

Clinical use of WHO haemoglobin colour scale: validation and critique

C F Ingram, S M Lewis

Abstract

Aim—The World Health Organisation (WHO) haemoglobin colour scale has been developed as a simple, inexpensive clinical device for diagnosing anaemia when laboratory based haemoglobinometry is not available. In an initial validation study at several health centres, scale readings were compared with measurements of haemoglobin by the laboratory. This showed the scale to have 90% sensitivity and 70% specificity in identifying whether anaemia was present or not. In addition, when present, the degree of anaemia was correctly classified in clinical terms as moderate, pronounced, or severe, with an overall sensitivity of 60% and specificity of 88%. Errors were mainly marginal—that is, between two adjacent categories—but there were also some major discrepancies, such as a blood with a haemoglobin of 6–7 g/dl being read as normal or vice versa. Because this would compromise the scale's reliability in practice, this study was undertaken to identify the causes of the discrepancies and to reassess the performance of the scale.

Methods—Venous blood samples were collected into potassium EDTA from patients attending selected clinics at three South African hospitals with good laboratory facilities. A prototype of the device was used unsupervised by nursing staff, doctors, and phlebotomists, who were told to follow the printed instructions. The blood specimens were then immediately sent to the laboratory where haemoglobin was measured by standardised automated blood cell counters. Any discrepancies > 1 g/dl were recorded and the tests were repeated by the same operators under supervision of the investigators.

Results—Almost all the errors that occurred resulted from the incorrect use of the device, namely: inadequate or excessive blood, reading the results too soon or too late (beyond the limit of two minutes), poor lighting, or holding the scale at the wrong angle. The accuracy improved dramatically when the tests were repeated under supervision and these faults were avoided: 95% of readings were within 1 g/dl of the reference measurements, and 97% within 1.5 g/dl. Anaemia screening showed 96% sensitivity and 86% specificity. Clinical judgement of pallor was frequently wrong, whereas the scale gave the correct diagnosis in more than 97% of cases.

Conclusion—The study confirmed the usefulness and reliability of the scale and its advantage over clinical signs for the diagnosis of anaemia, thus providing a clinically reliable near patient method in the absence of a laboratory. The instructions are easy to follow but must be strictly adhered to.

(J Clin Pathol 2000;53:933-937)

Keywords: diagnosis of anaemia; haemoglobinometry; haemoglobin colour scale; clinical assessment of anaemia

A simple and inexpensive device for providing a reliable indication of the presence and severity of anaemia would be of considerable value in situations where laboratory based haemoglobinometry is not readily available. The most simple method is to match the colour of a drop of blood on absorbent paper against a colour scale; however, previous devices of this type have been too unreliable to serve their intended purpose.

Stott and Lewis investigated the various factors responsible for the wide margin of error occurring with the available colour scales.¹ This led them, with collaboration by Wynn,² to identify Whatman 31ET chromatography paper as suitable for the preparation of test strips of 45 × 15 mm on which to absorb drops of blood. They established a set of standards consisting of blood samples, the haemoglobin concentration of which was measured by spectrometry with the ICSH³ (International Committee for Standardisation in Haematology; Expert Panel on Haemoglobinometry) reference method, and adjusted to obtain a range of 4–14 g/dl, in 2 g/dl steps. (It was thought that the intended users of the scale would be more familiar with this traditional unit for haemoglobin than g/l). The spectral characteristics of the colour produced by a drop of blood on a test strip from each of the haemoglobin standards were identified by a computerised analytical spectrophotometer. These specifications were reproduced in light resistant printing inks that were prepared from pigments of the three primary colours and a neutral diluent. The colour shades were then printed at a defined ink density on sheets of paper which, to avoid affecting the ink colour, was chemically neutral, unbleached, chlorine free, and resistant to UV light. After the ink was spread the paper was dried and varnished. The colour strips were mounted on a neutral grey surround with a PVC backing, which would allow any traces of blood to be wiped easily off the back of the scale. The viewing area was restricted to a circular aperture of 8–9 mm diameter in the cen-

Helen Joseph Hospital
Laboratory, South
African Institute for
Medical Research, PO
Box 1038,
Johannesburg 2000,
South Africa
C F Ingram

WHO Collaborating
Centre for
Haematology
Technology,
Department of
Haematology,
Hammersmith
Hospital, Imperial
College School of
Medicine, London
W12 0NN, UK
S M Lewis

