Elsevier Editorial System(tm) for The Lancet Global Health

Manuscript Draft

Manuscript Number: LANGLH-D-15-00864R1

Title: The WHO Haemoglobin Colour Scale can improve the accuracy of the diagnosis of anaemia in primary health care settings in low-income countries: a systematic review and meta-analysis

Article Type: Articles (Original Research)

Keywords: WHO Haemoglobin Colour Scale
low-income countries
anaemia
diagnosis
primary health care
systematic review
meta-analysis
diagnostic accuracy

Corresponding Author: Mr. Heiko Marn,

Corresponding Author's Institution: Institute of Tropical Medicine and International Health, Charité-Universitätsmedizin

First Author: Heiko Marn

Order of Authors: Heiko Marn; Julia A Critchley, MSc,DPhil

Manuscript Region of Origin: GERMANY

Abstract: 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 level outside the laboratory. We systematically reviewed the literature to assess accuracy of the HCS in primary care to diagnose anaemia, and compared this with clinical signs.
Methods
We searched the electronic databases including MEDLINE, EMBASE, SCOPUS, Web of Science, Cochrane library, CINAHL plus, Popline, Reproductive Health Library, Google Scholar and regional databases up to November 2014. Two reviewers independently screened studies, extracted data and assessed quality using the QUADAS-2 tool. Statistical analyses were carried out in STATA using the bivariate model.
Findings
We included 14 studies from Africa and Asia, most carried out in children and pregnant women. The pooled sensitivity of the HCS to diagnose anaemia was 80% (95% CI 68%-89%), significantly higher than sensitivity for clinical signs (52%, 95% CI 36%-67%; p=0·008). Specificity was similar for the HCS (80%, 95% CI 59%-91%) and clinical signs (75%, 95% CI 56%-88%, p=0·8395).
For severe anaemia, diagnostic accuracy was again higher overall for the HCS (p<0.0001); sensitivity was similar: 57% (95% CI 36%-76%) for HCS and 43% (95% CI 9%-85%) for clinical signs, but specificity appeared higher:

99.6% (95% CI 95%-100%) versus 93% (95% CI 56%-99%). Combining clinical signs and the HCS would result in higher sensitivity (anaemia: 92%, 95% CI 83% -97%; severe anaemia: 90%, 95% CI 33%-100%:), but at the expense of specificity (anaemia: 60%, 95% CI 33%-82%; severe anaemia: 84%, 95% CI 40%-98%). Individual studies were highly heterogeneous but pooled results did not differ markedly in a series of sensitivity analyses for indicators of study robustness.
Interpretation
Under "real life" primary health care conditions the HCS can significantly reduce misdiagnosis of anaemia compared to clinical assessment alone. Future research is required to optimise training, and evaluate clinical outcomes and cost-effectiveness.
Funding
None

Title

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

Authors: Heiko Marn*[a] and Julia Alison Critchley[b]

[a] *corresponding author: Institute of Tropical Medicine and International Health, Charité - Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, email: heiko.marn@charite.de Tel. +49 176 61080481

[b] Population Health Research Institute, St George's, University of London, London SW17 0RE, UK.

Abstract

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 level outside the laboratory. We systematically reviewed the literature to assess accuracy of the HCS in primary care to diagnose anaemia, and compared this with clinical signs.

Methods

We searched the electronic databases including MEDLINE, EMBASE, SCOPUS, Web of Science, Cochrane library, CINAHL plus, Popline, Reproductive Health Library, Google Scholar and regional databases up to November 2014. Two reviewers independently screened studies, extracted data and assessed quality using the QUADAS-2 tool. Statistical analyses were carried out in STATA using the bivariate model.

Findings

We included 14 studies from Africa and Asia, most carried out in children and pregnant women. The pooled sensitivity of the HCS to diagnose anaemia was 80% (95% CI 68%-89%), significantly higher than sensitivity for clinical signs (52%, 95% CI 36%-67%; p=0·008). Specificity was similar for the HCS (80%, 95% CI 59%-91%) and clinical signs (75%, 95% CI 56%-88%, p=0·8395).

For severe anaemia, diagnostic accuracy was again higher overall for the HCS (p<0.0001); sensitivity was similar: 57% (95% CI 36%-76%) for HCS and 43% (95% CI 9%-85%) for clinical signs, but specificity appeared higher: 99.6% (95% CI 95%-100%) versus 93% (95% CI 56%-99%). Combining clinical signs and the HCS would result in higher sensitivity (anaemia: 92%, 95% CI 83% -97%; severe anaemia: 90%, 95% CI 33%-100%:), but at the expense of specificity (anaemia: 60%, 95% CI 33%-82%; severe anaemia: 84%, 95% CI 40%-98%). Individual studies were highly heterogeneous but pooled results did not differ markedly in a series of sensitivity analyses for indicators of study robustness.

