Bulletin of the World Health Organization, 61 (6): 957-965 (1983) © World Health Organization 1983

Haemoglobin measurement: the reliability of some simple techniques for use in a primary health care setting

W. VAN LERBERGHE,[1] G. KEEGELS,[2] G. CORNELIS,[2] C. ANCONA,[2] E. MANGELSCHOTS,[3] & H. VAN BALEN[4]

The reliability of five simple techniques for the estimation of blood haemoglobin levels was studied under laboratory conditions. The methods tested were the Sahli technique, the Tallqvist method, and the Lovibond comparator with undiluted blood, blood diluted in Drabkin solution, and blood diluted in modified Drabkin solution. The Lovibond comparator technique using undiluted blood was satisfactory in terms of precision and accuracy, and seems suitable for the monitoring of individual patients and for screening for anaemia. The other techniques were less reliable.

A detailed procedure was worked out for estimation of the number of screening errors that can be expected with each of these techniques in any epidemiological situation.

Measurement of haemoglobin concentration is an important routine procedure in developing countries, but little is known about the results obtainable with simple techniques used in primary health care for monitoring patients under treatment and for screening for anaemia. These techniques should be reliable, cheap, simple, and easy to use, and should not require any electric power supply.

Previous studies evaluating such laboratory instruments (*1-4*) have provided little information on their reliability in specific epidemiological situations. This paper presents the results obtained with five simple techniques—the Sahli method, the Tallqvist method, and the Lovibond comparator used with blood diluted in the original Drabkin solution, in the modified Drabkin solution, and undiluted. The aim of the study was to determine the reliability of each of these techniques for monitoring the haemoglobin concentration of individual patients, and their sensitivity, specificity, and predictive value when they are used for screening populations for anaemia.

MATERIALS AND METHODS

Screening techniques[a]

The five methods used to estimate haemoglobin levels are described in detail below.

The Sahli method measures acid haematin. A sample of 0.02 ml of whole blood, taken with an automatic standardized manual pipette, is mixed with a small quantity of 0.1 mol/litre HCl. After 3 minutes, 0.1 mol/litre HCl is added drop by drop and mixed, until the colour of the solution matches the colour of two identical standards placed to the left and right of the dilution tube. Haemoglobin concentration is then read from the graduated scale on the dilution tube.

The Lovibond-Drabkin technique measures cyanmethaemoglobin, and makes use of a Lovibond comparator, original Drabkin solution, and discs 5/40 and 5/40X. A volume of 0.04 ml of whole blood is taken with an automatic standardized manual pipette, mixed with 10 ml of Drabkin solution, and allowed to stand for at least 3 minutes. The tube is then placed in the comparator, along with a tube containing pure Drabkin solution. The disc is then rotated to find the colour standard on the disc that matches the colour of the sample solution. The haemoglobin concentration is interpolated from the nearest corresponding values on the disc.

The Lovibond-modified Drabkin technique, using the Lovibond comparator, modified Drabkin solution and discs 5/40 and 5/40X, also measures cyanmethaemoglobin. The procedure is identical to that with the original Drabkin solution.

The Lovibond-undiluted technique uses undiluted blood placed in a separately provided cell of the Lovibond comparator, and disks 5/8 A and B. It compares the colour of 0.03 ml of whole blood with the colour standards of the discs. Before filling, the cell is cleaned and degreased with water and alcohol. The haemoglobin concentration is interpolated from the nearest corresponding values on the discs.

* From the Unit for Research and Education in Public Health, Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerp, Belgium.

[1] Assistant.

[2] Assistant libre.

[3] Chief Laboratory Technician.

[4] Unit Chief.

[a] All haemoglobinometers were manufactured either by Assistant Co., Sondheim Rhön, Federal Republic of Germany; or Tintometer Ltd, Salisbury, England.

The Tallqvist method. A drop of whole blood is placed on a strip of blotting paper. When the sheen of the blood has disappeared, the colour is compared with a set of colour standards on paper. The haemoglobin concentration is interpolated from the values corresponding to the closest colour standards.

General procedure

Measurements of haemoglobin level were made by five observers, two of whom were qualified professional laboratory workers with considerable experience. The other three were health workers with limited experience of laboratory procedures. All observers familiarized themselves with the different techniques during three preliminary sessions.

