Accuracy of the WHO Haemoglobin Colour Scale for the diagnosis of anaemia in primary health-care settings in low-income countries: a systematic review and meta-analysis

Heiko Marn, Julia Alison Critchley

Summary

Background Anaemia is a major cause of morbidity and mortality in low-income countries. Primary health-care workers in resource-poor settings usually diagnose anaemia clinically, but this is inaccurate. The WHO Haemoglobin Colour Scale (HCS) is a simple, cheap quantitative method to assess haemoglobin concentration outside of the laboratory. We systematically reviewed the literature to assess the accuracy of the HCS in primary care to diagnose anaemia, and compared this with clinical assessment.

Methods We searched the electronic databases including MEDLINE, EMBASE, SCOPUS, Web of Science, Cochrane library, CINAHL plus, Popline, Reproductive Health Library, and Google Scholar and regional databases up to Nov 14, 2014, "haemoglobin colour scale" in alternative spellings published in any language. Two reviewers independently screened studies, extracted data, and assessed quality using the QUADAS-2 instrument. Statistical analyses were carried out in STATA using the bivariate model.

Findings Of 141 records and abstracts screened, 14 studies were included. The pooled sensitivity of the HCS to diagnose anaemia was 80% (95% CI 68–88) compared with 52% for clinical assessment ([95% CI 36–67; p=0·008). Specificity was similar between the HCS (80% [95% CI 59–91]) and clinical assessment (75% [56–88; p=0·8250). For severe anaemia, diagnostic accuracy was again higher overall for the HCS (p<0·0001); sensitivity was 57% (36–76) for the HCS and 45% (95% CI 12–83) for clinical assessment, but specificity was 99·6% (95% CI 95–99·9) versus 92% (62–99). Combining clinical assessment and the HCS could result in higher sensitivity (anaemia: 91% [95% CI 81–96]); severe anaemia 83% (33–98), but at the expense of specificity (anaemia: 59% [35–79]; severe anaemia 90% [40–99]). Individual studies were highly heterogeneous but pooled results did not differ substantially in a series of sensitivity analyses for indicators of study robustness.

Interpretation In so-called real-life primary health-care conditions, HCS can significantly reduce misdiagnosis of anaemia compared with clinical assessment alone. Future research is required to optimise training, and assess clinical outcomes and cost-effectiveness.

Funding None.

Institute of Tropical Medicine and International Health, Charité—Universitätsmedizin Berlin, Berlin, Germany (H Marn MD)**; and Population Health Research Institute, St George's, University of London, London, UK** (Prof J A Critchley DPhil)

Correspondence to: Heiko Marn, Brüderstrasse 13, 20355 Hamburg, Germany heiko_marn@hotmail.com

Introduction

Anaemia is a major global cause of maternal, perinatal, and child mortality. Additionally, it causes low birthweight, impaired or delayed child physical and mental development, and an increased susceptibility to infections,[1] and contributes greatly to economic loss due to reduced productivity of workers.[2] About 1·62 billion people are affected.[1] Most are non-pregnant women (468·4 million), preschool age children (293·1 million), and pregnant women (56·4 million) predominantly in low-income countries, where prevalence rates are up to five times higher than in high-income countries and are inversely correlated with economic status.[3,4]

In these low-income societies, iron deficiency anaemia is believed to account for about 50% of all cases of anaemia,[5] but other causes are frequent and often co-exist, including malnutrition, micronutrient deficiencies, parasitic infections, other chronic inflammatory conditions, or hereditary haemoglobinopathies.[3]

Accurate quantitative point-of-care diagnostic tests are able to confirm the diagnosis of anaemia through measurement of a decreased amount of red blood cells or decreased haemoglobin concentration in the blood,[6] but these are not suitable in most primary health-care settings with very low resources, because they either require constant quality control by trained staff, use toxic or expensive reagents and consumables, or depend on an electricity supply.[7]

Diagnosis is thus often based on clinical signs alone such as conjunctival, palmar, and nailbed pallor. None of these signs, whether combined or singly, yield an acceptable diagnostic accuracy.[8] This leaves many cases undetected and untreated and also poses the risk of unnecessary and potentially harmful blood transfusions, increasing the risk of transmission of blood-borne

Research in context

Evidence before the study

The WHO Haemoglobin Colour Scale (HCS) became commercially available in 2001 as an instrument for health-care workers in resource-poor settings, who usually have to base the diagnosis of anaemia on signs and symptoms, to quantatively assess the anaemia status of their patients. The first and only systematic review to date to assess the diagnostic accuracy of the HCS was published in 2005, which included 14 studies, but most of these were laboratory-based with only four taking place in primary care in low-income settings, under which the HCS is supposed to be used in practice. The reported estimates of diagnostic accuracy from this 2005 review were very heterogeneous (sensitivity 75–97% and specificity 41–98% for the detection of anaemia), and were less accurate in the four field studies (sensitivity 76–88%; specificity 41–100%). The authors did not compute summary estimates from individual studies, except for the five laboratory studies.

Added value of this study

We restricted our systematic review to real life studies (n=14), identifying ten more than available at the time of the previous review. We were also able to compare the performance of the HCS directly against the diagnosis of anaemia by clinical signs, because most studies directly compared these two tests. This is important because clinical assessment is the standard procedure to diagnose anaemia in most primary health-care settings in low-income countries. We also estimated diagnostic accuracy for simultaneous testing (HCS and clinical signs). Despite heterogeneous outcomes, we undertook meta-analysis of individual studies using the bivariate random effects model, and we used an evidence informed tool (QUADAS 2) for the assessment of the methodological quality of studies, allowing a series of sensitivity analyses.

Implications of all the available evidence

There is sound evidence that the HCS can improve the accuracy of diagnosis of anaemia and severe anaemia by primary health-care workers under resource-poor conditions. This finding is consistent in a variety of sensitivity analyses accounting for study quality and threshold effects. The HCS is significantly more sensitive for the diagnosis of anaemia than assessment of clinical signs, and the improvement in sensitivity could be clinically important in practice. Evidence concerning how training and supervision might affect the overall performance of the device, as well as its cost-effectiveness in reducing anaemia-related mortality and morbidity in practice, is lacking.

pathogens, and wasting resources in case of mis-diagnosed severe anaemia.

In response to the need for a "simple, cheap, and robust device to measure haemoglobin by health workers outside the laboratory"[9,10] the WHO Haemoglobin Colour Scale (HCS) was developed and has been produced and distributed under licence agreement by Copack (Oststeinbek, Germany) since 2001.[10–12] The scale comprises a small card of six shades of red (lighter to darker), each representing a haemoglobin concentration of 40 g/L, 60 g/L, 80 g/L, 100 g/L, 120 g/L, and 140 g/L, respectively. A drop of blood absorbed onto a standardised chromatography filter paper is compared with the colour scale, allowing assessment of the patient's haemoglobin concentration, including an estimation of intermediate results, in 10 g/L steps.[13]

The usefulness of the device in practice has been disputed,[14,15] but in 2005 a systematic review of 14 studies showed that, under ideal conditions, the HCS might improve diagnosis of mild and moderate anaemia with reasonable accuracy (sensitivities from 85% to 99% and specificities from 91% to 100% in five laboratory-based studies).[16] Ideal conditions are defined as studies taking place in a laboratory setting, including trained laboratory staff operating or supervising the HCS measurements after intensive training, from blood samples of hospital populations or blood donors. The diagnostic accuracy tended to be lower in the four so-called real-life studies (sensitivities 76–88%, apart from one outlier, and specificities from 41% to 100%), leading to the conclusion

that further research was needed to assess the usefulness of the HCS in real-life situations. Real life conditions are defined as studies that were carried out in patient populations attending routine primary health clinics or public schools, with the HCS undertaken by primary health-care workers or a person with comparable skills or training. Only a minority (5 of 14) compared the accuracy of HCS with clinical diagnosis. We are aware of no systematic reviews of the performance of HCS since 2005, although additional "real life" studies have been published.

We aimed to do an updated systematic review to assess the accuracy of the HCS to diagnose anaemia and severe anaemia in resource-poor primary health-care settings compared with the accuracy of diagnosis by clinical assessment, wherever such data are available.

Methods

Search strategy and selection criteria

Following PRISMA guidelines, we searched the electronic databases MEDLINE, EMBASE, SCOPUS, Web of Science, Cochrane library, CINAHL plus, Popline, Reproductive Health Library, TRIP Database, ADOLEC, BDENF, DESASTRES, HISA, MedCarib, LILACS, IMEMR, IMSEAR, WPRIM, and Google Scholar, all from inception up to Nov 14, 2014. To increase sensitivity of the search strategy,[17] we searched only the keywords "haemoglobin colour scale" without any filters using alternative spellings in English, Spanish, and French. A citation search on "Critchley and Bates 2005 systematic

review[16]" was done in Medline+Embase (Ovid), Scopus, Web of Science, Cinahl plus, and Google scholar. Both authors independently screened the titles and abstracts of all records retrieved and checked the reference lists of eligible articles for further studies; any disagreements were resolved by discussion (appendix p 1).

We included all studies comparing the diagnostic accuracy of the HCS with any reference method (gold standard) to diagnose anaemia under real life conditions as defined before. There were no restrictions based on sample size, location, background morbidities, or anaemia prevalence. Studies done in hospitals, laboratories, or blood banks were excluded because they are not generalisable towards primary health-care in low-resource settings.

Data extraction and quality assessment
Both authors independently extracted data including the main study outcomes, study characteristics, and quality-related information based on WHO recommendations for HCS evaluations (appendix p 2).[12]

We assigned tailored quality-relevant criteria to the domains "patient selection", "index test", "reference standard test", and "flow and timing", as proposed in the QUADAS-2 instrument[18] and applied customised signalling questions (table 1) to each individual study to judge whether the risk of bias and applicability concerns to our review objectives were either "high" or "low". The rating "unclear" was only used when the publication did not report quality-relevant data, when the inter-rater reliability was not assessed, or if only one operator did all HCS readings. Again, both authors independently extracted data for all these aspects of quality using a standardised form. Any disagreements were resolved by discussion between authors.