Correspondence to:
Dr Lewis
smleml@intonet.co.uk

Accepted for publication
9 June 2000

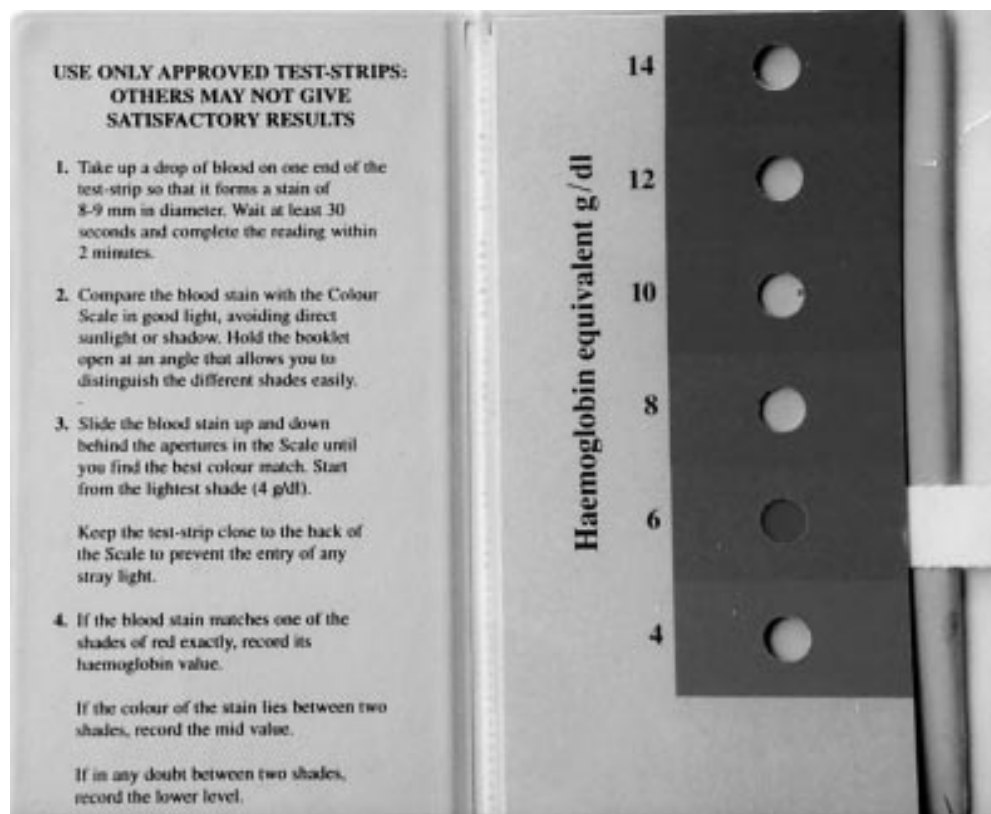


Figure 1 The World Health Organisation haemoglobin colour scale.

tre of each colour standard (fig 1). Optimal conditions were established for colour matching: the best results were obtained with the scale held at an angle of about 45° with the light (daylight or artificial tungsten or fluorescent light) coming over the shoulder of the observer. Although the scale is a clinical device, not intended to be compared with laboratory based measurement, the latter was necessary to provide reference measurements for validating the reliability of the scale. By this means, a laboratory based study on 1213 blood samples showed that with training it was possible to detect anaemia reliably with the scale, and to classify the degree of anaemia correctly in clinical terms as moderate (8 to < 12 g/dl), pronounced (6 to < 8 g/dl), or severe (< 6 g/dl).²

To test the feasibility of using the scale in field conditions a study was undertaken in a rural hospital with a group of nurses, medical students, and untrained assistants. After a 30 minute training session, they were able to identify the degree of anaemia correctly in a batch of blood samples, albeit with some individual discrepancies in readings.⁴ The device was also used with satisfactory results in a WHO training course on malaria control,⁵ peripheral health clinics in the Gambia,⁶ a rural antenatal clinic in Malawi,⁷ and a survey of anaemia in preschool children.⁸