Interpretation

Under “real life” primary health care conditions the HCS can significantly reduce misdiagnosis of anaemia compared to clinical assessment alone. Future research is required to optimise training, and evaluate clinical outcomes and cost-effectiveness.

Funding

None

Background

Anaemia is a major global cause of maternal, perinatal, and child mortality. Additionally, it causes low birth weight, impaired or delayed child physical and mental development, and a higher susceptibility to infections,[1] and contributes considerably to economic loss due to reduced productivity of workers.[2] Approximately 1·62 billion people are affected.[1] The majority are non-pregnant women (468·4 million), pre-school-age children (293·1 million) and pregnant women (56·4 million) predominantly in low-income countries, where prevalence rates are up to 5-fold 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 approximately 50% of all anaemia cases,[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 by measuring a decreased amount of red blood cells or decreased haemoglobin (Hb) 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 pathogens, and wasting resources in case of misdiagnosed severe anaemia.

In response to the need for a "simple, cheap, and robust device for measuring 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, Germany, since 2001.[10-12] It comprises a small card of six shades of red, each representing a haemoglobin level of 4,6,8,10,12 and 14 g/dl, respectively. A drop of blood absorbed onto a standardized chromatography filter paper is compared with the colour scale, allowing assessment of the patient´s haemoglobin level, including estimation of intermediate results, in 1g/dl steps.[13]

The usefulness of the device in practice has been disputed,[14, 15] but in 2005 a systematic review of 14 studies found that under ideal conditions, the HCS may 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] The diagnostic accuracy tended to be lower in the four "real-life" studies (sensitivities 76%-88%, apart from one outlier, and specificities from 41% -100%) leading to the conclusion that further research was needed to evaluate the usefulness of the HCS in real-life situations. Only a minority (n=5) compared the accuracy of HCS with clinical diagnosis. There have been no systematic reviews of HCS performance that we are aware of since this time though additional 'real life' studies have been published since 2005.

Objectives

The HCS aims to improve the diagnosis of anaemia in clinical settings with poor resources, where clinical diagnosis remains the standard diagnostic procedure. This systematic literature review therefore assesses the accuracy of the HCS to diagnose anaemia and severe anaemia in resource-poor primary health care settings. The accuracy of clinical signs for the diagnosis of anaemia alone is also compared to the HCS method, whenever such data is available.

Methods

Types of studies

We included all studies comparing the diagnostic accuracy of the HCS with any reference method ("gold standard") to diagnose anaemia under "real life" conditions, i.e. in patient populations attending routine primary health clinics, with the HCS performed by primary health care workers or a person with comparable skills or training. There were no restrictions based on sample size, location, background morbidities or anaemia prevalence. Studies carried out in hospitals, laboratories or blood banks were excluded because they are not generalizable towards primary health care in low resource settings.

Search methods

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 November 14th 2014. To increase sensitivity of the search strategy,[17] we searched only the key words "haemoglobin colour scale" without any filters using alternative spellings in English, Spanish and French. A citation search on "Critchley and Bates 2005 systematic review" was carried out 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. (See table 1; online supplement for search history).

Data extraction and management

Both authors independently extracted data including: main study outcomes, study characteristics and quality related information based on WHO recommendations for HCS evaluations.[12] (See the template data extraction form, table 2 in the online supplement)

Assessment of methodological quality

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 tool[18] and applied customized 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 performed all HCS readings. Again, both authors independently extracted data on all these aspects of quality using a standardised form. Any disagreements were resolved by discussion between authors.

Table 1 (signalling questions)

Data synthesis and statistical analyses

Both authors independently extracted the study outcomes for true positive, true negative, false positive and false negative test results into 2x2 tables. The haemoglobin cut-off level in children aged 6-59 months and during pregnancy for diagnosing anaemia was 11 g/dl and 7 g/dl for severe anaemia according to WHO recommendation.[19] Studies with a different threshold for anaemia and severe anaemia were included into the meta-analysis, but excluded in a sensitivity analysis. We assessed heterogeneity between studies by creating forest plots and summary ROC curves.

Overall summary estimates of diagnostic accuracy of the HCS and clinical signs

We used the bivariate random effects model to combine data across all studies included. This analyses pairs of sensitivity and specificity 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 for the HCS and "clinical signs" 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 which did not adjust for multiple readings of HCS results from the same patient, studies using different cut-offs for anaemia and severe anaemia would affect the pooled accuracy estimates. We then also repeated analyses restricted to the 10 studies that compared the HCS directly with clinical diagnosis, to assess whether confounding by study was affecting comparisons.