Each observer made two measurements on each of 10 blood samples; (one observer made measurements on 2 sets of samples using the Lovibond-Drabkin and modified Drabkin techniques). A different set of blood samples was used for each technique. The blood samples had a haemoglobin concentration ranging from 30 to 150 g/l, and were obtained by diluting blood, collected the same day in conventional 5-ml EDTA[b]-coated tubes, with its autologous plasma, obtained by centrifugation. The haemoglobin content of the samples was determined from automatized measurements[c] of cyanmethaemoglobin by professional experienced laboratory personnel. Two such measurements were done for each sample, one between 11h 00 and 12h 00, before the observers started their estimations, and one afterwards, at 16h 30. The mean of these two measurements was taken as the reference haemoglobin concentration for each sample. The reference values were not known by the observers, who were given no information about the range of values of the samples.

The 10 samples were first assessed in a random order and the results collected. A new random identification number was then allocated to each sample, and the 10 samples were returned to the observers for the second measurement. This procedure ensured blindness of the second measurement with respect to the first. Throughout the experiment the samples were kept on an electric shaking apparatus of the toppling type.

Daylight from a north-eastern window was used as the light source, without a light-diffusing screen; the light intensity was thus not uniform.

RESULTS

Fig. 1 shows the mean differences between the two measurements of each sample, for each observer and

[b] Ethylenediaminetetraacetic acid.

[c] Using a haemoglobinometer manufactured by Coulter Electronics Co., Harpenden, Hertfordshire, England.

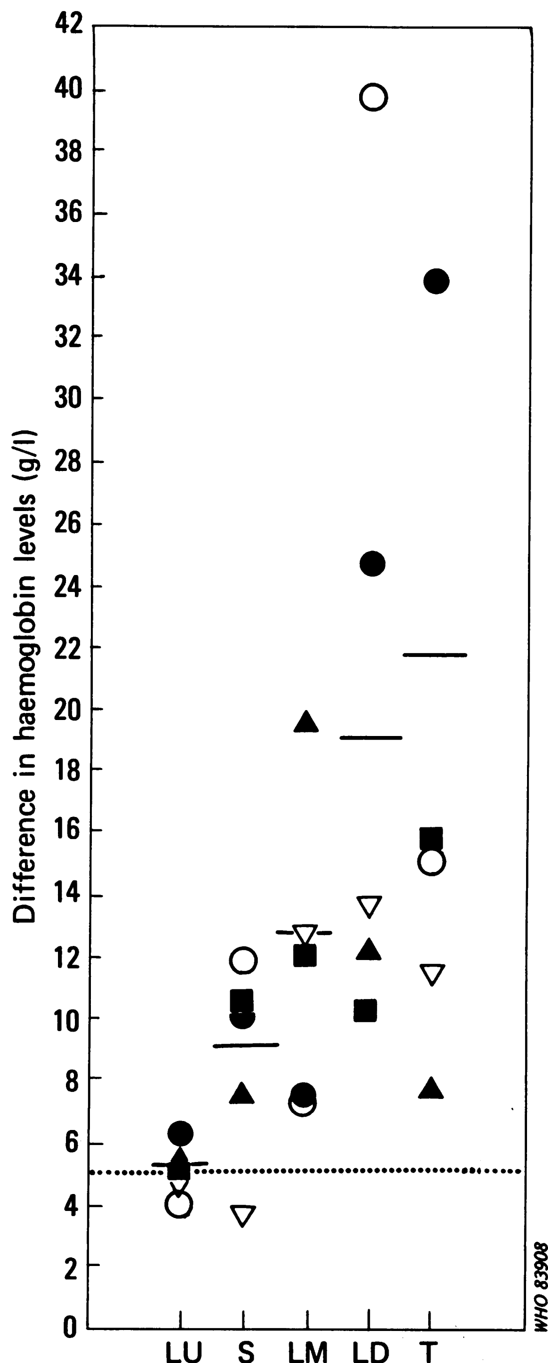

Fig. 1. Mean absolute difference between the two measurements on each blood sample, showing results for each observer and the overall mean for each technique (horizontal line). The dotted line represents the mean absolute difference between readings taken with the automatized photoelectric method. Key: LU = Lovibond-undiluted; S = Sahli; LM = Lovibond-modified Drabkin; LD = Lovibond-Drabkin; T = Tallqvist.

each technique. This permits a comparison of the precision achieved by each observer with each technique. The overall mean difference for each technique reflects the precision of the technique.

