sampled from whichever arm had the most accessible vein. The infants came from poor families in the urban area of San Pedro Sula, the second largest city in Honduras. The blood was collected by a trained phlebotomist into a Monovette tube (Sarstedt, Newton, NC) containing EDTA as an anticoagulant. Immediately afterward, one drop of whole blood was transferred from the collection tube onto a small piece of parafilm, drawn by capillary action into 2 separate microcuvettes, and analyzed in the PHM by a specially trained laboratory assistant. At the same time, 0.5 mL whole blood was transferred to a glass tube and analyzed in duplicate within the hour in a hematology analyzer (Cell-Dyn 601; Abbott Diagnostics, Mount View, CA). Written, informed consent for the study procedures was obtained from the mothers of the infants within 1 wk of delivery.

In all 3 studies described above, special care was taken to ensure high-quality data from the PHM readings. The PHMs were checked each day by using the special microcuvettes provided by the manufacturer and sampling was conducted exactly as recommended by the manufacturer. In particular, each full cuvette was carefully checked for air bubbles and the plastic containers of unused cuvettes were immediately resealed after use to avoid degradation from heat, humidity, or both.

Simulation model

In addition to the 3 empirical data sets, we present results from simulated data with 20 000 observations and 2 variables: "true" hemoglobin status and apparent hemoglobin status in the presence of instrument imprecision and biological variability. The parameters for the simulation data were derived from the 3 empirical data sets. We assumed that true hemoglobin values were normally distributed about their mean with an SD of 10 g/L and that the within-subject variability was also normally distributed and uncorrelated with the true value. The mean of the true values was allowed to take 3 different values (110, 120, and 130 g/L) and the

within-person SD was allowed to vary from 4 g/L (CV of ≈3%) to 8 g/L (CV of 6–7%). The between-person SD of the true mean values was set at 10 g/L, roughly the average of the SDs found in our 3 empirical data sets. The simulation data were used to evaluate the effects of different levels of error on the sensitivity and specificity of a diagnosis of low hemoglobin (<120 g/L), as well as the predictive values of a positive (<120 g/L) and negative (≥120 g/L) screening result. We also assessed the magnitude of biases in anemia prevalence that would result from these levels of error.

Data analysis

Reliability is defined in this paper as the extent to which a measurement in a given individual is reproducible over time (5) or at a single time across different sites. *Unreliability* (1 − reliability) is assumed to be synonymous with within-subject variability and is the sum of imprecision (error due to measurement) and undependability (error due to short-term within-individual random fluctuations). *Accuracy* is defined as lack of error due to the instrument and refers to the comparison of the method with a gold standard.

Unreliability was estimated with data from studies A and B. Study A measured unreliability due to differences in hemoglobin concentration at different body sites and study B measured unreliability due to changes over time. Site-to-site unreliability was assessed by taking blood from the left and right hands of the same person within a period of 5 min and short-term temporal unreliability was assessed by taking blood from the same individual on 4 consecutive days. Within-subject SDs were estimated by using a one-way random-effects analysis of variance model with no covariates (9). The intraclass correlation coefficient (10) provided a direct measure of reliability (1 − unreliability). The within-subject CV (%) was calculated as follows:

$$\sigma_{within}/\text{mean} \times 100 \qquad (1)$$

For study A, which had only 2 measurements, we also calculated the mean of the paired differences and its SE, and the limits of agreement ($\bar{x} \pm 2$ SDs of the differences) (11).

Precision was assessed by contrasting repeat measures obtained by using the same analytic method and blood. The means of the paired differences, the limits of agreement, and the within-pair CVs were estimated (9, 11). Pearson correlation coefficients between pairs of readings are also presented for comparability with the published literature. A more appropriate measure of correspondence—also presented—is the concordance coefficient, which, in addition to measuring the agreement between 2 readings, also measures the departure from the 45° line through the origin (11, 12).

Accuracy was assessed by using the same statistical approach, but by comparing the PHM method with the standard laboratory method based on the same (venous) blood samples. All analyses were conducted by using STATA (version 5.0; STATA Corp, College Station, TX).

RESULTS

The 3 populations studied had different hemoglobin profiles (**Table 2**), which was to be expected considering that they included different age and sex groups and were conducted at different altitudes and on different continents. However, within each study the means and SDs of the various hemoglobin measures were very similar.