Statistical analyses
Both authors independently extracted the study outcomes for true positive, true negative, false positive, and false negative test results into 2×2 tables. The haemoglobin cut-off level in children aged 6–59 months and during pregnancy for diagnosing anaemia was 110 g/L and for diagnosing severe anaemia was 70 g/L according to WHO recommendation.[19] Studies with a different threshold for anaemia and severe anaemia were included in the meta-analysis, but excluded in a sensitivity analysis. We assessed heterogeneity between studies through creation of forest plots and summary ROC curves.

Overall summary estimates
We used the bivariate random effects model to combine data across all included studies. This model analyses pairs of sensitivity and specificity estimates jointly, accounting for possible correlation between both measures within (using a random effects model) and between studies (assuming normal distribution), hence preserving the two-dimensional nature of the original data.[20] We pooled data

	Risk of bias	Applicability concerns
Patient selection	Was a consecutive or randomised sample of cases enrolled Did the study avoid inappropriate exclusions	Did included patients match the target population
Index test	Was the WHO certified HCS kit used Were the HCS results interpreted without the knowledge of the reference test results Were the results of HCS readings reliable across different raters	Did the HCS operator match the review's "real life" objective Was the training appropriate for resource-poor situations (at least 1 h, at most 1 day) Was the cut-off for anaemia according to WHO recommendations (haemoglobin <110 g/L)
Reference test	Was the reference test likely to correctly diagnose anaemia	Did the reference test allow the assessment of the HCS accuracy
Flow and timing	Was the sampling of HCS and reference test concurrent	..

HCS= Haemoglobin colour scale.

Table 1: Signalling questions for risk of bias and applicability judgement (QUADAS-2) by domain

for the HCS and clinical assessment separately. In a series of sensitivity analyses, we excluded different subsets of studies to explore whether the exclusion of studies with high risk of bias, studies that did not adjust for several readings of HCS results from the same patient, and studies using different cut-offs for anaemia and severe anaemia would affect the pooled accuracy estimates. We then also repeated analyses restricted to the studies that compared the HCS directly with clinical diagnosis, to assess whether confounding by study was affecting comparisons.

See **Online** for appendix

Comparison of the diagnostic accuracy
We compared the diagnostic accuracy of the HCS with clinical diagnosis in a meta-regression analysis (adding test as a covariate), allowing for covariance both between and within these two "tests". Again, we used a bivariate random effects model. We accounted for the correlation expected when two different tests take place in the same study population, and also tested whether the variances of the random effects differed between tests. For severe anaemia, this full model did not converge due to the smaller number of studies. We thus entered the type of "test" as a covariate with random effects; an approach that has been shown to produce similar results,[21] but with the limitation that we can only test for overall differences in diagnostic accuracy rather than specifying whether it is the expected sensitivities or specificities that differ. We undertook these models in all studies initially and then only in those studies that examined the performance of both methods. This also allowed us to estimate a pooled accuracy for simultaneous testing, which we assumed to be routine practice.

Meta-regression analysis
Using the same bivariate random effects model, we undertook meta-regression analysis with the addition of covariates in sequence to assess whether the following variables could explain any of the heterogeneity between studies: (1) level of training (greater or less than half a

	Population	Sample size	Study setting	Study design	Reference standard test (blood sample)	Operators	Training	Cut-off anaemia/severe anaemia (g/L)	Prevalence anaemia/severe anaemia	HCS sensitivity/specificity	Clinical signs sensitivity/specificity
van den Broek (1999)[36]	Pregnant women	1066 observations from 643 samples	5 rural antenatal clinics (3 rural hospitals and 2 health centres) in Malawi	Prospective diagnostic accuracy study comparing HCS, HemoCue and conjunctival colour	Electronic coulter counter (venous)	44 nurse-midwives from 5 different sites	1 day	Anaemia: <110 Severe anaemia: <60	Anaemia: 0.58 Severe anaemia: 0.006	Anaemia, sensitivity: 0.78 (0.74–0.81), Anaemia, specificity: 0.50 (0.46–0.55), Severe anaemia, sensitivity: 0.50 (0.12–0.88), Severe anaemia, specificity: 0.98 (0.98–0.99)	Anaemia, sensitivity: 0.33 (0.29–0.38), Anaemia, specificity: 0.84 (0.79–0.88), Severe anaemia, sensitivity: 0.67 (0.09–0.99), Severe anaemia, specificity: 0.74 (0.70–0.77)
Montresor (2000)[30]	Pre-school children (16–83 months)	535	Mother and child health clinics, presumably rural area of Zanzibar (Tanzania)	Prospective diagnostic accuracy study comparing HCS and clinical pallor signs in children recruited for deworming and iron supplementation intervention study	HemoCue (venous)	6 members of Helminth Control Programme (2 "highly skilled laboratory technicians" did 95% of the readings)	1 day on an average of 15 blood samples	Anaemia: <110 Severe anaemia: <70	Anaemia: 0.79 Severe anaemia: 0.04	Anaemia, sensitivity: 0.85 (0.81–0.88), Anaemia, specificity: 0.75 (0.66–0.83), Severe anaemia, sensitivity: 0.74 (0.49–0.91), Severe anaemia, specificity: 1.00 (0.99–1.00)	Anaemia, sensitivity: 0.22 (0.18–0.26), Anaemia, specificity: 0.92 (0.86–0.96), Severe anaemia, sensitivity: 0.63 (0.38–0.84), Severe anaemia, specificity: 0.84 (0.80–0.87)
Barduagni (2003)[26]	School-children, mean age 8·4 years (range 6–11 years)	150	Qena Governorate, upper Egypt	Prospective diagnostic accuracy study comparing HCS and Sahli's haemoglobin-meter	HemoCue (capillary)	1 nurse	Unclear	Anaemia: <120 Severe anaemia: <70	Anaemia: 0.17 No cases	Anaemia, sensitivity: 0.88 (0.70–0.98), Anaemia, specificity: 0.49 (0.40–0.58)	NA
Montresor (2003)[31]	Pregnant women	1529	8 dispensaries on Pemba Island, Zanzibar (Tanzania)	Prospective diagnostic accuracy study, 2-part hospital/field study, (only field data used for this review)	HemoCue (capillary)	13 HCW at dispensaries (HCS); different 8 HCW (pallor signs)	2 days	Anaemia: <110 Severe anaemia: <70	Anaemia: 0.83 Severe anaemia: 0.2	Anaemia, sensitivity: 0.95 (0.94–0.96), Anaemia, specificity: 0.14 (0.10–0.19), Severe anaemia, sensitivity: 0.82 (0.78–0.87), Severe anaemia, specificity: 0.86 (0.84–0.88)	Anaemia, sensitivity: 0.41 (0.39–0.44), Anaemia, specificity: 0.76 (0.71–0.81), Severe anaemia, sensitivity: 0.78 (0.72–0.82), Severe anaemia, specificity: 0.65 (0.62–0.68)
Gies (2003)[28]	Pregnant women	403	Urban health centre in Assawa, southern Ethiopia, Rift valley, 1700 m altitude	Prospective diagnostic accuracy study comparing HCS and conjunctival pallor and developing risk score based on symptoms and complaints	HemoCue (capillary)	4 midwives, 1 principal investigator	2 afternoon sessions	Anaemia: <110 Severe anaemia: <70	Anaemia: 0.15 Severe anaemia: 0.003	Anaemia, sensitivity: 0.44 (0.31–0.57), Anaemia, specificity: 0.87 (0.83–0.90), No data	Anaemia, sensitivity: 0.44 (0.32–0.58), Anaemia, specificity: 0.79 (0.74–0.83), No data

(Table 2 continues on next page)