The accuracy of each observer with each technique can be seen in Fig. 2, which shows the mean difference between the measurements and the relevant reference value; this gives an indication of interobserver variability of accuracy.

A better representation of the accuracy and precision of each technique is given in Fig. 3-7, which show the scatter diagrams of the individual measurements, with the corresponding regression lines. Analysis of the residuals yielded no evidence for curvilinearity or dependence of the variance on the haemoglobin concentration, except in the case of the Lovibond-Drabkin technique, where variance appeared to increase at higher haemoglobin values.

The departure of the regression line from the line of perfect agreement is an indicator of the accuracy of

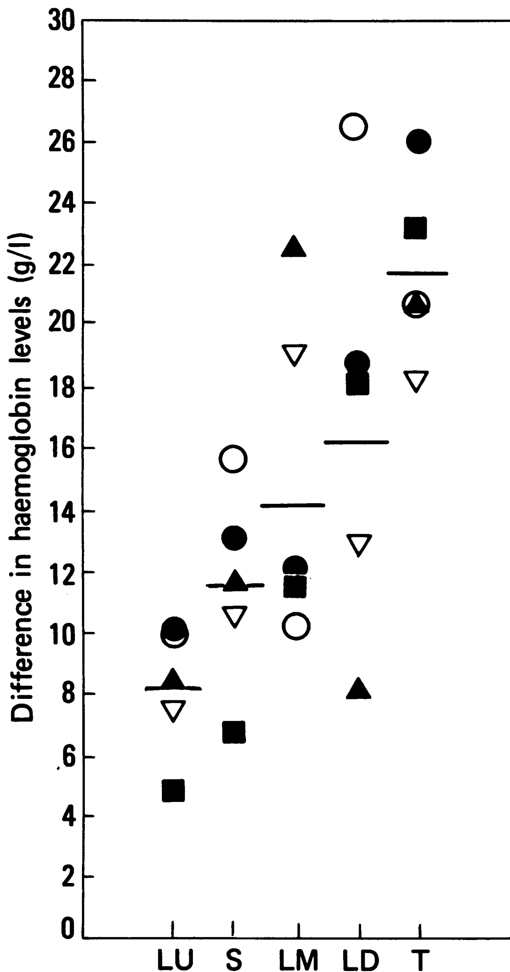

Fig. 2. Mean absolute difference between the measured and the reference values for the blood samples, showing results for each reader, and the overall mean for each technique (horizontal line). For key, see legend to Fig. 1.

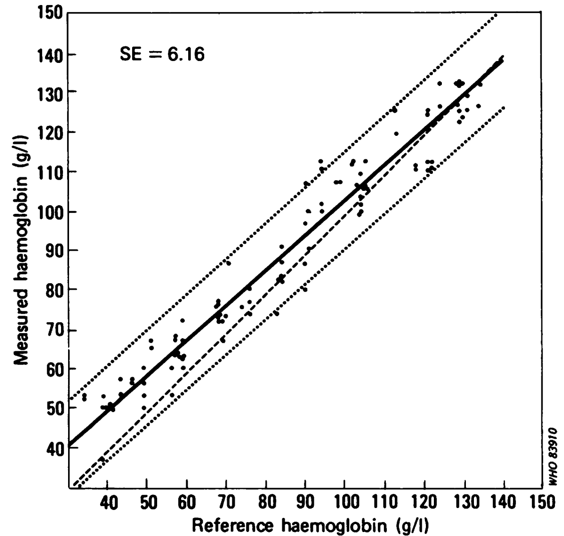

Fig. 3. Scatter diagram of all measurements made with the Lovibond-undiluted technique, showing the regression line and 95% confidence belt (dotted lines). The broken line is the line of perfect agreement.

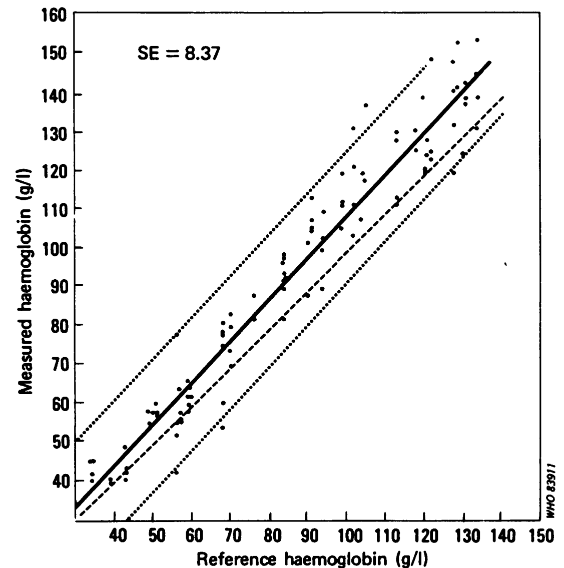

Fig. 4. Scatter diagram of all measurements made with the Sahli technique, showing the regression line and 95% confidence belt (dotted lines). The broken line is the line of perfect agreement.