TABLE 2
Hemoglobin status of subjects from the 3 different studies[1]

Study and method	Hemoglobin concentration
	g/L
A, Honduras, western region (*n* = 87)	
Capillary (PHM)	
Left hand	131.0 ± 14.9
Right hand	130.5 ± 14.9
B, Central Bangladesh (*n* = 73)	
Capillary (PHM)	
Day 1	117.8 ± 11.9
Day 2	116.4 ± 11.4
Day 3	117.4 ± 11.1
Day 4	117.0 ± 11.6
C, San Pedro Sula, Honduras (*n* = 141)	
Venous (LM)	
First duplicate	112.0 ± 7.8
Second duplicate	111.9 ± 7.9
Venous (PHM)	
First duplicate	111.7 ± 7.8
Second duplicate	111.8 ± 7.8

[1] $\bar{x}$ ± SD. PHM, portable hemoglobinometer; LM, laboratory method.

Information on reliability (for studies A and B) and precision and accuracy (for study C) is presented in **Table 3**. Large discrepancies were found between the PHM values obtained from capillary blood sampled from the left and right hands of women in study A. Although the mean of the paired differences between the 2 samples did not differ significantly from zero, the within-subject CV was 6.3% and the reliability was just 69.3%. The concordance between the 2 values was 0.69. The wide limits of agreement indicate that 2 samples from different fingers of the same person could have hemoglobin concentrations differing by ≥20 g/L. The estimated within-subject SD was 8.3 g/L (not shown). Similarly, in study B, in which finger-prick blood samples were taken from the same individuals on 4 consecutive days, reliability was just 49.7%. The within-subject CV was 7.0% and the estimated within-subject SD was 8.2 g/L (not shown).

Precision of the PHM method was excellent and comparable with that of the laboratory method (concordance coefficients of 0.99 for both methods; CVs of 0.8% and 0.6%, respectively). The accuracy of the PHM method compared with the laboratory method gold standard was also excellent when venous blood was used to fill the PHM cuvettes (concordance coefficient: 0.98).

The likely effects of different levels of error on the sensitivity and specificity of a diagnosis of low hemoglobin (<120 g/L) with use of the simulation data set described in the Methods section are given in **Table 4**. The predictive values of a positive (<120 g/L) and negative (≥120 g/L) result with use of the capillary sample are also shown, as are the estimated population prevalences of low hemoglobin concentrations in each case. This analysis showed that when the population mean hemoglobin concentration is equal to the cutoff point for diagnosis of anemia (120 g/L), the true prevalence of low hemoglobin is equal to 50% and use of a single capillary sample provides an unbiased estimate of this quantity, irrespective of the level of within-subject error. At the level of individuals, an adequately sensitive and specific (≥80%) diagnosis of low hemoglobin is obtained even with within-subject error as high as 6.0 g/L (SD, between sampling sites). This scenario corresponds with a within-person CV of 5%.

TABLE 3
Assessment of reliability, precision, and accuracy[1]

Study and method	Paired difference	Limits of agreement	CV	Correlation coefficient[2]	Concordance coefficient
	g/L	*g/L*	*%*		
Reliability					
A, Honduras, western region, PHM: left hand compared with right hand	0.49 ± 1.26[3]	−23.0, 24.0[4]	6.3[5]	0.69[6]	0.69
B, Central Bangladesh, PHM: 4 consecutive days	NA	NA	7.0[5]	0.50[6]	NA
Precision[7]					
C, San Pedro Sula, Honduras					
LM: duplicate assays	0.08 ± 0.10	−2.36, 2.52	0.6	0.99	0.99
PHM: duplicate assays	−0.11 ± 0.08	−2.12, 1.90	0.8	0.99	0.99
Accuracy (PHM compared with LM)					
C, San Pedro Sula, Honduras: mean of duplicates compared with mean of duplicates[7]	0.23 ± 0.13	−2.93, 3.38	NA	0.98	0.98

[1] PHM, portable hemoglobinometer; LM, laboratory method; NA, not applicable.
[2] Pearson correlation coefficient unless otherwise indicated.
[3] $\bar{x}$ ± SEM.
[4] $\bar{x}$ ± 2 SDs.
[5] Within-subject.
[6] Intraclass correlation coefficient (reliability).
[7] All laboratory analyses used the same venous blood sample.

With a level of error of 8 g/L (CV: 6.7%), sensitivity, specificity, and predictive values are all in the range of 75–80%.

With the population mean hemoglobin concentration 1 SD below the anemia cutoff point, all levels of error result in biased estimates (underestimates) of the true prevalence of anemia (84%). The absolute magnitude of the bias increases with increasing levels of error, reaching 6% when the between-finger SD is set at 8 g/L. The specificity, and particularly the predictive value of a negative result, are reduced so that the predictive value of a negative result would not exceed 53% when the level of error was high (eg, 8.0 g/L; CV: 7.3%). When the population mean is 1 SD above the mean, the sensitivity and the predictive value of a positive result are similarly affected, and overestimates of the true prevalence of anemia (16%) are generated. Thus, the absolute size of the bias in the estimate of the prevalence of anemia is dependent both on the reliability of the method and on the true prevalence in the population. This bias never exceeds the values shown in the table for true prevalences of 16% and 84%, because true prevalences closer to 50%, 0%, or 100% are all associated with smaller biases (data not shown).