	Population	Sample size	Study setting	Study design	Reference standard test (blood sample)	Operators	Training	Cut-off anaemia/severe anaemia (g/L)	Prevalence anaemia/severe anaemia	HCS sensitivity/specificity	Clinical signs sensitivity/specificity
(Continued from previous page)											
Lindblade (2006) p[29]	Pregnant women	643	Rural communities in Gem, Nyanza Province, Kenya	Prospective diagnostic accuracy study, 2-part hospital/field, mixed population: children and pregnant women (assessed separately in this review) comparing HCS and clinical pallor signs	HemoCue (capillary)	6 CHW (limited formal training in traditional birth attending and community health)	4·5 h (1·5 h explaining the study and 3 h practicing the HCS on 5 specimen with known Hb level)	Anaemia: <110 / Severe anaemia: <70	Anaemia: 0·52 / Severe anaemia: 0·025	Anaemia, sensitivity: 0·60 (0·55–0·66); Anaemia, specificity: 0·94 (0·90–0·96); Severe anaemia, sensitivity: 0·44 (0·20–0·70); Severe anaemia, specificity: 1·00 (0·99–1·00)	Anaemia, sensitivity: 0·67 (0·61–0·72); Anaemia, specificity: 0·55 (0·49–0·60); Severe anaemia, sensitivity: 1·00 (0·79–1·00); Severe anaemia, specificity: 0·45 (0·41–0·49)
Lindblade (2006) c[29]	Children (2–24 m)	438	Rural communities in Gem, Nyanza Province, Kenya	Prospective diagnostic accuracy study, 2-part hospital/field, mixed population: Children and pregnant women (assessed separately in this review) comparing HCS and clinical pallor signs	HemoCue (capillary)	6 CHW (limited formal training in traditional birth attending and community health)	4·5 h (1·5 h explaining the study and 3 h practicing the HCS on 5 specimen with known Hb level)	Anaemia: <110 / Severe anaemia: <70	Anaemia: 0·74 / Severe anaemia: 0·10	Anaemia, sensitivity: 0·79 (0·75–0·84); Anaemia, specificity: 0·85 (0·77–0·91); Severe anaemia, sensitivity: 0·63 (0·47–0·77); Severe anaemia, specificity: 0·97 (0·95–0·99)	Anaemia, sensitivity: 0·64 (0·59–0·69); Anaemia, specificity: 0·59 (0·49–0·68); Severe anaemia, sensitivity: 0·88 (0·75–0·96); Severe anaemia, specificity: 0·45 (0·40–0·50)
van Rheenen (2007)[37]	Newborn (at birth, follow-up at 2 months and 4 months)	250	Mpongwe rural district Mission Hospital (at birth) and Mpongwe mother and child clinic, Copperbelt district, Zambia	Prospective diagnostic accuracy study	HemoCue (umbilical cord and capillary)	1 investigator (author)	Unclear	Anaemia at birth: <125 g/L; at 2 months: <95 g/L; at 4 months <104 g/L / Severe anaemia: no data	Anaemia: 0·12 / Severe anaemia: no data	Anaemia, sensitivity: 0·40 (0·23–0·59)); Anaemia, specificity: 0·96 (0·92–0·98); Severe anaemia: no data	Anaemia: not assessed; Severe anaemia: not assessed
Sinha (2008)[35]	Children (6–35 months)	772	67 villages of 3 health centres (Anji, Gaul, Talegaon; total population: 88187), Wardha district, Central India	Prospective diagnostic accuracy study comparing HCS and palmar pallor embedded into a larger morbidity survey	Filter paper cyanomet-haemoglobin method (FPCM; capillary)	Investigator (not clear)	Unclear	Anaemia <110 / Severe anaemia: <70	Anaemia: 0·80 / Severe anaemia: 0·013	Anaemia, sensitivity: 0·90 (0·87–0·92); Anaemia, specificity: 0·97 (0·93–0·99); Severe anaemia, sensitivity: 0·00 (0·00–0·31); Severe anaemia, specificity: 1·00 (1·00–1·00)	Anaemia, sensitivity: 0·67 (0·63–0·70); Anaemia, specificity: 0·98 (0·94–1·00); Severe anaemia, sensitivity: 0·00 (0·00–0·31); Severe anaemia, specificity: 1·00 (1·00–1·00)

(Table 2 continues on next page)

	Population	Sample size	Study setting	Study design	Reference standard test (blood sample)	Operators	Training	Cut-off anaemia/ severe anaemia (g/L)	Prevalence anaemia/ severe anaemia	HCS sensitivity/ specificity	Clinical signs sensitivity/ specificity
(Continued from previous page)											
Rusnawatinigtyas (2009)[33]	Elementary school children, mean age 9 years	124	Elementary school in Karangrejo, Jogjakarta, Indonesia	Prospective diagnostic accuracy study	Hematology analyzer (HmX) (venous)	1 paediatric resident, 1 paramedic; blood samples taken by a trained paramedic	Intensive training included initial pilot study	Anaemia 115; Severe anaemia: no data	Anaemia: 0.12; Severe anaemia: no data	Anaemia, sensitivity: 0.93 (0.68–1.00); Anaemia, specificity: 1.00 (0.97–1.00); Severe anaemia: no data	Anaemia: not assessed; Severe anaemia: not assessed
Bala (2011)[35]	Pregnant women	129	Randomly selected urban health centres in Ahmedabad, Gujarat State, Western India	Prospective diagnostic accuracy study comparing HCS and clinical pallor signs	Sahli's hemometer (unclear)	trained multi-purpose health worker or health visitor	No training reported, but HCS introduced before	Anaemia <110; Severe anaemia: <70	Anaemia: 0.70; Severe anaemia: 0.016	Anaemia, sensitivity: 0.83 (0.74–0.90); Anaemia, specificity: 0.33 (0.19–0.50); Severe anaemia, sensitivity: 1.00 (0.16–1.00); Severe anaemia, specificity: 0.98 (0.94–1.00)	Anaemia, sensitivity: 0.91 (0.83–0.96); Anaemia, specificity: 0.13 (0.04–0.27); Severe anaemia, sensitivity: 1.00 (0.16–1.00); Severe anaemia, specificity: 0.98 (0.93–1.00)
Prathapan (2011)[32]	Pregnant women	101	Field ante-natal clinics in 11 out of 13 MOH areas in the Colombo district, Sri Lanka	Prospective diagnostic accuracy study as secondary objective; part of quality survey of primary health care services	Spectrometry method at "quality assured laboratory" (venous)	Medical officers at antenatal field clinic	No training reported, but HCS introduced before	Anaemia <110; Severe anaemia: no data	Anaemia: 0.21; Severe anaemia: no data	Anaemia, sensitivity: 0.62 (0.38–0.82); Anaemia, specificity: 0.86 (0.77–0.93); Severe anaemia: no data	Anaemia: not assessed; Severe anaemia: not assessed
Chathurani (2012)[27]	Pregnant women	115	MOH field clinics Anuradhapura district, Sri Lanka	Cross sectional health survey; retrospective diagnostic accuracy study of HCS as secondary objective, comparing historical HCS values with current ref. test; current pallor signs compared to ref. stand.	Cyanmet-haemoglobin method in reference laboratory (venous)	PHM or public health nursing sisters	Unclear	Anaemia <110; Severe anaemia: <70	Anaemia: 0.16; Severe anaemia: no cases	Anaemia, sensitivity: 0.50 (0.26–0.74); Anaemia, specificity: 0.76 (0.67–0.84); Severe anaemia: no cases	Anaemia, sensitivity: 0.19 (0.11–0.29); Anaemia, specificity: 0.88 (0.85–0.91); Severe anaemia: no cases
Aldridge (2012)[24]	Pre-school children (2–59 m)	1050 observations from 799 samples	Primary health care services (6 mother and child health clinics) Pemba island of Zanzibar archipelago, Tanzania	Prospective diagnostic accuracy study comparing HCS and clinical pallor signs	HemoCue (capillary)	9 HCW (3 nurses, 1 nurse prescriber, 1 midwife, 2 public health nurses, 1 laboratory technician, 1 psychiatric nurse)	1 h	Anaemia <110; Severe anaemia: <50	Anaemia: 0.71; Severe anaemia: 0.0067	Anaemia, sensitivity: 0.33 (0.294–0.36); Anaemia, specificity: 0.87 (0.83–0.91); Severe anaemia, sensitivity: 0.14 (0.00–0.58); Severe anaemia, specificity: 1.00 (0.99–1.00)	Anaemia, sensitivity: 0.58 (0.54–0.62); Anaemia, specificity: 0.55 (0.48–0.61); Severe anaemia, sensitivity: 0.00 (0.00–0.46); Severe anaemia, specificity: 1.00 (0.99–1.00)

(Table 2 continues on next page)

	Population	Sample size	Study setting	Study design	Reference standard test (blood sample)	Operators	Training	Cut-off anaemia/severe anaemia (g/L)	Prevalence anaemia/severe anaemia	HCS sensitivity/specificity	Clinical signs sensitivity/specificity
(Continued from previous page)											
Shah (2014)[34]	Women of reproductive age (15–45 years)	501	8 villages of Jhagadia block located in Gujarat, India	Prospective diagnostic accuracy study	HemoCue (capillary)	Village-based CHW (mean age 31 y, at least primary education) mean duration of experience 6·5 y	0·5 day	Anaemia <120; Severe anaemia: <70	Anaemia: 0·71; Severe anaemia: 0·024	Anaemia, sensitivity: 0·96 (0·4–0·98) Anaemia, specificity: 0·22 (0·15–0·29) Severe anaemia, sensitivity: 0·83 (0·52–0·98) Severe anaemia, specificity: 0·99 (0·98–1·00)	Anaemia: not assessed Severe anaemia: not assessed

A=anaemia. SA=severe anaemia. CHW=community health worker. HCS=Haemoglobin colour scale. HCW=health-care workers. MOH=ministry of health. PHM=public health midwife.

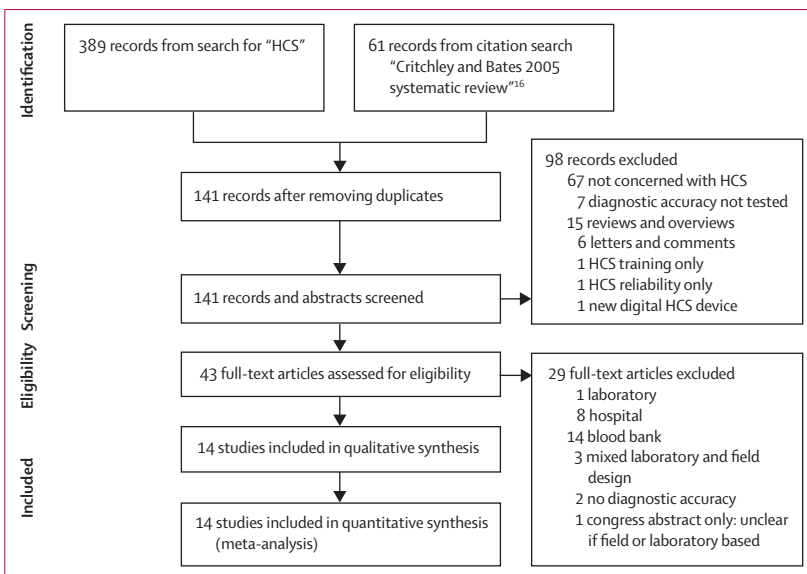*Table 2:* Study characteristics and main outcomes

Figure 1: Study selection
HCS=Haemoglobin colour scale.

day); (2) type of reference test (standard laboratory test or point-of-care test); (3) whether both the HCS and reference test used the same type of blood sample (ie, both used capillary blood or both used venous blood) or a different sample; (4) the population type (women or children); (5) anaemia prevalence (40% or higher compared with less than 40%). In this meta-regression, we assumed that training levels were "low" for the four studies that did not report this and that the type of blood sample was different for the three studies that did not state this clearly.