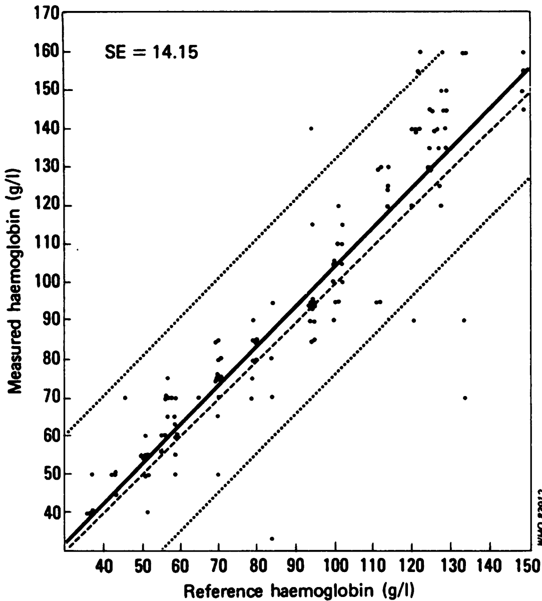

Fig. 5. Scatter diagram of all measurements made with the Lovibond-modified Drabkin technique, showing the regression line and 95% confidence belt (dotted lines). The broken line is the line of perfect agreement.

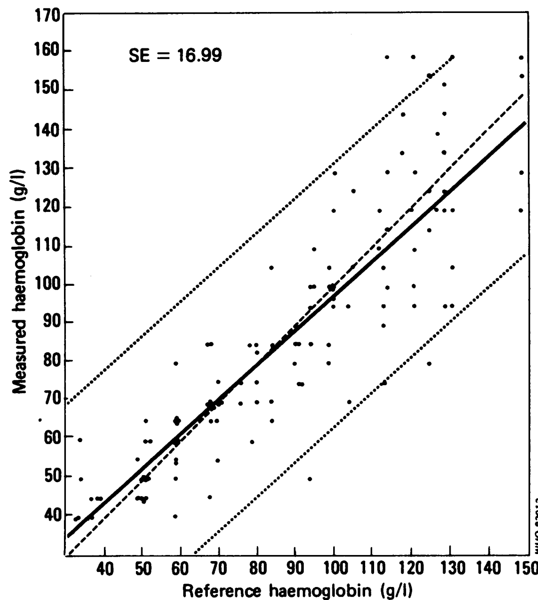

Fig. 6. Scatter diagram of all measurements made with the Lovibond-Drabkin technique, showing the regression line and 95% confidence belt (dotted lines). The broken line is the line of perfect agreement.

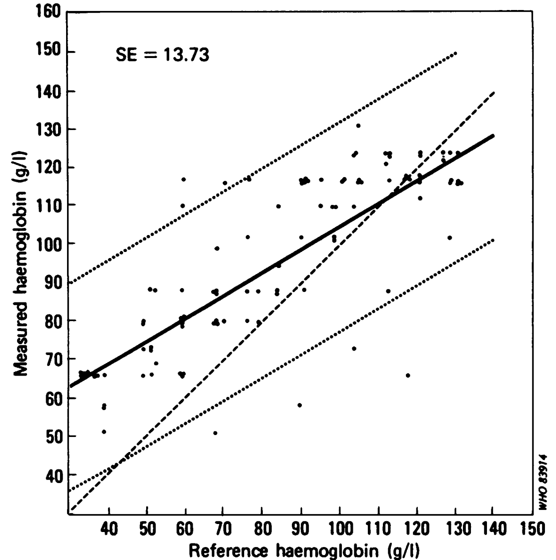

Fig. 7. Scatter diagram of measurements made with the Tallqvist technique, showing the regression line and 95% confidence belt (dotted lines). The broken line is the line of perfect agreement.

the technique. The 95% confidence belt around the predicted measurements is calculated from the standard error of the estimate (SE); its width is indicative of the variability of accuracy, which is not dependent on bias but on the lack of precision. There were significant differences between the SEs for each technique (F-test, $P < 0.05$), except the Tallqvist and Lovibond-modified Drabkin techniques.