Comparison of the diagnostic accuracy of the HCS with clinical signs

We compared the diagnostic accuracy of the HCS with clinical diagnosis in a meta-regression analysis (adding test as a co-variate), 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 which 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 performed these models in all studies initially and then only in those studies which 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 to explore heterogeneity between studies estimating the diagnostic accuracy of the HCS

Using the same bivariate random effects model, we performed meta-regression analysis adding covariates in sequence to assess whether the following variables could explain any of the heterogeneity between studies. These were:

a) level of training (greater or less than half a day),

b) the type of reference test (standard laboratory test or point-of-care test),
c) whether both the HCS and reference test used the same type of blood sample (i.e. both used capillary blood or both used venous blood) or a different sample;
d) the population type (women or children)
e) anaemia prevalence (40% or higher, compared with less than 40%),

In this meta-regression, we assumed that training levels were "low" for the 4 studies that did not report this and that the type of blood sample was different for the 3 studies that did not state this clearly. We used the statistical software Review manager version 5.3 and STATA 12 statistical software packages metandi, gllamm and xtmelogit for meta-analysis and meta-regression modelling.[20, 22] (See online appendix 1 for further details)

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.

Results

Results of search

A total number of 141 records were screened for eligibility based on titles and abstracts. We excluded 98 papers (see flow-chart figure 1 and table 3 of the online supplement) based on titles and abstracts. The remaining 43 full-text articles were assessed for eligibility. Most of the 29 excluded articles did not meet the previously defined "real-life" inclusion criteria: they were performed in blood banks (n=14), hospitals (n=8), a laboratory (n=1), had a mixed field/laboratory design (n=3), or did not report diagnostic accuracy data (n=2). For one congress abstract[23] information whether it was field or lab based could not be obtained (See table 4 online supplement). 14 real-life studies remained and are included into this review.[24-37]

Figure 1 (flow diagram)

Study characteristics

Five of the 14 included studies have been carried out 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 (see 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 the 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 (age distribution 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% to 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%-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 excluded these two studies from pooled estimates because they inappropriately analysed all ratings of the scale, rather than patients assessed (see online Appendix 1 for further details).

Training intensity varied widely from one hour[24] to two days,[28, 31] including one case, where 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 didn´t 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]

HemoCue (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 Hemometer[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.[24, 25, 27-31, 35, 36]

Table 2 (study characteristics and main outcomes)

Methodological quality of included studies

We detected high risk of bias in five[25, 27, 32, 35, 37] and had severe applicability concerns for nine of the 14 studies. In all but in two studies[24, 28] incomplete reporting demanded an “unclear” rating in one or more quality relevant domains (see figure 2 and table 3). See appendix 2 and table 5 in the online supplement for further details.

Figure 2

Table 3 (QUADAS-2 judgements)

Findings

Anaemia

The diagnostic accuracy of the HCS for diagnosing anaemia varied widely across individual studies; sensitivities ranged from 33% to 96%, specificities from 14% to 100% (figure 3).
The meta-analysis from 13 statistically unbiased studies, i.e. excluding those with multiple counts from the same sample[24, 36] showed a higher pooled sensitivity of 80% (95% CI 68%-89%) for the HCS (figure 5) compared with 52% (95% CI 36%-67%) for clinical signs, p=0·008 (figure 4). Pooled specificities were similar at 80% (95% CI 59%-91%) for the HCS and 75% (95% CI 56-88%) for clinical signs, p=0·8395.

Figure 3
Figure 4

When we included only those ten studies that explicitly compared the HCS with clinical signs to diagnose anaemia (median anaemia prevalence: 70%) the pooled results were very similar. Whether we included all studies or excluded studies which 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 which differed from 11g/dl made little difference to the results (See table 4 for sensitivity analysis).

Table 4 (sensitivity analysis)

Severe anaemia

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 7 and 8). In the meta-analysis the HCS again appeared better than clinical signs (p<0.0001) yielding 57% (95% CI 36%-76%) sensitivity compared with 43% (95% CI 9%-85%) by clinical signs. Specificity for the HCS was 99.5% (95% CI 98%-100%), significantly higher than the estimate of 93% (95% CI 56%-99%) for clinical signs (figures 9 and 10), again we saw little differences in the sensitivity analysis (table 4).