Data were analysed with Review manager version 5.3 and STATA 12 statistical software packages metandi, gllamm, and xtmelogit for meta-analysis and meta-regression modelling (appendix p 3).[20,22]

Role of the funding source

There was no external funding for this study. The funding institution of JC had no role in the design and development, data extraction, analysis and interpretation of the data, or preparation, review, or approval of the paper. HM had full access to all data. HM and JC both had the decision to submit for publication.

Results

Of 141 records screened for eligibility based on titles and abstracts, 98 papers were excluded based on titles and abstracts, and 43 full-text articles were assessed for eligibility (figure 1; appendix p 4). 29 articles were excluded because they did not meet the previously defined real-life inclusion criteria: 14 were undertaken in blood banks, eight in hospitals, one in a laboratory, three had a mixed field or laboratory design, and two did not report diagnostic accuracy data. For one congress abstract[23] information about whether it was field or laboratory

	Risk of bias							Applicability concerns				
	Patient selection		Index test			Reference test*	Flow and timing†	Patient selection‡	Index test			Reference test§
	Randomisation or consecutive cases	No inappropriate exclusions	WHO certified HCS¶	Blinding HCS vs reference test	Reliability of HCS readings				HCS operator matches review objective	Training intensity at least 1 h, at most 1 day	WHO according cut-off (110 g/L)	
van den Broek (1999)	Unclear	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Montresor (2000)	Unclear	Low	Low	Low	Low	Low	Low	Low	High	Low	Low	Low
Barduagni (2003)	Low	Low	Low	Low	Unclear	Low	Low	Low	Low	Unclear	High	Low
Montresor (2003)	Unclear	Low	Low	Low	Low	Low	Low	Low	Low	High	Low	Low
Gies (2003)	Low	Low	Low	Low	Low	Low	Low	Low	Low	High	Low	Low
Lindblade (2006c)	Low	Low	Low	Unclear	Low	Low	Low	Low	Low	Low	Low	Low
Lindblade (2006p)	Low	Low	Low	Unclear	Low	Low	Low	Low	Low	Low	Low	Low
van Rheenen (2007)	Low	High	Low	Low	Unclear	Low	Low	High	High	Unclear	High	Low
Sinha (2008)	Low	Low	Low	Low	Unclear	High	Low	Low	High	Unclear	Low	High
Rusmawatiningtyas (2009)	Low	Unclear	Low	Low	Low	Low	Low	Low	High	High	High	Low
Bala (2012)	Low	Low	High	Low	Low	High	Low	Low	Low	Unclear	Low	High
Prathapan (2011)	Low	High	Unclear	Low	Unclear	Low	Low	Low	Low	Unclear	Low	Low
Chathurani (2012)	High	High	Unclear	Low	Unclear	Low	Low	Low	Low	Unclear	Low	Low
Aldridge (2012)	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Shah (2014)	Low	Low	Low	Low	Unclear	Low	Low	Low	Low	Low	High	Low

HCS=Haemoglobin colour scale. *Test likely to correctly diagnose anaemia. †Concurrent sampling of HCS and reference test. ‡Included patients match the target population. §Test allows assessment of HCS accuracy. ¶The Haemoglobin Colour Scale from the Indian manufacturer Kruise Path (Ahmedabad, India) is not certified by the original German manufacturer Copack GmbH (appendix p 2 gives more details).

Table 3: QUADAS-2 quality judgments about each domain for each included study

based could not be obtained (appendix p 8). 14 real-life studies remained and are included in this review.[24–37]

Five of the 14 included studies were done in low-income countries and nine in lower middle-income countries: seven in sub-Saharan Africa, one in upper Egypt, three in India, two in Sri Lanka, and one in Indonesia. All but two[25,28] were located in rural areas (table 2). Two studies were embedded into larger morbidity surveys,[32,35] and one study retrospectively investigated the use of HCS as part of a general survey of quality of primary health-care services in Sri Lanka.[27] Two studies examined patients attending hospitals and primary health-care facilities in rural communities.[29,31] In both cases only the data from the field studies were included in this review. One study examined both children and pregnant women.[29] For practical reasons we regarded these data as two separate studies: one in children (Lindblade 2006c) and the other one in pregnant women (Lindblade 2006p).

Seven studies[24,26,29,30,33,35,37] included children (aged from neonates to 11 years), seven studies enrolled pregnant women[25,27–29,31,32,36] and one included women of reproductive age irrespective of their pregnancy status.[34]

The absolute range of anaemia prevalence was 2–83% (median 58%). Only 11 of 15 studies assessed severe anaemia; in two of these studies no cases were found either by HCS or the reference test.[26,27] In the remaining nine studies with available data, 20% was the highest

prevalence reported in one outlier;[31] in the remaining studies prevalence of severe anaemia varied between 0·6% and 10% (median 2%).

Sample sizes ranged between 101 and 1529. In two studies[24,36] the samples were read more than once by different assessors. We report main results excluding these two studies because they inappropriately analysed all ratings of the scale, rather than patients assessed (appendix p 3).

Training intensity varied widely from 1 h[24] to 2 days,[28,31] including one case in which the main study was only started after two raters had reached excellent agreement in a preliminary training pilot.[33] Six studies did not report any information about training.[25–27,32,35,37]

Nine studies used capillary blood samples for the HCS test,[24,26,28,29,31,34–36] three studies did not report which kind of samples were used,[25,27,30] one used venous blood for both the HCS and the reference test,[33] and one used umbilical cord blood at birth and capillary blood in the follow-up for both tests.[37]

Ten studies used the same kind of sample for both tests,[24,26,28,29,31,33–35,37] in four studies venous blood samples for the reference test were tested in distant laboratories,[27,32,33,36] two of these against capillary blood samples for the HCS.[32,36] In three studies, the origin of the blood sample was not disclosed for either one or both tests.[25,27,30]

	Anaemia		Severe anaemia	
	Haemoglobin colour scale	Clinical assessment	Haemoglobin colour scale	Clinical assessment
All studies				
Participants (studies)	7805 (15)	6413 (10)	6663 (9)	5476 (8)
Prevalence	0·58 (0·12–0·83)	0·70 (0·15–0·83)	0·024 (0·006–0·2)	0·02 (0·006–0·2)
Sensitivity (%; 95% CI)	77 (64–86)	52 (36–67)	54 (36–71)	45 (12–83)
Specificity (%; 95% CI)	79 (61–90)	75 (56–88)	99·5 (98–99·9)	92 (62–99)
PV+	0·84	0·83	0·73	0·11
PV–	0·71	0·40	0·99	0·99
All studies without multiple HCS testing*				
Participants (studies)	5813 (13)	6413 (10)	4547 (7)	5476 (8)
Prevalence	0·52 (0·12–0·83)	0·70 (0·15–0·83)	0·025 (0·013–0·2)	0·02 (0·006–0·2)
Sensitivity (%; 95% CI)	80 (68–88)	52 (36–67)	57 (36–76)	45 (12–83)
Specificity (%; 95% CI)	80 (59–91)	75 (56–88)	99·6 (95–99·9)	92 (62–99)
PV+	0·81	0·83	0·79	0·11
PV–	0·79	0·40	0·99	0·99
Studies without high risk of bias† or multiple HCS testing*				
Participants (studies)	4322 (8)	4977 (7)	3646 (5)	4575 (6)
Prevalence	0·62 (0·12–0·83)	0·71 (0·15–0·83)	0·036 (0·024–0·2)	0·028 (0·006–0·2)
Sensitivity (%; 95% CI)	84 (70–92)	46 (34–58)	68 (55–79)	62 (22–90)
Specificity (%; 95% CI)	76 (43–93)	74 (61–83)	99 (95–99·8)	80 (46–95)
PV+	0·85	0·81	0·72	0·08
PV–	0·74	0·36	0·99	0·99
All comparative studies‡ (HCS vs clinical assessment) without multiple HCS testing*				
Participants (studies)	4564 (8)	6413 (10)	4046 (6)	5476 (8)
Prevalence	0·72 (0·15–0·83)	0·70 (0·15–0·83)	0·03 (0·013–0·2)	0·02 (0·006–0·2)
Sensitivity (%; 95% CI)	79 (64–88)	52 (36–67)	53 (27–78)	45 (12–83)
Specificity (%; 95% CI)	77 (52–91)	75 (56–88)	99·6 (93–99·9)	92 (62–99)
PV+	0·90	0·83	0·80	0·11
PV–	0·59	0·40	0·99	0·99
All studies with common threshold for anaemia (Hb <110 g/L) and severe anaemia (Hb <70 g/L)§				
Participants (studies)	6781 (11)	6413 (10)	4547 (7)	4045 (6)
Prevalence	0·70 (0·15–0·83)	0·70 (0·15–0·83)	0·025 (0·013–0·2)	0·03 (0·013–0·2)
Sensitivity (%; 95% CI)	74 (60–84)	52 (36–67)	57 (36–76)	54 (16–88)
Specificity (%; 95% CI)	77 (59–89)	75 (56–88)	99·6 (95–99·9)	91 (44–99)
PV+	0·88	0·83	0·79	0·16
PV–	0·56	0·40	0·99	0·99