DISCUSSION

The techniques tested were chosen because they are simple and require no electric light source, two conditions that are essential for instruments used for field work in developing countries. Both the Sahli and the Lovibond-modified Drabkin techniques are commonly used in developing countries. The Lovibond-Drabkin technique is not often recommended, but it was included in the present study because in many places it is easier to obtain the original than the modified Drabkin solution. The Lovibond-undiluted and Tallqvist techniques were included because they do not require dilution of the blood, a commendable advantage under field conditions. They offer a further advantage in that they do not require fragile cuvettes and pipettes.

The tests were carried out in a way that reflected field procedures, except that automatic standardized

manual pipettes were used for the dilutions. This permitted an evaluation of the technique under optimal conditions. The use of normal (mouth-suction) pipettes would no doubt give inferior results, particularly as the pipettes used in operational circumstances are often of inferior quality. Thus it should be kept in mind throughout that the results for the techniques using diluted blood may give an over-optimistic impression of what would happen in the field.

It should also be noted that all the readings were done on venous rather than capillary blood. The latter is recommended by the manufacturers and might give better results (*5*).

Precision and accuracy: interobserver variability and influence of haemoglobin concentration

The Lovibond comparator technique using undiluted blood was by far the most precise of the methods tested. The confidence belt around its regression line was significantly smaller than those for the other techniques. The mean absolute difference between measurements was only slightly higher than that for the automatic spectrophotometer. Moreover, the difference in precision between the observers was negligible. The Sahli technique also performed quite well, the mean absolute difference remaining smaller than 10 g/l. The other techniques, however, were markedly less precise, both in terms of average results and of interobserver variation. It should be noted that observer precision did not seem to be related to experience of laboratory work, but rather to the care taken in executing the techniques.

The Lovibond-undiluted technique was also the most accurate, with a mean absolute difference of 8.0 g/l between the measurements and their reference values. The difference in accuracy between the observers was also smallest for this technique. The Lovibond-undiluted technique was followed closely by the Sahli method. The Lovibond techniques that used diluted blood were less accurate and showed large differences in accuracy between observers. The Tallqvist method was consistently inaccurate.

The scatter diagrams (Fig. 3–7) show that for both the Lovibond-modified Drabkin and Lovibond-Drabkin techniques, precision did not change significantly with the haemoglobin concentration (although, as was remarked earlier, the Drabkin method appears to be somewhat less precise for higher values). The haemoglobin concentration did not have a significant impact on accuracy (except in the Tallqvist technique, which is accurate only for a very small haemoglobin range, around 110 g/l), although the Lovibond-undiluted technique seems somewhat less accurate in the lower values and the Sahli method in the higher.

Monitoring of individual patients

Table 1 shows, for each technique, the minimum difference between two measurements that cannot be ascribed to random errors of measurement ($P < 0.05$). As these differences have been calculated on the basis of the SE of the regression equation for each technique, they are expressed in g of haemoglobin per litre, as measured by the technique. The corresponding "true" minimum difference, calculated from the regression equation, is also given in Table 1.

It has been suggested (e.g., by the manufacturers of the Lovibond comparator) that in order to obtain an accurate estimate of haemoglobin level, one should make two readings on each blood sample and take the mean value. Regression lines were calculated for the means of the paired measurements and the reference values, and the minimum difference that cannot be attributed to error is also shown in Table 1. As could be expected, there is a marked improvement only for the techniques with poor precision.

Only the Lovibond-undiluted and Sahli techniques would permit reliable monitoring of haemoglobin changes in individual patients. If the estimations are

Table 1. The minimum difference between two estimations, based on single or double measurements, that cannot be attributed to measurement error ($P < 0.05$) and the corresponding "true" haemoglobin concentration

Technique	Single measurement		Double measurement	
	Minimum measured difference (g/l)	Corresponding "true" difference (g/l)	Minimum measured difference (g/l)	Corresponding "true" difference (g/l)
Lovibond-undiluted	17	20	16	18
Sahli	23	22	20	18
Lovibond-modified Drabkin	40	38	36	35
Lovibond-Drabkin	48	53	37	41
Tallqvist	39	64	28	47

made on the basis of two measurements, both techniques can detect differences of haemoglobin concentration of 18 g/l. The other techniques clearly are not satisfactory, even when the determination of haemoglobin concentration is based on two consecutive measurements.