DISCUSSION

Anemia affects 30–40% of women and young children worldwide (13) and has major effects on cognitive development, work performance, and pregnancy outcome (14). Once diagnosed, it can be treated with a combination of iron and folate supplementation and deworming and antimalarials, as appropriate. The PHM offers

TABLE 4
Simulation model of bias in estimates of the prevalence of anemia, and sensitivity, specificity, and predictive values of hemoglobin screening for anemia[1]

Simulation variables	Between-finger variation in hemoglobin (SD)		
	4 g/L	6 g/L	8 g/L
		%	
Population mean (±SD): 120 ± 10 g/L (true prevalence of Hb < 120 g/L = 50%)			
Apparent prevalence of Hb < 120 g/L	50	50	50
Sensitivity	88	82	77
Specificity	88	84	79
Predictive value of a positive result	88	83	79
Predictive value of a negative result	88	83	79
Population mean (±SD): 110 ± 10 g/L (true prevalence of Hb < 120 g/L = 84%)			
Apparent prevalence of Hb < 120 g/L	82	80	78
Sensitivity	94	91	87
Specificity	82	77	72
Predictive value of a positive result	96	95	94
Predictive value of a negative result	72	61	53
Population mean (±SD): 130 ± 10 g/L (true prevalence of Hb < 120 g/L = 16%)			
Apparent prevalence of Hb < 120 g/L	18	20	22
Sensitivity	80	75	72
Specificity	95	91	88
Predictive value of a positive result	73	61	52
Predictive value of a negative result	96	95	94

[1] Hb, hemoglobin concentration; positive result, Hb < 120 g/L; negative result, Hb ≥ 120 g/L.

public health practitioners a means of screening for anemia that is rapid, easy to use, and inexpensive and was shown previously to be both precise and accurate when compared with standard laboratory procedures (1–3). Our study in San Pedro Sula, Honduras (study C), tends to confirm that these findings are also valid for settings in developing countries, although this study examined only the precision and accuracy of the PHM method when used with venous blood. Undoubtedly, much of the appeal of the PHM method derives from the fact that the blood can be obtained from a finger prick, which is less resource-intensive than venipuncture and better accepted by patients and the community. Until now, scant attention has been paid to the reliability of hemoglobin assessments of capillary blood or to the programmatic implications of short-term, within-subject fluctuations in hemoglobin concentrations.

Our studies in the western region of Honduras and in Bangladesh (studies A and B) showed large within-subject variation in the hemoglobin concentration of blood sampled from fingertip capillaries. In the first case, this variation reflected differences in the site of sampling the hemoglobin concentration of capillary blood (left or right hand), whereas in the second case it reflected short-term, within-subject temporal variation over a 4-d period (and possibly the site of sampling also). The CVs estimated from these data—6.3% in study A and 7.0% in study B—are strikingly consistent with the findings of Boulton et al (6), although only 6 individuals were investigated in that study. On the other hand, studies quantifying the within-person variability of hemoglobin concentrations of venous blood (7, 8) have found lower within-subject variability (2.2–3.5%).

We believe that it is biologically plausible that capillary blood would be more variable in hemoglobin concentration than venous blood because inclusion of extracellular fluid would decrease the concentration of components present in the red cell fraction, and the amount of extracellular fluid present in finger-prick samples is likely to be very sensitive to the technician's handling of the patient's finger. Although technicians were rigorously trained to avoid "milking" the finger in all of our studies, some contamination with extracellular fluid seems inevitable in any programmatic application of the PHM or capillary blood sampling method. More research is needed to establish whether within-subject CVs in capillary blood hemoglobin of the order of 6–7% represent the true norm for biological variation or whether further standardization of the sampling method can reduce this variability. Our analysis of study C suggested that imprecision (instrument error) can be ruled out as a significant component of the error due to unreliability (sum of error due to undependability + imprecision).

Our simulation model suggested that estimates of the prevalence of low hemoglobin are unbiased when the true prevalence is 50%, regardless of the degree of unreliability of the method. Bairagi (15) used a theoretical approach to show that the absolute size of the bias increases until the diagnostic cutoff is equal to the population mean ± 1 SD (equivalent to true prevalences of 16% or 84%) and then falls again as the difference between the cutoff and the population mean becomes even greater (true prevalences $<16\%$ or $>84\%$). We have shown that the levels of unreliability found in our 2 data sets would result in absolute biases in the prevalence of low hemoglobin values of no more than 6% if the mean population hemoglobin were 1 SD above or below the diagnostic cutoff. A bias of this magnitude will often be acceptable; when it is not, it may be possible to assess the scale of error in a small subsample and use existing methods to correct the prevalence estimate (16).