An international study with the prototype device was organised by the WHO Programme on Health Technology to assess its validity when used by health clinic staff who were provided with detailed written instructions but no individually supervised training. Scale readings

were judged against the measurement of haemoglobin on the blood samples by laboratories using haemoglobinometers calibrated with the international haemoglobinocyanide standard. A total of 3600 tests were performed at the clinics and also 2800 tests at blood donor sessions. The results were similar to those in the original evaluation,² with 90% sensitivity and 70% specificity in identifying whether anaemia was present or not. In blood donor screening, 97% were correctly identified at a 12 g/dl discrimination value. When anaemia was present, its severity was assessed as moderate, pronounced, or severe with an overall sensitivity of 60% and specificity of 88% (unpublished report to the WHO). There were also some unexpected mistakes: in some centres there were differences as great as 3–4 g/dl in an otherwise accurate batch of readings in a run, and although these discrepancies occurred in only a small proportion of the total number of tests, they put into question the reliability of the scale. Accordingly, a study was devised to assess the usefulness of the scale in practice at busy health clinics.

Methods

The project was carried out with the prototype scale in general medical and antenatal clinics at three hospitals in the neighbourhood of Johannesburg. The participants included nursing staff (senior nurses, midwives, student nurses), doctors, and phlebotomists. The tests were performed on a total of 548 patients who presented routinely at various outpatient clinics. Using samples from venous blood collected by a phlebotomist, the participants measured their

Table 1 Comparisons of scale and reference measurements

Reference (g/dl)	Colour scale (g/dl)	Initial reading		Repeat reading Incorrect*
		Correct*	Incorrect*	
12 or more	12-14	274	19 (6.5%)	4 (1.4%)
≥11	≥11	331	33 (9.1%)	12 (3.3%)
10 or 11	10 or 11	86	29 (25.2%)	10 (8.7%)
8 or 9	8 or 9	72	14 (12.8%)	9 (4.7%)
6-8	6-8	66	18 (21.4%)	14 (6.7%)
≤6	4-6	35	7 (16.7%)	1 (2.4%)

*Compared with reference measurements.

haemoglobin with the scale independently and without supervision; their readings of haemoglobin values were recorded on a report form ("initial readings"). The specimens and forms were then sent without delay to the hospital laboratory (a branch of the South African Institute for Medical Research), where the haemoglobin was measured on an automated blood cell analyser, which was calibrated by reference to the ICSH international haemoglobin (HiCN) standard.³ These results were entered on the report forms, which were then reviewed by the investigators (CFI and SML). Discrepancies of > 1 g/dl between the scale reading and the haemoglobin measurement were noted and the test was repeated on the original specimens by the same participants (who were not told the laboratory results), but now supervised by the investigators who were vigilant for any possible discrepancies in technique and for any difficulty in distinguishing between the colour shades.

It was impractical to retest the entire set of specimens, but about 10% of the correctly read specimens were included randomly in the batches for the repeated tests. None of these gave different repeat readings; this was consistent with the intra-observer reproducibility (SD, < 1 g/dl) recorded in the original study.²

At some of the clinics the clinicians (doctors or nurses) assessed clinical signs of anaemia, examining conjunctiva, palms, nail bed, and gums for pallor, in accordance with their standard routine practice. These observations were judged against the laboratory haemoglobin measurements on the patients' venous blood samples as reference; a comparison was made with the ability of the scale to identify anaemia correctly.

Table 3 Diagnostic usefulness and predictive values at different haemoglobin (Hb) values using best scale readings where discrepancies occurred (n = 548)

Reference Hb (g/dl)	Scale readings				
	<6	6 to <8	8 to <10	10 to <12	12 or more
<6	34	2	0	0	0
6 to <8	2	27	5	0	0
8 to <10	0	3	87	8	3
10 to <12	0	0	4	74	36
12 or more	0	0	0	9	254
TP	34	27	87	74	254
FP	2	5	9	17	39
FN	2	7	14	40	9
TN	509	508	438	396	245
Sensitivity (%)	94.4	79.4	86.0	64.9	96.6
Specificity (%)	99.6	99.0	98.0	95.9	86.3
PPV (%)	94.4	84.3	90.6	81.3	86.7
NPV (%)	99.6	98.6	96.9	90.8	96.5
LR	0.94	0.78	0.84	0.61	0.83

FN, false negative; FP, false positive; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

Table 2 Discrepancies between scale and reference haemoglobins in individual cases