Screening for anaemia

Screening for anaemia with any of these techniques will result in a certain number of false positives and false negatives. Their actual number can be calculated and depends on the slope of the regression line, the dispersion around it, and the cut-off level, relative to the frequency distribution of haemoglobin levels in the population. Details on how the proportion of false positives and false negatives can be calculated for each haemoglobin level are given in Annex 1.

Table 2 presents the distribution of the proportion of false positives and negatives to be expected in each haemoglobin class relative to a given cut-off point.

It is not possible to calculate the sensitivity, specificity, and predictive values of each of these techniques without taking into account the distribution of the haemoglobin concentration in the population being screened. This is illustrated in Fig. 8, which shows sensitivity, specificity, and predictive values of positive and negative test results that would be obtained, with each of these techniques, using two cut-

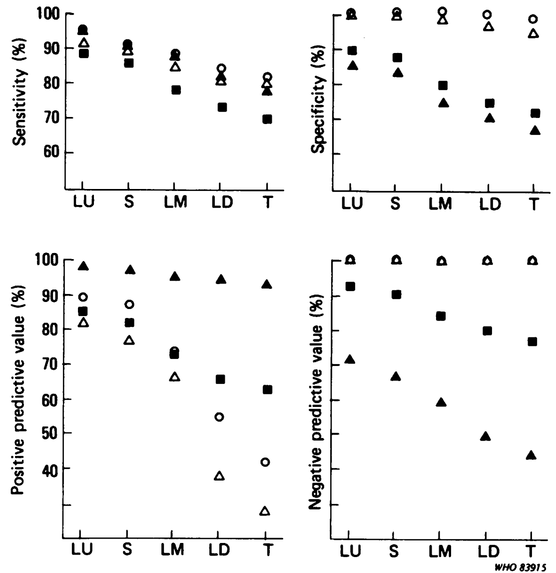

Fig. 8. Sensitivity, specificity, and predictive values of positive and negative results of the five techniques when screening (*a*) a population with a high prevalence of anaemia, at cut-off levels 90 g/l (▲) and 110 g/l (■), and (*b*) one with a low prevalence of anaemia at the same cut-off levels of 90 g/l (O) and 110 g/l (Δ). For key, see legend to Fig. 1.

Table 2. The expected proportion of false positives (above the dotted line) and false negatives (below the dotted line) within each haemoglobin class when screening for anaemia with a cut-off point of 100 g/l[a]

Haemoglobin class (g/l)	Lovibond-undiluted	Sahli	Lovibond-modified Drabkin	Lovibond-Drabkin	Tallqvist
150–159				0.002	0.008
140–149				0.009	0.024
130–139			0.005	0.031	0.064
120–129		0.001	0.034	0.092	0.136
110–119	0.016	0.035	0.136	0.212	0.255
100–109	0.239	0.274	0.356	0.396	0.413
90– 99	0.239	0.274	0.356	0.396	0.413
80– 89	0.016	0.035	0.136	0.212	0.255
70– 79		0.001	0.034	0.092	0.136
60– 69			0.005	0.031	0.064
50– 59				0.009	0.024
40– 49				0.002	0.008
30– 39					

[a] For other cut-off points, the distribution of false results around the cut-off value will be identical.

off levels: 110 g/l, the WHO-recommended cut-off point for the screening of pregnant women, and 90 g/l, which in many developing countries is a more realistic criterion. These cut-off points are applied to a population of pregnant women from the United Republic of Tanzania with a high prevalence of low haemoglobin values (*6*) (86% are below 110 g/l, 41% below 90 g/l), and to a population of adult men from Central America with a low prevalence of low haemoglobin values (*7*) (2.6% were below 110 g/l and 1.2% were below 90 g/l).