Data are n (n), median (range), unless otherwise stated. HCS=Haemoglobin colour scale. PV+=positive predictive value. PV–=negative predictive value. *Aldridge (2012)[24] and van den Broek (1999)[36] allowed multiple observers to assess the same HCS specimen from some of the participants, see main text for details. We report this result as the main pooled analysis since it only includes statistically unbiased studies. The difference between the sensitivity of the HCS and clinical assessment to diagnose anaemia is statistically significant (p=0·008). The difference between the specificity of the HCS and clinical assessment to diagnose anaemia was not statistically significant (p=0·825). For severe anaemia the overall diagnostic accuracy of the HCS is significantly higher than for clinical assessment (p<0·0001). †van Rheenen (2007),[37] Sinha (2008),[35] Bala (2012),[25] Prataphan (2011),[32] and Chathurani (2012)[27] were excluded for high risk of bias. See appendix p 14 for details. ‡Barduagni (2003),[26] van Rheenen (2007),[37] Rusmawatinigtyas (2009),[33] Prataphan (2011),[32] and Shah (2014)[34] did not assess anaemia by clinical assessment. §The following studies used thresholds different from the WHO recommendations in school-age children and pregnant women for the diagnosis of anaemia (<110 g/L): Barduagni (2003;[26] <120 g/L), van Rheenen (2007;[37] different age-specific thresholds for newborn babies), Rusmawatinigtyas (2009),[33] (<115 g/L), and severe anaemia (<70 g/L): van den Broek (1999;[36] <60 g/L), Aldridge (2012;[24] <50 g/L). Shah (2014)[34] tested pregnant and non-pregnant women at the same threshold (<120 g/L).

Table 4: Sensitivity analysis of pooled estimates for HCS and clinical assessment accuracy

HemoCue (HemoCue AB, Ängelholm, Sweden) from capillary blood samples was the most frequent reference standard test (n=9) for practical reasons. Two studies used inappropriate point-of-care methods as reference tests: Sahli's haemometer[25] and the filter paper cyanmethaemoglobin method.[35] In ten studies, the investigators directly compared the performance of clinical assessment for anaemia with the HCS (table 2).[24,25, 27–31,35,36]

We detected high risk of bias in five studies[25,27,32,35,37] and had severe applicability concerns about nine of the 14 studies. In all but two studies[24,28] incomplete reporting demanded an unclear rating in one or more quality relevant domains (figure 2; table 3; appendix p 10, 14).
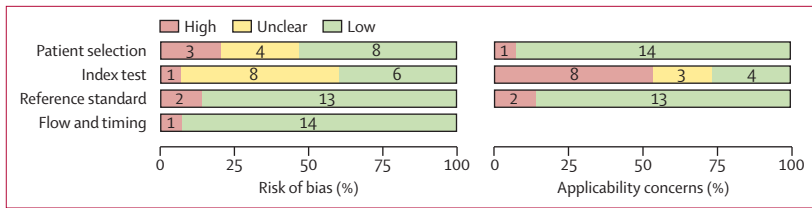
Figure 2: QUADAS-2 judgments about each domain presented as percentages across included studies

	Sensitivity	Specificity	p value
Prevalence of anaemia			
Very high (≥40%)	79	69	0·3068
Low to moderate (<40%)	68	90	..
Population			
Children	83	93	0·3153
Women	66	68	..
Training*			
High (≥0·5 day)	78	74	0·8091
Low (<0·5 day)	74	83	..
Reference test			
Point-of-care test†	78	76	0·5897
Laboratory test‡	66	88	..
Blood sample§			
Same	79	84	0·2721
Different	66	67	..

*We did the meta-regression analysis for training under the assumption that studies without information on training had less than half day (low) of training. †Point-of-care tests included: HemoCue, filter paper cyanmethaemoglobin method, and Sahli's haemometer. ‡Laboratory tests were done in clinical laboratories and included: electronic coulter counter, Hematology Analyzer (HmX), spectrometry method, and laboratory-based cyanmethaemoglobin method. §Blood samples for the HCS and the reference test had either the same origin (capillary, venous, or umbilical cord) or different sources (eg, capillary vs venous). In cases where it was unclear whether the origin was the same, we assumed that the sources of the blood sample were different.

Table 5: Meta-regression analysis of the effect of covariates on HCS accuracy by potential sources of heterogeneity

The diagnostic accuracy of the HCS to diagnose anaemia varied widely across individual studies; sensitivities ranged from 33% to 96% and specificities from 14% to 100% (figure 3).

The meta-analysis from 13 statistically unbiased studies—ie, excluding those with multiple counts from the same sample[24,36]—showed a higher pooled sensitivity of 80% (95% CI 68–88) for the HCS compared with 52% (36–67) for clinical assessment (p=0·008; figure 4). Pooled specificities were similar at 80% (95% CI 59–91) for the HCS and 75% (56–88) for clinical assessment (p=0·8250).

When we included the eight statistically unbiased studies (without multiple HCS testing) that explicitly compared the HCS with clinical assessment within the same study to diagnose anaemia (median anaemia prevalence: 70% [range 15–83], the pooled results were very similar: HCS sensitivity 79% (95% CI 64–88) vs clinical assessment sensitivity 52% (36–67; p=0·0289)

and HCS specificity 77% (52–91) vs clinical assessment specificity 75% (56–88; p=0·8649). Whether we included all studies or excluded studies that had an unacceptable number of exclusions or withdrawals of participants, did not use an appropriate reference standard, used a non-certified version of the HCS, or a cut-off for anaemia that differed from 110 g/L, made little difference to the results (table 4).

For the diagnosis of severe anaemia, the diagnostic accuracy across individual studies showed a similar heterogeneity (specificities 19% to 91%; sensitivities 13% to 98%; figure 5). In the meta-analysis, the HCS again performed better (p<0·0001), yielding 57% (95% CI 36–76) sensitivity compared with 45% (12–83) by clinical assessment (figure 6). Specificity for the HCS was 99·6% (95% CI 95–99·9)—higher than the estimate of 92% (62–99) for clinical assessment; again we saw little differences in the sensitivity analysis (table 4).

In practice, it is likely that primary health-care workers would use both the HCS and clinical assessment to diagnose anaemia, resulting in a net gain in sensitivity. In studies examining both methods, the sensitivity of a positive result on either the HCS or clinical assessment for anaemia rose to 91% (95% CI 81–96) after excluding studies with inappropriate multiple assessments[24,36] and an unacceptable amount of missing HCS values.[27] However, to rule out anaemia, results from both methods would have to be negative, which leads to a net loss of specificity to 59% (95% CI 35–79) for simultaneous testing.[38] For severe anaemia, simultaneous testing would yield a pooled net sensitivity of 83% (95% CI 33–98) in the six comparative studies without multiple assessments for the HCS, whereas the specificity would decrease to 90% (95% CI 40–99).

Meta-regression analyses did not show a significant effect of the covariates population group, anaemia prevalence, reference test, training quantity, and source of blood sample (table 5; appendix p 15), although this could be due to incomplete reporting—eg, for training, or small numbers of studies (with use of appropriate laboratory reference tests).

Discussion

We systematically reviewed the literature to assess the accuracy of the HCS to diagnose anaemia and severe anaemia when used by primary health-care workers in resource-poor settings, and compared this with the accuracy of assessment by clinical signs alone. Publication bias can never be ruled out completely, but the search was comprehensive and no studies were excluded due to language of publication.

We have identified substantial heterogeneity of accuracy outcomes between the selected 14 studies, with sensitivities ranging from 33% to 96% and specificities from 14% to 100% for the HCS. We could not fully account for this heterogeneity, possibly because of the small number of studies or incomplete reporting of key
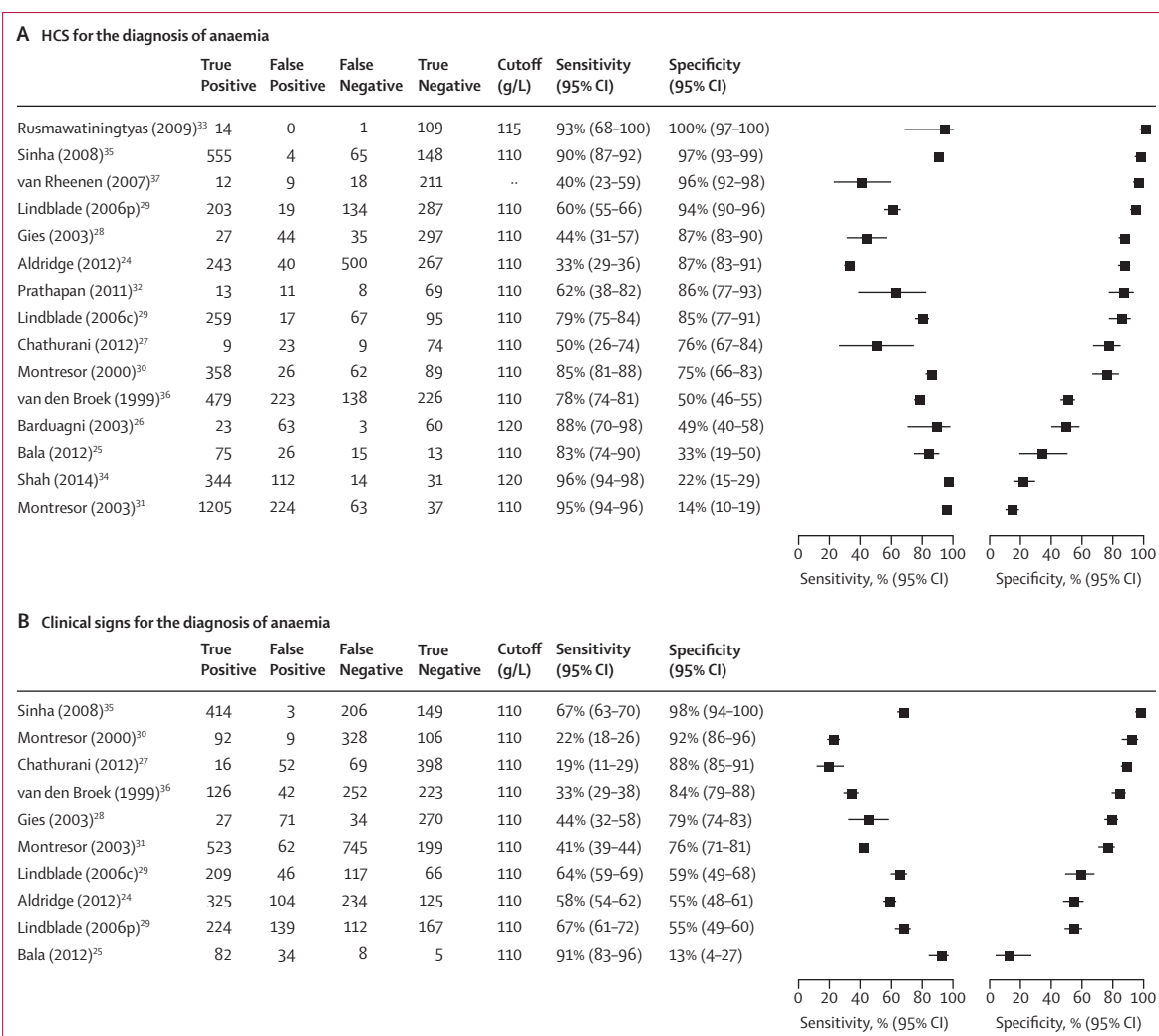
A HCS for the diagnosis of anaemia