Variance from reference (g/dl)	Initial reading	Repeat reading
1 or less	340 (65.8%)	489 (94.6%)
1.1-1.4	75 (14.5%)	11 (2.1%)
1.5-1.9	56 (11.2%)	9 (1.7%)
2 (one colour shade)	44 (8.5%)	8 (1.6%)
3 (two colour shades)	15 (2.9%)	0 (0%)

Results

There were two sets of results for the scale: initial readings and repeat readings. The ICSH reference method was used for haemoglobinometry.³ These measurements were recorded in g/dl to one decimal point, but when they were above 14 g/dl they were analysed as 14.0. Tables 1 and 2 compare the scale readings and the reference measurements.

PAIRED *t* TEST OF SCALE *v* REFERENCE MEASUREMENTS

Initial scale readings: 0.3743.

Repeat scale readings: 0.0805.

These results indicate a 70% probability that there is no significant difference between the reference measurements and the initial scale readings, and a 95% probability of no significant differences with the repeat readings. This indicates the occurrence of some random errors in the initial readings, with no specific bias.

DIAGNOSTIC USEFULNESS OF THE SCALE

This was assessed by comparing readings with the reference haemoglobin measurements in four clinically demarcated groups, namely: no anaemia (≥ 12 g/dl), moderate anaemia (8 to < 12 g/dl), pronounced anaemia (6 to < 8 g/dl), and severe anaemia (< 6 g/dl), the repeat readings being used if there had previously been discrepancies (table 3).

Diagnostic sensitivity and positive predictive value are indices of the frequency with which a positive result is correct, whereas diagnostic specificity and negative predictive value indicate the frequency with which a negative result is correct.⁹ The results in table 3 indicate that there was a high degree of accuracy in identifying when anaemia was present and in estimating the degree of anaemia in the clinical cut off points. The likelihood ratio was calculated by Youden's method,¹⁰ where values from +0.1 to +1.0 indicate an increasing probability that the test gives correct diagnostic information; the results show that the ratio was satisfactorily high at all values, especially in anaemia with haemoglobin < 6 g/dl.

COMPARISON OF THE SCALE WITH CLINICAL

DIAGNOSIS OF ANAEMIA

Tables 4 and 5 compare the scale with the clinical ability to diagnose anaemia from physical signs. In severe anaemia (haemoglobin, < 6 g/dl) no judgmental errors were made on physical examination, but with higher haemoglobins the physical features were increasingly difficult to detect and to interpret. Even when haemoglobin was normal, pallor was diagnosed in 16.6% of cases, whereas only 3.4% of the normals were recorded by the scale as having

Table 4 Assessment of clinical signs of anaemia (n = 335)

Ref Hb scores (g/dl)	Conjunctiva			Mucosa			Nail bed			Palms			Total score
	2	1	0	2	1	0	2	1	0	2	1	0	
12 or more	2	24	130	3	23	130	4	9	143	3	9	144	26
10 to <12	5	18	51	1	18	55	2	9	63	1	6	67	23
8 to <10	7	31	25	6	36	21	6	17	40	5	12	46	42
6 to <8	9	5	4	4	10	4	2	12	4	3	8	7	14
<8	10	7	4	5	12	4	3	14	4	4	10	7	17
<6	1	2	0	1	2	0	1	2	0	1	2	0	3

Scores: 2, definitely anaemic; 1, probably anaemic; 0, not anaemic.
Ref Hb, reference haemoglobin.

Table 5 Comparison of clinical signs and colour scale

Reference Hb (g/dl)	Clinical signs present (n/set)*	Anaemia indicated by scale reading (n/set)
12 or more	26/56 (16.6%)	9/261 (3.4%)**
10 to <12	23/74 (31.1%)	78/114 (68.4%)***
8 to <10	42/63 (66.6%)	98/101 (97.0%)
6 to <8	14/18 (77.7%)	34/34 (100%)
<10	59/84 (70.2%)	169/171 (98.8%)
<8	17/21 (80.9%)	70/70 (100%)
<6	3/3 (100%)	36/36 (100%)

*The patients were assessed clinically for anaemia only in some of the clinics (314 cases), whereas the scale assessments of anaemia included all the patients in our study.