Figure 5
Figure 6

HCS combined with clinical diagnosis

In practice, it is likely that primary health care workers would use both the HCS and clinical signs 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 signs for anaemia rose to 92 % (95% CI 83%-97%) 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 60% (95% CI 33%-82%) for simultaneous testing.[38]

For severe anaemia, simultaneous testing would yield a pooled net sensitivity of 90% (95% CI 33%-100%) in the six comparative studies without multiple assessments for the HCS, while the specificity would decrease to 84% (95% CI 40%-98%).

Sources of heterogeneity

Meta-regression analyses could not demonstrate any statistically significant effect of the covariates population group, anaemia prevalence, reference test, training quantity and source of blood sample (table 5; also appendix 3 and figures 1-5 in the online supplement), although this may be due to incomplete reporting, e.g. for training, or small numbers of studies (using appropriate laboratory reference tests).

Table 5 (meta-regression covariates)

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 very comprehensive and no studies were excluded due to language of publication.

We have identified considerable 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 due to the small number of studies or incomplete reporting of key methods. Heterogeneity may 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.

It is unclear whether the use of different blood samples (capillary, venous, umbilical cord) between studies could have been one reason for heterogeneity. Discrepancies between the standard test and the HCS may 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 sub-optimal storage could potentially have damaged blood specimens in four studies[27, 32, 33, 36] where the reference test was carried out 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 relationship between training and accuracy outcomes. However, during HCS development, it was shown that trainee's 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 two hours on two consecutive days. Others have shown inter-rater variations even if adhering to the protocol[39] and some have suggested 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 due to 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]
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 have 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 have 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% patients would miss the correct diagnosis of mild to moderate anaemia. The HCS alone may reduce the number of anaemic patients missing the right diagnosis significantly to 20%. While 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 signs) for the single methods to 92% if the diagnosis of anaemia was considered with either or both methods being positive (severe anaemia: net sensitivity 90%). However, the potential cost of using both methods simultaneously would be a loss of specificity.
The public health relevance is best illustrated by an example: 80% of Malawi´s 15 million people live in rural areas, among these are 6·5 million (m) women, of whom 2·7 m suffer from anaemia (anaemia prevalence 45%). Nearly every second woman, i.e. 1·3 m would miss the correct diagnosis by assessing clinical signs only. The HCS alone would reduce the number of under diagnosed women from 1·3 m to 0·5 m hence 800,000 additional women would receive the appropriate diagnosis and potentially correct care. If use of both clinical signs and HCS was combined, over 1 million additional women would be diagnosed correctly.
Unfortunately, the reduction of under diagnosis diminishes when anaemia becomes severe. In this case the HCS leaves 43% undetected, whilst the assessment of clinical signs leaves 57% undetected. The HCS is able to significantly reduce the number of those falsely diagnosed with severe anaemia (0.4% vs. 7%), 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%. Over-diagnosis 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, while the clinical assessment is purely qualitative. Although the available studies do not allow an inference about the influence of the knowledge of continuous values on clinical decisions, most likely those will be influenced stronger 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, i.e. the impact on clinical outcomes, or its cost-effectiveness.

Conclusion

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 haemoglobin level 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 disconcordant results have to decide whether to rely on their clinical judgement or the HCS. Taking into account the potential clinical and economic consequences of misdiagnosis and in light of the evidence that the HCS yields a significantly better sensitivity but 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 overrule the clinical judgement in most cases, but for severe anaemia a positive HCS might be overruled if clinical signs are missing. It remains to be assessed if a short-term follow up of those patients with disconcordant or borderline results would improve their clinical outcome.

Public health decision makers should be aware that the use of the HCS may require more training and supervision than technically more sophisticated devices.

To tap the full potential of the HCS an evidence-based standardized 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 using the HCS in routine practice.

Declaration of interests

We declare no competing interest

Research in context

Evidence before the study

The WHO Haemoglobin Colour Scale (HCS) became commercially available in 2001 as a tool for health care workers in resource-poor settings, who usually have to base the diagnosis of anaemia on signs and symptoms, to quantitavely 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 13 studies, but most of these were laboratory based with only 4 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 4 „field" studies (sensitivity 76%-88%; specificity 41%-100%). The authors did not compute summary estimates from individual studies, except for the 5 laboratory studies.

Added value of this study

We restricted our new systematic review to "real life" studies (n=14), identifying 10 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, since most studies directly compared these two tests. This is important as clinical assessment which remains 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 were able to perform a 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.

Implication of the totality of evidence

There is sound evidence that the HCS can improve the accuracy to diagnose anaemia and severe anaemia of 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 diagnosing anaemia compared with assessing clinical signs, and the improvement in sensitivity would be clinically important in practice. There remains a lack of evidence concerning how training and supervision may affect the overall performance of the device, as well as its cost-effectiveness in reducing anaemia related mortality and morbidity in practice."