	True Positive	False Positive	False Negative	True Negative	Cutoff (g/L)	Sensitivity (95% CI)	Specificity (95% CI)
Rusmawatiningtyas (2009)[33]	14	0	1	109	115	93% (68–100)	100% (97–100)
Sinha (2008)[35]	555	4	65	148	110	90% (87–92)	97% (93–99)
van Rheenen (2007)[37]	12	9	18	211	··	40% (23–59)	96% (92–98)
Lindblade (2006p)[29]	203	19	134	287	110	60% (55–66)	94% (90–96)
Gies (2003)[28]	27	44	35	297	110	44% (31–57)	87% (83–90)
Aldridge (2012)[24]	243	40	500	267	110	33% (29–36)	87% (83–91)
Prathapan (2011)[32]	13	11	8	69	110	62% (38–82)	86% (77–93)
Lindblade (2006c)[29]	259	17	67	95	110	79% (75–84)	85% (77–91)
Chathurani (2012)[27]	9	23	9	74	110	50% (26–74)	76% (67–84)
Montresor (2000)[30]	358	26	62	89	110	85% (81–88)	75% (66–83)
van den Broek (1999)[36]	479	223	138	226	110	78% (74–81)	50% (46–55)
Barduagni (2003)[26]	23	63	3	60	120	88% (70–98)	49% (40–58)
Bala (2012)[25]	75	26	15	13	110	83% (74–90)	33% (19–50)
Shah (2014)[34]	344	112	14	31	120	96% (94–98)	22% (15–29)
Montresor (2003)[31]	1205	224	63	37	110	95% (94–96)	14% (10–19)

Sensitivity, % (95% CI)   Specificity, % (95% CI)

B Clinical signs for the diagnosis of anaemia

	True Positive	False Positive	False Negative	True Negative	Cutoff (g/L)	Sensitivity (95% CI)	Specificity (95% CI)
Sinha (2008)[35]	414	3	206	149	110	67% (63–70)	98% (94–100)
Montresor (2000)[30]	92	9	328	106	110	22% (18–26)	92% (86–96)
Chathurani (2012)[27]	16	52	69	398	110	19% (11–29)	88% (85–91)
van den Broek (1999)[36]	126	42	252	223	110	33% (29–38)	84% (79–88)
Gies (2003)[28]	27	71	34	270	110	44% (32–58)	79% (74–83)
Montresor (2003)[31]	523	62	745	199	110	41% (39–44)	76% (71–81)
Lindblade (2006c)[29]	209	46	117	66	110	64% (59–69)	59% (49–68)
Aldridge (2012)[24]	325	104	234	125	110	58% (54–62)	55% (48–61)
Lindblade (2006p)[29]	224	139	112	167	110	67% (61–72)	55% (49–60)
Bala (2012)[25]	82	34	8	5	110	91% (83–96)	13% (4–27)

Sensitivity, % (95% CI)   Specificity, % (95% CI)

Figure 3: **Forest plots of all studies diagnosing anaemia by HCS and clinical assessment**
HCS=Haemoglobin colour scale.

methods. Heterogeneity might be explained by differences in the quality of methods, anaemia prevalence, training intensity, the choice of the reference test, and the source of the blood sample.

Whether the use of different blood samples (capillary, venous, or umbilical cord) between studies could have been a reason for heterogeneity is unclear. Discrepancies between the standard test and the HCS might have been exaggerated by the fact that the origins of the blood samples also varied within at least two studies.[32,36] Transport conditions or suboptimal storage could potentially have damaged blood specimens in four studies[27,32,33,36] in which the reference test was done in a distant laboratory, although this was not mentioned in the studies.

Intensity of training varied substantially and was poorly reported. We could not identify a relation between training and accuracy outcomes. However, during HCS development, it was shown that trainees' performance improved significantly with further familiarisation, even

after receiving an initial 30 min demonstration.[10] Consequently, the original training protocol required two training sessions of about 2 h on 2 consecutive days. Others have shown inter-rater variations even if adhering to the protocol[39] and some have suggested that easy-to-read instructions, cartoons, and coloured test strips might improve accuracy.[40] Unfortunately, once the HCS became commercially available, no further evidence was collated to refine the training protocol, possibly explaining the variation in training across the included studies.

Although laboratory-based methods remain the gold standard for the measurement of haemoglobin,[5] most studies used the HemoCue, which is easy to use, battery powered, and requires only a small amount of blood because of the use of microcuvettes. Although its accuracy compared against the gold standard is good, venous and arterial samples yield more accurate results than those obtained from capillary blood[5,7,41] and high humidity might alter the functionality of the microcuvettes.[42]
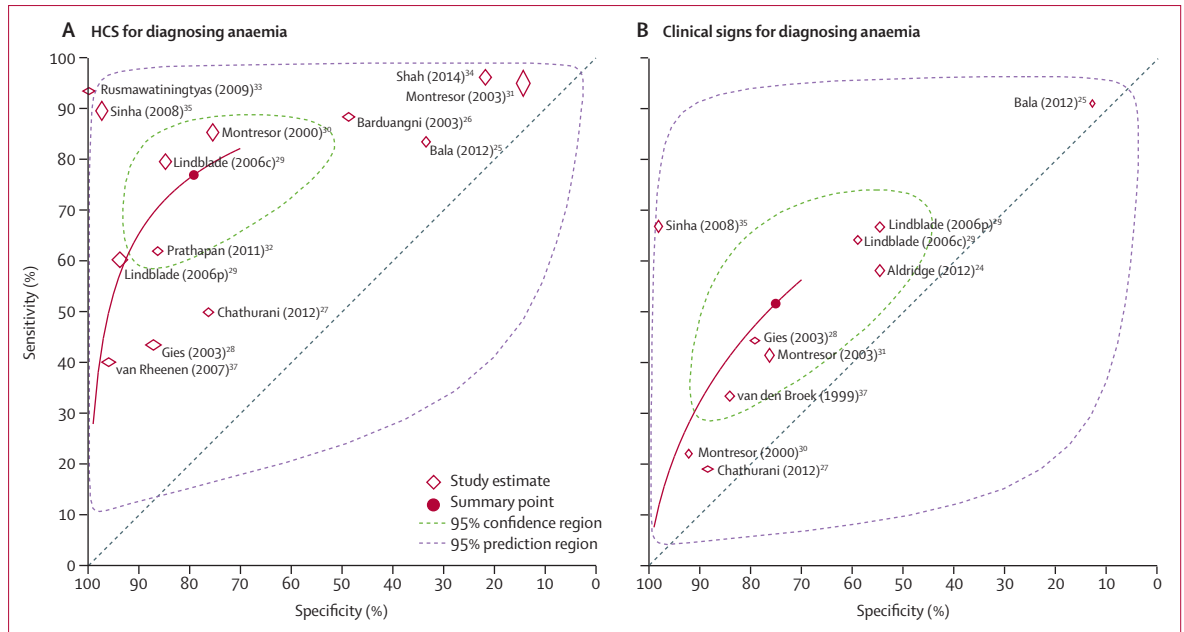
Figure 4: Summary ROC plot of studies diagnosing anaemia by HCS and clinical assessment
Note: Aldridge (2012) and van den Broek (1999) excluded from summary estimate for the HCS assessment for allowing observers to assess the same HCS specimen from some of the participants, see main text for details. The weights for analysis are inverse variance, size points of individual studies represent sample sizes.
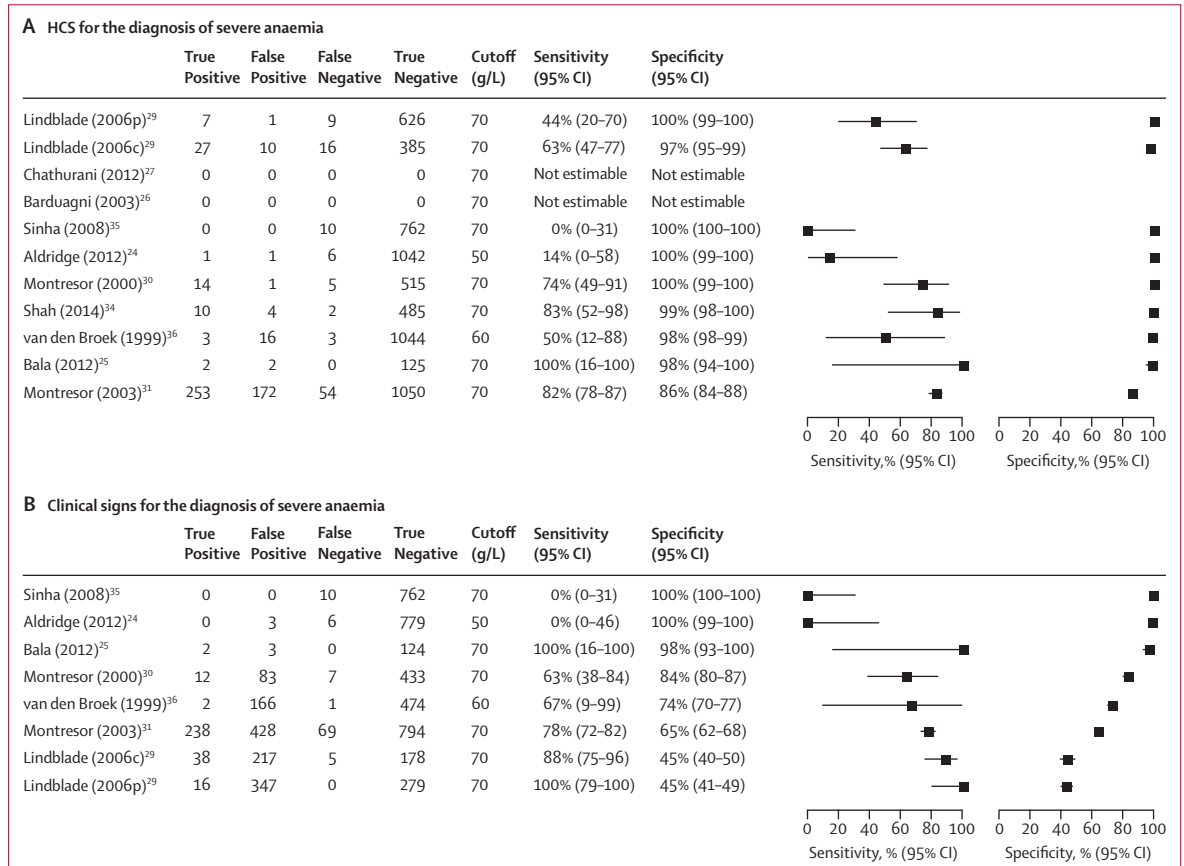HCS=Haemoglobin colour scale.