**The scale reading indicated only a mild anaemia (10 or 11 g/dl) in the nine cases.

***The scale reading indicated a mild anaemia (10 or 11 g/dl) in 74 cases, moderate anaemia (8 or 9 g/dl) in four cases, and no anaemia (12 g/dl) in the remaining 36 cases.
Hb, haemoglobin.

mild, clinically unimportant anaemia, with haemoglobin readings of 10 or 11 g/dl. On the other hand, when the reference measurements indicated moderate anaemia (haemoglobin, 8–9 g/dl), this was not detected in one third of the patients by physical examination, whereas the scale identified moderate anaemia correctly in 97% of cases.

Discussion

The haemoglobin colour scale was developed as a clinical device to provide a means for identifying whether an individual is anaemic and, in a broad classification, the severity of anaemia. It does not aim to compete with a haemoglobinometer in the laboratory, but it is intended for use when the latter is not available or practical.

Previous studies, including a multicentre evaluation organised by the WHO Programme on Health Technology, indicated its clinical reliability in general medical and antenatal clinics for detecting anaemia and discriminating between moderate, pronounced, and severe anaemia. A study of anaemia in pregnancy in Malawi provided broadly similar results, with a likelihood ratio of correct diagnosis from 1.4 to 3.⁷ However, in all these studies there were some unexpected discrepancies and even some gross mistakes. Thus, for the scale to be depended upon at all times it became necessary to investigate possible causes of such faulty readings.

Our present study was undertaken in medical and antenatal clinics in the neighbourhood of Johannesburg with medical and nursing staff and phlebotomists. Initially, they performed the tests on their own in accordance with the written instructions used in the earlier WHO study, and some of them also received a preliminary brief demonstration. Liaison with

the laboratory ensured that the reference haemoglobin measurements were obtained soon after the scale was read. Where there were discrepancies, the scale test was repeated on the original blood by the original users under expert supervision. In some cases, the participants found it helpful to undertake an exercise with blood samples of known haemoglobin values before repeating the test. This checking resulted in a significant reduction in reading errors (tables 1 and 2), confirming that the scale is fundamentally reliable as long as it is used correctly. Accordingly, it was considered justifiable to use the repeat measurements in assessing the usefulness of the scale. From these results (table 3), it can be seen that the scale is sensitive and specific, with a high degree of discrimination. However, to ensure its reliability in practice it was important to identify the causes of errors and for this information to be incorporated in the instructions to users.

Several factors appear to contribute to incorrect readings. (1) The instructions state that colour matching must be made only after waiting for at least 30 seconds, and that the test must be completed within two minutes because the blood stain changes colour (becoming paler) after this time. In most cases where an error occurred, this was the result of not waiting for the requisite time; subsequently, a wait of one minute has been proposed. Conversely, in some cases, readings were delayed beyond the stipulated two minutes, and it was necessary to prevent users from adding the drops of blood simultaneously to test strips for several consecutive tests. (2) The size of the drop is important. In a small number of cases the initial reading was made on a test strip with too little blood, leading to inadequate spread, with a white periphery at the adjacent matching area. Conversely, too much blood led to a thick spread and insufficient drying in the prescribed time. The solution to this problem was to take up blood to a distance of about 3 cm in a capillary tube for delivery on to the test strip. (3) It is essential to have good light. This can be either daylight (but not direct sunlight) or artificial light, or a mixture of both. The study was carried out in midsummer, but outside light varied from bright sunshine to heavy rain clouds. The tests were performed in clinics with daylight coming through windows of different sizes, supplemented by fluorescent light. Scale readings were not affected by any of these variable conditions except on one occasion when the tests were performed in a poorly lit basement examination room—30% of the results were discrepant—but were corrected when the tests were repeated by the same operators after adjourning to a better lit room. Problems also occurred when the users stood in their own shadows or did not hold the scale at the recommended angle of vision. Similarly, there were occasional errors from failure to ensure close apposition of the test strip to the apertures on the scale, because this cast a shadow on the colour strip. (4) Some training is required, albeit minimal: careful understanding of the instructions and a few minutes of

practice with some blood samples of known haemoglobin content is all that is necessary.

Clerical errors with incorrect entering of results and specimen mix up are recognised as major causes of error in any laboratory, and this occurred on two occasions in our study during the stress of busy clinics; this problem is not unique to scale readings, and should not be attributed to scale failure.