Depending on the epidemiological situation, the cut-off point, and the technique, the sensitivity ranges between 94.2% (Lovibond-undiluted, United Republic of Tanzania, 110 g/l) and 70.4% (Tallqvist, United Republic of Tanzania, 90 g/l); for the specificity, the maximum and minimum figures are 99.9% and 66.3%. The differences in predictive values are even greater: as many as 97.6% of positive results are true positives when the Lovibond-undiluted technique is used in the Tanzanian population at a cut-off point of 110 g/l. This figure goes down to 82.0% when the same test is used, at the same cut-off point, in the Central American population. The lowest predictive values for positive tests are obtained with the Tallqvist technique, i.e., 27.6–93.5%. Predictive values for negative tests range from 99.9% to 34.0%, depending on technique, epidemiological setting and cut-off point.

The Lovibond-undiluted technique thus undoubtedly performs best; moreover, it is an easy and quick method, the apparatus is robust and does not require a regular supply of chemicals, and it is not more expensive than the Lovibond techniques that use diluted blood. The Tallqvist technique is much cheaper in terms of capital investment, but this advantage is negligible when one considers its inferior performance and the fact that a constant supply of papers has to be ensured; the difference in price would soon be outweighed by the cost of unnecessary transfusions, referrals, or other measures that the use of the Tallqvist technique would involve. The Lovibond techniques that use diluted blood are as expensive as the Lovibond-undiluted method; they are less precise, less accurate, less reliable and more cumbersome in use. Only the Sahli technique could be considered as a valid alternative to the Lovibond-undiluted method, but it has the disadvantages that it is more cumbersome in use, and requires a supply of hydrochloric acid. The performance of the Sahli technique is expected to be much less satisfactory under operational circumstances, chiefly because of dilution problems involved in the use of ordinary manual pipettes instead of automatic standardized pipettes.

The Lovibond-undiluted method therefore appears to be a satisfactory technique for routine screening. In operational circumstances, its results are likely to be comparable to those obtained with a (substantially more expensive) colorimeter of the EEL type, which requires a power supply. The optimal cut-off point to be used with this technique will depend on the objectives of the screening, and the cost of treating true and false positives and of not treating true and false negatives. This requires a knowledge not only of the reliability of the technique, but also of the local epidemiological situation.

In situations with few resources, and where the technical management of a photoelectrical technique is not of the highest quality (e.g., because of problems with the power supply or with the necessary dilutions, etc.), simple techniques should certainly be considered. In many operational circumstances they might actually perform better than more sophisticated methods.

The relative reliability of a technique is relevant only in as much as the consequences of a wrong decision are important. Only when the reliability of the technique has been quantified is it possible to decide whether these consequences are acceptable or not; it is likely that the Lovibond-undiluted technique will meet the needs of most primary health care programmes.

CONCLUSION

Simple techniques for the measurement of haemoglobin levels may be quite satisfactory for the follow-up of individual patients as well as for screening for anaemia. Of the techniques tested, the Lovibond comparator with undiluted blood undoubtedly performed best. The sensitivity, specificity, and predictive values of each test can be calculated for each specific epidemiological situation, thereby providing the information necessary for decision-making concerning the choice of technique and the cut-off points to be used when screening for anaemia.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Mr O. Mwan-Deke for his enthusiastic collaboration, as well as to Professor G. Van Ros who kindly provided the reference measurements.

RÉSUMÉ

LE TAUX D'HÉMOGLOBINE: FIABILITÉ DE QUELQUES TECHNIQUES SIMPLES UTILISABLES DANS LE CONTEXTE DES SOINS DE SANTÉ PRIMAIRES

Cinq techniques simples de dosage de l'hémoglobine ont été testées: le 'Sahli', le 'Tallqvist', le 'Lovibond comparator' avec sang non dilué, du sang dilué dans une solution de Drabkin et du sang dilué dans une solution Drabkin modifiée. De ces cinq techniques, le Lovibond utilisant du sang non dilué, une des techniques les plus maniables, donne de loin les meilleurs résultats en termes de précision et d'exactitude. Cette technique permet de discriminer des différences de taux d'hémoglobine de 20 g/l chez un malade, avec un risque d'erreur inférieur à 5%. L'utilité relative des différentes techniques pour le dépistage de l'anémie dans une population est discutée. Une procédure est proposée qui permet, pour chaque technique, de calculer le nombre de faux positifs et de faux négatifs, ainsi que la sensibilité, la spécificité et la valeur prédictive auxquelles il faudra s'attendre lors d'un dépistage dans une situation épidémiologique donnée et pour un niveau-seuil d'hémoglobine déterminé.