A HCS for the diagnosis of severe anaemia

	True Positive	False Positive	False Negative	True Negative	Cutoff (g/L)	Sensitivity (95% CI)	Specificity (95% CI)
Lindblade (2006p)[29]	7	1	9	626	70	44% (20–70)	100% (99–100)
Lindblade (2006c)[29]	27	10	16	385	70	63% (47–77)	97% (95–99)
Chathurani (2012)[27]	0	0	0	0	70	Not estimable	Not estimable
Barduagni (2003)[26]	0	0	0	0	70	Not estimable	Not estimable
Sinha (2008)[35]	0	0	10	762	70	0% (0–31)	100% (100–100)
Aldridge (2012)[24]	1	1	6	1042	50	14% (0–58)	100% (99–100)
Montresor (2000)[30]	14	1	5	515	70	74% (49–91)	100% (99–100)
Shah (2014)[34]	10	4	2	485	70	83% (52–98)	99% (98–100)
van den Broek (1999)[36]	3	16	3	1044	60	50% (12–88)	98% (98–99)
Bala (2012)[25]	2	2	0	125	70	100% (16–100)	98% (94–100)
Montresor (2003)[31]	253	172	54	1050	70	82% (78–87)	86% (84–88)

B Clinical signs for the diagnosis of severe anaemia

	True Positive	False Positive	False Negative	True Negative	Cutoff (g/L)	Sensitivity (95% CI)	Specificity (95% CI)
Sinha (2008)[35]	0	0	10	762	70	0% (0–31)	100% (100–100)
Aldridge (2012)[24]	0	3	6	779	50	0% (0–46)	100% (99–100)
Bala (2012)[25]	2	3	0	124	70	100% (16–100)	98% (93–100)
Montresor (2000)[30]	12	83	7	433	70	63% (38–84)	84% (80–87)
van den Broek (1999)[36]	2	166	1	474	60	67% (9–99)	74% (70–77)
Montresor (2003)[31]	238	428	69	794	70	78% (72–82)	65% (62–68)
Lindblade (2006c)[29]	38	217	5	178	70	88% (75–96)	45% (40–50)
Lindblade (2006p)[29]	16	347	0	279	70	100% (79–100)	45% (41–49)

Figure 5: Forest plots of all studies diagnosing severe anaemia by HCS and clinical assessment
HCS=Haemoglobin colour scale.
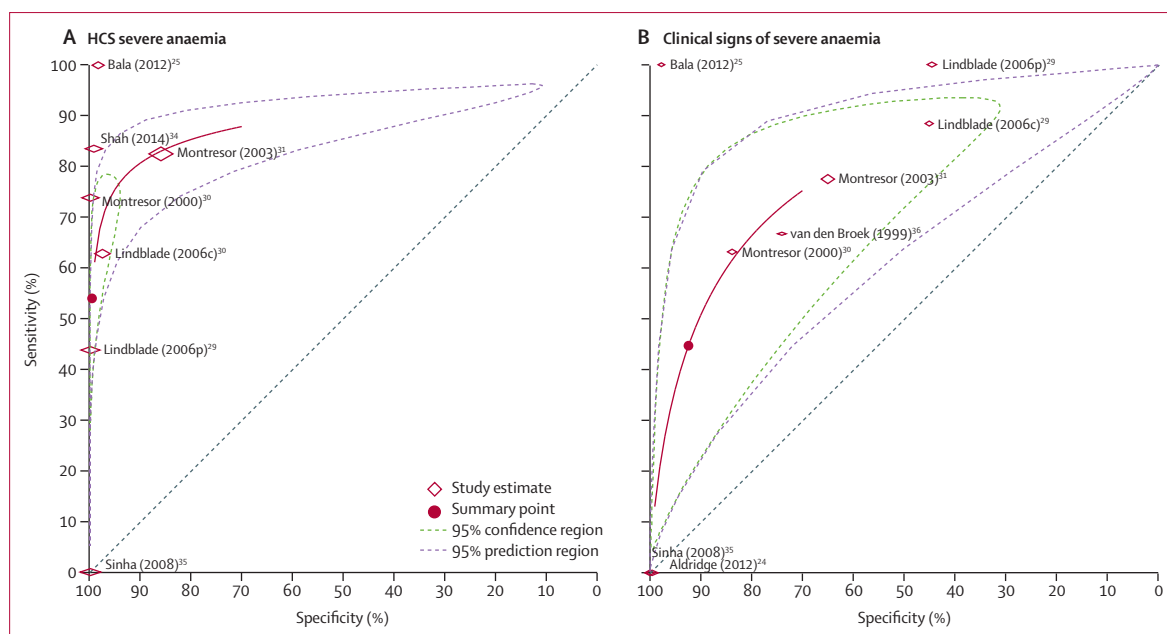
Figure 6: **Summary ROC plot of studies assessing severe anaemia by HCS and clinical assessment**
Note: Aldridge (2012) and van den Broek (1999) excluded from summary estimate for the HCS assessment for allowing observers to assess the same HCS specimen from some of the participants, see main text for details. The weights for analysis are inverse variance, size points of individual studies represent sample sizes. HCS=Haemoglobin colour scale.

It was also unavoidable that our selection criteria allowed four studies to be included that did not completely comply with the real-life approach with respect to the person who did the HCS assessment. Four studies used cut-offs for the definition of anaemia that were not in line with WHO recommendations[19] and we identified five studies that had introduced a high risk of bias, which we handled by excluding them in a sensitivity analysis (table 4). Two studies introduced statistical bias including multiple counts from the same sample in their analysis, which obliged us to exclude them from the summary estimates, but in most studies (n=12) the possibility of bias was hard to assess due to incomplete reporting of methods.

Despite these limitations, our pooled estimates suggest that in real-life circumstances the HCS significantly improves the accuracy of the diagnosis of anaemia. By clinical examination alone, 48% of patients with mild-to-moderate anaemia would be missed. The HCS alone might significantly reduce this proportion to 20%. Although in study settings both methods were assessed independently, in reality they would be combined as simultaneous tests in addition to the patient's history. We would expect a net gain in sensitivity from 80% (HCS) and 52% (clinical assessment) for the single methods to 91% if the diagnosis of anaemia was considered with either or both methods being positive (severe anaemia: net sensitivity 83%). However, the potential cost of use of both methods simultaneously would be a loss of specificity.

The public health relevance is best shown by an example: 80% of Malawi's 15 million people live in rural areas; among these are 6·5 million women, of whom 2·7 million have anaemia (anaemia prevalence 45%). Nearly every second woman—ie, 1·3 million—would have the correct diagnosis missed through assessment of the clinical signs only. The HCS alone would reduce the number of underdiagnosed women from 1·3 million to 0·5 million, hence 800 000 additional women would receive the appropriate diagnosis and potentially correct care. If use of both clinical assessment and HCS was combined, more than 1 million additional women would be diagnosed correctly.

Unfortunately, the reduction of underdiagnosis diminishes when anaemia becomes severe. In this case, the HCS leaves 43% undetected, whereas the assessment of clinical signs leaves 55% undetected. The HCS is able to significantly reduce the number of those falsely diagnosed with severe anaemia (0·4% *vs* 7·6%), hence preventing a large number of patients from unnecessary and potentially harmful blood transfusions or cost-intensive referrals.