The relative frequency of the different faults does not imply that some are more important than others, but only indicates how individuals performed. It demonstrates the importance of adhering to the instructions, which take account of all the causes of error listed above.

Because the scale is intended to be a clinical facility, the study included a comparison with the clinical ability to diagnose anaemia from physical signs, based on the presence of pallor. With severe anaemia the clinical features were clearly positive, but with moderate anaemia pallor was difficult to detect: when the reference haemoglobin was 8–9 g/dl, anaemia was not detected in one third of the patients by physical examination, whereas the scale identified moderate anaemia correctly in 97% of cases. Conversely, pallor was recorded (misleadingly) in 16.6% of cases when haemoglobin was normal, whereas 96.6% were normal by the scale reading, and only 3.4% were read as mild anaemia with haemoglobin readings of 10–11 g/dl.

Without exception the scale was well received, and the clinic staff indicated their wish to use it as soon as possible in their routine practice. Most participants found the scale and the test strips easy to handle, and the test easy to perform and user friendly. Some found it difficult to distinguish between 10 and 12 g/dl. Consequently, these adjacent shades have been adjusted in the production version of the scale to ensure that discrimination will be easier. Confidence in selecting the correct shade increased with experience and most users had no difficulty in judging intermediate values.

Conclusion

Our study confirmed the usefulness and reliability of the scale and its great advantage over clinical signs for diagnosing anaemia, and

providing a reliable near patient method for screening anaemia in the absence of a laboratory. Most health workers should be able to obtain reliable readings of haemoglobin, accurate within 1 g/dl, as long as the correct procedure is strictly adhered to. It is useful, but not essential, to have a brief introductory training with a batch of blood samples of known haemoglobin values, and even trained users might find it helpful to have a brief refresher training course with such bloods from time to time. It is essential to accompany the scale and test strips with a leaflet containing clearly written instructions for use, emphasising the potential causes of error. This study has been valuable in identifying these errors, and thus provides guidance for both the specifications of the scale and the correct technique for its use.

We thank Professor J Hofmeyr, Head of Obstetrics at Coronation Hospital; Professor Van Gelderen, Head of Obstetrics at Chris Hani Baragwanath Hospital; and Professor MacPhail, Head of Medicine at Helen Joseph Hospital for providing the facilities to carry out this study. We are grateful to Dr J Hall, Consultant Obstetrician at Chris Hani Hospital for her support, and the nursing staff, phlebotomists, clinicians, and laboratory technologists at the three centres who made the study possible by their collaboration and enthusiastic support, despite having to cope simultaneously with their heavy routine workload.

- 1 Stott GJ, Lewis SM. A simple and reliable method for estimating haemoglobin. *Bull World Health Organ* 1995;73:369–73.
- 2 Lewis SM, Stott GJ, Wynn KL. An inexpensive and reliable new haemoglobin colour scale for assessing anaemia. *J Clin Pathol* 1998;51:21–4.
- 3 International Committee for Standardisation in Haematology: Expert Panel on Haemoglobinometry. World Health Organisation. Recommendations for reference methods for haemoglobinometry in human blood (ICSH standard 1986) and specifications for international haemoglobinocyanide reference preparation (3rd edition). *Clin Lab Haematol* 1987;9:73–9.
- 4 Munster M, Lewis SM, Erasmus LK, *et al.* Field evaluation of a novel haemoglobin measuring device designed for use in a rural setting. *S Afr Med J* 1997;87:1522–6.
- 5 Beales PF. Anaemia in malaria control: a practical approach. *Ann Trop Med Parasitol* 1997;91:713–18.
- 6 Gosling R, Walraven G, Manneh F, *et al.* Problems in training health workers to use the WHO haemoglobin colour scale. *Trop Med Int Health* 2000;5:214–21.
- 7 Van den Broek NR, Ntonya C, Mhango E, *et al.* Diagnosing anaemia in pregnancy in rural clinics: assessing the potential of the haemoglobin colour scale. *Bull World Health Organ* 1999;77:15–21.
- 8 Montresor A, Albionico M, Khalfan N, *et al.* Field trial of a haemoglobin colour scale: an effective tool to detect anaemia in preschool-children. *Trop Med Int Health* 2000;5:129–30.
- 9 Galen RS, Gambino S. *Beyond normality: the predictive value and efficiency of medical diagnoses*. New York: Wiley, 1975.
- 10 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.