REFERENCES

1. BIGGS, R. & MACMILLAN, R. L. The errors of some haematological methods as they are used in a routine laboratory. *Journal of clinical pathology*, **1**: 269-287 (1948).
2. ELWOOD, P. C. & JACOBS, A. Haemoglobin estimation: a comparison of different techniques. *British medical journal*, **1**: 20-24 (1966).
3. GAMMON, A. & BAKER, S. J. Studies in methods of haemoglobin estimation suitable for use in public health programmes. *Indian journal of medical research*, **65**: 150-156 (1977).
4. VAN BALEN, H. & VAN DEN BULCKE, W. Utilité de la méthode Tallqvist pour le dosage d'hémoglobine au premier échelon du service de santé. *Annales de la Société belge de Médecine tropicale*, **53**: 651-657 (1973).
5. HARRISON, G. A. Estimation of haemoglobin in undiluted blood, using a Lovibond comparator. *Lancet*, **2**: 621-622 (1938).
6. MWANIKUZI, E. & NHONOLI, A. Anaemia in expectant mothers. *East African medical journal*, **49**: 101 (1972).
7. VITERI, F. E. & GUZMÁN, M. A. Haematological status of the Central American population: prevalence of individuals with haemoglobin levels below 'normal'. *British journal of haematology*, **23**: 725-735 (1972).

Annex 1

ESTIMATION OF THE NUMBER OF SCREENING ERRORS

Let the measurement $\hat{m}$ that can be expected when blood with haemoglobin value Hb is measured with a certain technique be given by the regression equation:

$$\hat{m} = a + b \cdot Hb$$

Assuming a normal distribution of the measurements m around their expected measurement $\hat{m}$, the standard error of this distribution is given by:

$$SE = \sqrt{\frac{\Sigma(m - \hat{m})^2}{n-2}}$$

where n = number of measurements.

If one intends to use the haemoglobin level Hb_{co}, and the corresponding $\hat{m}_{co}$ as cut-off point, some blood samples with haemoglobin values less than Hb_{co} will erroneously be considered negative, since some of the corresponding measurements m_i will be larger than $\hat{m}_{co}$. The proportion of false negatives for blood samples with haemoglobin value Hb_i, i.e., the proportion with corresponding $m_i > \hat{m}_{co}$, can be calculated as follows.

The proportion p_i of measurements m_i larger than $\hat{m}_i$ but smaller than $\hat{m}_{co}$ can be read from a Z-table at z_i, where:

$$z_i = \frac{\hat{m}_{co} - \hat{m}_i}{SE} = \frac{b\,(Hb_{co} - Hb_i)}{SE}$$

As 50% of the m_i will be above $\hat{m}_i$, the proportion of false negatives, fn_i, among measurements of blood samples with haemoglobin value Hb_i, is given by:

$$fn_i = 0.5 - p_i$$

A similar reasoning can be made for the false positives (fp_i) among blood samples with haemoglobin values $Hb_i > Hb_{co}$.

Thus, for a given cut-off point, the proportion of false positives and false negatives at each haemoglobin level can be estimated from the regression equation and its standard error.

In order to calculate sensitivity, specificity, and predictive values of a screening test used on a population, these proportions fp_i and fn_i have to be applied to the relative frequency rf_i of each haemoglobin value in the population. The proportions of false positives (FP) and false negatives (FN) that will be found in that population are given by:

$$FP = \sum_i fp_i \cdot rf_i, \quad \text{for } Hb_i > Hb_{co}$$

$$FN = \sum_i fn_i \cdot rf_i, \quad \text{for } Hb_i < Hb_{co}$$

The proportion of true positives (TP) and true negatives (TN) that will be found in that population can easily be calculated from rf_i, FP, and FN.

We can then calculate:

$$\text{the predictive value of a positive test} = \frac{TP}{TP + FP};$$

$$\text{the predictive value of a negative test} = \frac{TN}{TN + FN};$$

$$\text{sensitivity} = \frac{TP}{TP + FN}; \text{ and}$$

$$\text{specificity} = \frac{TN}{TN + FP}.$$

Predictive values, sensitivity, and specificity are thus dependent on the frequency distribution of haemoglobin values in the population and the relation of the cut-off level to this distribution. No absolute values of sensitivity, specificity, and predictive value can be given for any of the techniques: they have to be recalculated in each epidemiological situation and for each cut-off point. This can be done with the help of Table 2.