Both methods do not significantly differ between the amount of non-anaemic patients being wrongly diagnosed with mild-to-moderate anaemia, which would be the case by clinical assessment in 25% and with the HCS in 20%. Overdiagnosis of mild-to-moderate anaemia is predominantly an economic issue. It increases expenses for unneeded supplementation therapy or unnecessary further diagnostic investigations in settings where resources are already poor.

However, one advantage of the HCS is that it delivers quantitative results, whereas the clinical assessment is purely qualitative. Although the available studies do not allow an inference about the effect of the knowledge of

continuous values on clinical decisions, such decisions will probably be more strongly affected by borderline results close to the defined thresholds of severe anaemia than by the clinical assessment alone. Unfortunately, none of the studies assessed the effectiveness of the HCS, such as the effect on clinical outcomes or its cost-effectiveness.

Almost 15 years after it became commercially available, the HCS remains the most simple to use and affordable point-of-care device to assess the concentration of haemoglobin quantitatively. However, clinical outcomes depend on the management decisions made by primary health-care workers who have diagnosed anaemia, regardless of the method used. The results from the HCS are prone to individually erroneous readings by individual health-care workers, who in case of discordant results have to decide whether to rely on their clinical judgment or the HCS. Taking into account the potential clinical and economic consequences of misdiagnosis and in view of the evidence that the HCS yields a significantly better sensitivity and a similar specificity for mild-to-moderate anaemia, but a similarly poor sensitivity and a better specificity for severe anaemia, we recommend that the HCS result should over-rule the clinical judgment in most cases, but for severe anaemia a positive HCS might be over-ruled if clinical signs are missing. Whether a short-term follow-up of patients with discordant or borderline results would improve their clinical outcome remains to be assessed.

Public health decision makers should be aware that the use of the HCS might require more training and supervision than technically more sophisticated devices. To tap the full potential of the HCS, an evidence-based standardised training protocol that has to be as short and cost-effective as possible under the pressure of poor resources is urgently needed. Future research should also address endpoints beyond the diagnostic accuracy of the HCS, such as its potential to reduce morbidity and mortality associated with anaemia and the cost-effectiveness of use of the HCS in routine practice.

References
1 Ezzati M, Lopez AD, Rodgers A, Murray CJ. Comparative Quantification of Health Risks, Vol.1: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Geneva: World Health Organization, 2004.
2 Haas JD, Brownlie Tt. Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship. *J Nutr* 2001; **131:** 676S–88S.
3 Balarajan Y, Ramakrishnan U, Ozaltin E, Shankar AH, Subramanian SV. Anaemia in low-income and middle-income countries. *Lancet* 2011; **378:** 2123–35.
4 WHO. Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia. Geneva, Switzerland: World Health Organization, 2008.
5 WHO. Assessing the iron status of populations, 2007. http://www. who.int/nutrition/publications/micronutrients/anaemia_iron_deficiency/9789241596107/en/ (accessed June 21, 2015).
6 WHO. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System World Health Organization, 2011. http://www.who.int/vmnis/indicators/haemoglobin (accessed Jan 18, 2015).
7 Briggs C, Kimber S, Green L. Where are we at with point- of- care testing in haematology? *Br J Haematol* 2012; **158:** 679–90.
8 Chalco JP, Huicho L, Alamo C, Carreazo NY, Bada CA. Accuracy of clinical pallor in the diagnosis of anaemia in children: a meta-analysis. *BMC Pediatr* 2005; **5:** 46.
9 WHO. Report of consultation on standardization in haematology. Geneva: World Health Organization, 1975.
10 Lewis SM, Stott GJ, Wynn KJ. An inexpensive and reliable new haemoglobin colour scale for assessing anaemia. *J Clin Pathol* 1998; **51:** 21–24.
11 Stott GJ LS. A simple and reliable method for estimating haemoglobin. *Bull World Health Organ* 1995; **73:** 369–73.
12 WHO. Haemoglobin Colour Scale: Operational Research Agenda and Study Design, 2004. http://www.who.int/medical_devices/publications/en/HbCS_Research_AgendaStudyDesign.pdf (accessed Jan 18, 2015).
13 Copack. How to screen anaemia in the absence of laboratory-based haemoglobinometry? In: Manufacturer, editor. manufacturer's product brochure. Oststeinbeck, Germany: Copack medical branch, 2015.
14 Paddle JJ. How objective are the supporters of the Haemoglobin Colour Scale? *Bull World Health Organ* 2002; **80:** 987.
15 Cherian M, Emmanuel JC, Lewis SM, et al. Evaluation of the haemoglobin colour scale. *Bull World Health Organ* 2002; **80:** 839.
16 Critchley J, Bates I. Haemoglobin colour scale for anaemia diagnosis where there is no laboratory: a systematic review. *Int J Epidemiol* 2005; **34:** 1425–34.
17 Beynon R, Leeflang MM, McDonald S, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database Syst Rev* 2013; **9:** MR000022.
18 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155:** 529–36.
19 WHO. Iron deficiency anaemia: assessment prevention and control, 2001. http://www.who.int/nutrition/publications/micronutrients/anaemia_iron_deficiency/WHO_NHD_01.3/en/index.html (accessed Nov 12, 2012).
20 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; **58:** 982–90.
21 Dwamena B. Multivariate mixed models for metanalysis of paired-comparison studies of two medical diagnostic tests. Stata Users Group, 2008.
22 Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009; **9:** 211–29.
23 Asante K, Owusu-agyei S, Dosoo D, Agyei G. An valuation of the WHO haemoglobin colour scale among young children in the Kintampo District of Ghana MIM-KA-25200. *Acta Trop* 2005; **95:** S381–88S.
24 Aldridge C, Foster HM, Albonico M, Ame SM, Montresor A. Evaluation of the diagnostic accuracy of the Haemoglobin Colour Scale to detect anaemia in young children attending primary healthcare clinics in Zanzibar. *Trop Med Int Health* 2012; **17:** 423–29.
25 Bala DV, Vyas S, Shukla A, Tiwari H, Bhatt G, Gupta K. Validity and reliability of haemoglobin colour scale and its comparison with clinical signs in diagnosing anaemia in pregnancy in Ahmedabad, India. *East Mediterr Health J* 2012; **18:** 749–54.
26 Barduagni P, Ahmed AS, Curtale F, Raafat M, Soliman L. Performance of Sahli and colour scale methods in diagnosing anaemia among school children in low prevalence areas. *Trop Med Int Health* 2003; **8:** 615–18.

27 Chathurani U, Dharshika I, Galgamuwa D, Wickramasinghe ND, Agampodi TC, Agampodi SB. Anaemia in pregnancy in the district of Anuradhapura, Sri Lanka—need for updating prevalence data and screening strategies. *Ceylon Med J* 2012; **57:** 101–06.

28 Gies S, Brabin BJ, Yassin MA, Cuevas LE. Comparison of screening methods for anaemia in pregnant women in Awassa, Ethiopia. *Trop Med Int Health* 2003; **8:** 301–09.

29 Lindblade KA, Mwololo K, van Eijk AM, et al. Evaluation of the WHO Haemoglobin Colour Scale for diagnosis of anaemia in children and pregnant women as used by primary health care nurses and community health workers in western Kenya. *Trop Med Int Health* 2006; **11:** 1679–87.

30 Montresor A, Albonico 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:** 129–33.

31 Montresor A, Ramsan M, Khalfan N, et al. Performance of the Haemoglobin Colour Scale in diagnosing severe and very severe anaemia. *Trop Med Int Health* 2003; **8:** 619–24.

32 Prathapan S, Lindmark G, Fonseka P, Lokubalasooriya A, Prathapan R. How good is the quality of antenatal care in the Colombo district of Sri Lanka in diagnosing and treating anaemia? *Qual Prim Care* 2011; **19:** 245–50.

33 Rusmawatiningtyas Desy SD, Mulatsih Sri , Sutaryo. Early detection of anemia among school children using the World Health Organization Hemoglobin Color Scale 2006. *Paediatr Indones* 2009; **49:** 135–38.

34 Shah PP, Desai SA, Modi DK, Shah SP. Assessing diagnostic accuracy of Haemoglobin Colour Scale in real-life setting. *J Health Popul Nutr* 2014; **32:** 51–57.

35 Sinha N, Deshmukh PR, Garg BS. Evaluation of WHO haemoglobin colour scale & palmar pallor for screening of anaemia among children (6-35 months) in rural Wardha, India. *Indian J Med Res* 2008; **128:** 278–81.

36 van den Broek NR, Ntonya C, Mhango E, White SA. Diagnosing anaemia in pregnancy in rural clinics: assessing the potential of the Haemoglobin Colour Scale. *Bull World Health Organ* 1999; **77:** 15–21.

37 van Rheenen PF, de Moor LT. Diagnostic accuracy of the haemoglobin colour scale in neonates and young infants in resource-poor countries. *Trop Doct* 2007; **37:** 158–61.

38 Gordis L. Assessing the validity and reliability of diagnostic and screening tests. Epidemiology, 4th edn. Philadelphia: Elsevier; 2009: pp 92–96.

39 Munster M, Lewis SM, Erasmus LK, Mendelow BV. Field evaluation of a novel haemoglobin measuring device designed for use in a rural setting. *S Afr Med J* 1997; **87:** 1522–26.

40 Gosling R, Walraven G, Manneh F, Bailey R, Lewis SM. Training health workers to assess anaemia with the WHO haemoglobin colour scale. *Trop Med Int Health* 2000; **5:** 214–21.

41 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.

42 Nguyen HT. High humidity affects HemoCue cuvette function and HemoCue haemoglobin estimation in tropical Australia. *J Paediatr Child Health* 2002; **38:** 427–28.