We also examined the likely consequences of error in hemoglobin concentrations in capillary blood on the screening for anemia and found that if the population values were symmetrically distributed around the cutoff value for anemia, a single fingertip sample would provide an adequately sensitive and specific ($\geq 80\%$) diagnosis of low hemoglobin concentration, even if the level of error was at the top of the range considered (CV from $\approx 3\%$ to 6–7%). However, if the population mean were 1 SD below the cutoff point, specificity (and the predictive value of a negative, or high, result) would be jeopardized, whereas with the population mean 1 SD above the mean, sensitivity and the predictive value of a positive (low) result would be similarly affected. As explained by Himes (17), the reliability of measures subject to within-subject error can be increased by using the mean of several replicate measures instead. Indeed, the Spearman-Brown Prophesy formula (10) allows investigators to calculate how many replicates would be needed to achieve a desired level of reliability. Our data from western Honduras (study A) indicated that no less than 4 replicates would be required to estimate individual hemoglobin concentrations with 90% reliability, whereas our Bangladesh data (study B), which included within-week temporal variation, suggested that 9 replicates would be required. Although we found double sampling to be well accepted at our western Honduras study site, we doubt that >2 replicates would be feasible in most circumstances.

In conclusion, we believe that the PHM method with use of capillary blood samples will continue to be a useful screening tool for anemia, but that unreliability poses a significant problem whenever individual-level estimates are required. In these circumstances, measures such as replicate sampling should be used to reduce the effect of within-subject error, and even then, residual unreliability should be assumed.

We thank Bo Lönnerdal, Howarth Bouis, and Juan Manuel Medina Banegas for their contributions to the studies reported and Suzanne Filteau for insights into possible biological mechanisms.

REFERENCES

1. Bridges N, Parvin RM, van Assendelft OW. Evaluation of a new system for hemoglobin measurement. Am Clin Products Rev 1987;April:22–5.
2. Cohen AR, Seidl-Friedman J. HemoCue system for hemoglobin measurement. Evaluation in anemic and nonanemic children. Am J Clin Pathol 1988;90:302–5.
3. von Schenck H, Falkensson M, Lundberg B. Evaluation of "HemoCue," a new device for determining hemoglobin. Clin Chem 1986;32:526–9.
4. Neville RG. Evaluation of portable haemoglobinometer in general practice. BMJ 1987;294:1263–5.
5. Marks GC, Habicht JP, Mueller WH. Reliability, dependability, and precision of anthropometric measurements. The Second National Health and Nutrition Examination Survey 1976–1980. Am J Epidemiol 1989;130:578–87.
6. Boulton FE, Nightingale MJ, Reynolds W. Improved strategy for screening prospective blood donors for anaemia. Transfus Med 1994;4:221–5.
7. Statland BE, Winkel P, Harris SC, Burdsall MJ, Saunders AM. Evaluation of biologic sources of variation of leukocyte counts and other hematologic quantities using very precise automated analyzers. Am J Clin Pathol 1978;69:48–54.
8. Looker AC, Sempos CT, Liu K, Johnson CL, Gunter EW. Within-person variance in biochemical indicators of iron status: effects on prevalence estimates. Am J Clin Nutr 1990;52:541–7.
9. Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames, IA: Iowa State University Press, 1980.
10. STATACORP. STATA reference manual. Release 5. Vol 2 G-O. College Station, TX: Stata Press, 1997.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
12. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–68.
13. DeMaeyer E, Adiels-Tegman M. The prevalence of anaemia in the world. World Health Stat Q 1985;38:302–16.
14. Yip R, Dallman PR. Iron. In: Ziegler EE, Filer LJ Jr, eds. Present knowledge in nutrition. 7th ed. Washington, DC: ILSI Press, 1996:277–92.
15. Bairagi R. Effects of bias and random error in anthropometry and in age on estimation of malnutrition. Am J Epidemiol 1986;123:185–91.
16. Subcommittee on Criteria for Dietary Evaluation, Coordinating Committee on Evaluation of Food Consumption Surveys, Food and Nutrition Board, Commission on Life Sciences, National Research Council. Nutrient adequacy. Assessment using food consumption surveys. Washington, DC: National Academy Press, 1986.
17. Himes JH. Reliability of anthropometric methods and replicate measurements. Am J Phys Anthropol 1989;79:77–80.