Acknowledgements: We thank Imelda Bates for her comments on a draft of this manuscript and Barbara Butland for assistance with the statistical methods. We thank two authors for their assistance with obtaining additional analyses and data [P F van Rheenen and N R van den Broek].

Funding: We had no specific support for this project. JC is funded by the High Education Funding Council for England. HM did not receive any funding.

Author Contributions: HM and JC conceived of the study, designed the search strategy, data collection instruments and analysis plan and performed analyses. HM wrote the initial draft of the manuscript and JC commented on this draft.

Reference list

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**(2S-2): 676S-88S; discussion 88S-90S.
3. Balarajan Y, Ramakrishnan U, Ozaltin E, Shankar AH, Subramanian SV. Anaemia in low-income and middle-income countries. *Lancet* 2011; **378**(9809): 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 6/21/2015).
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 1/18/2015).

7. Briggs C, Kimber S, Green L. Where are we at with point- of- care testing in haematology? *Br J Haematol* 2012; **158**(6): 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**(1): 21-4.
11. Stott GJ LS. A simple and reliable method for estimating haemoglobin. *Bulletin World Health Organization* 1995; **73**: 369-73.
12. WHO. Haemoglobin Colour Scale: Operational Research Agenda and Study Design2004. http://www.who.int/medical_devices/publications/en/HbCS_Research_AgendaStudyDesign.pdf (accessed 1/18/2015).
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**(12): 987.
15. Cherian M, Emmanuel JC, Lewis SM, et al. Evaluation of the haemoglobin colour scale. *Bull World Health Organ* 2002; **80**(10): 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**(6): 1425-34.
17. Beynon R, Leeflang MM, McDonald S, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *The Cochrane database of systematic reviews* 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**(8): 529-36.
19. World Health Organization. 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 12.11.2012).
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. *Journal of clinical epidemiology* 2005; **58**(10): 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**(2): 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-S.
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**(4): 423-9.
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**(7): 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. *Tropical Medicine and International Health* 2003; **8**(7): 615-8.
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**(3): 101-6.
28. Gies S, Brabin BJ, Yassin MA, Cuevas LE. Comparison of screening methods for anaemia in pregnant women in Awassa, Ethiopia. *Tropical Medicine and International Health* 2003; **8**(4): 301-9.
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**(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**(2): 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**(7): 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**(4): 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**(3): 135-8.
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**(1): 51-7.
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. *The Indian journal of medical research* 2008; **128**(3): 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**(1): 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**(3): 158-61.
38. Gordis L. Assessing the validity and reliability of diagnostic and screening tests. Epidemiology. 4th ed. Philadelphia: Elsevier; 2009: 92-6.
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**(11): 1522-6.
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**(3): 214-21.
41. Sanchis-Gomar F, Cortell-Ballester J, Pareja-Galeano H, Banfi G, Lippi G. Hemoglobin Point-of-Care Testing: The HemoCue System. *Journal of laboratory automation* 2013; **18**(3): 198-205.
42. Nguyen HT. High humidity affects HemoCue cuvette function and HemoCue haemoglobin estimation in tropical Australia. *J Paediatr Child Health* 2002; **38**(4): 427-8.

Table 1

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

	RISK OF BIAS	APPLICABILITY CONCERNS
DOMAIN	**SIGNALING QUESTIONS**	
Patient Selection	Was a consecutive or randomized 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 hour, at most 1 day)? Was the cut-off for anaemia representative for practice (Hb<11g/dl)?
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?	

Table 2: Study characteristics and main outcomes

Study	Population	Sample Size	Study setting	Study design	Reference Standard Test (blood sample)	Operators	Training	Cut-off Anaemia / Severe Anaemia (g/dl)	Prevalence Anaemia/ Severe anaemia	HCS Sensitivity/ Specificity (95%CI)		Clinical signs Sensitivity/ Specificity (95%CI)	
van den Broek 1999	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	A: <11	0·58	0·78	(0·74,0·81)	0·33	(0·29,0·38)
										0·50	(0·46,0·55)	0·84	(0·79,0·88)
								SA: <6	0·006	0·50	(0·12,0·88)	0·67	(0·09,0·99)
										0·98	(0·98,0·99)	0·74	(0·70,0·77)
Montresor 2000	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 supple-mentation 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	A: <11	0·79	0·85	(0·81,0·88)	0·22	(0·18,0·26)
										0·77	(0·69,0·85)	0·92	(0·86,0·96)
								SA: <7	0·04	0·74	(0·49,0·91)	0·63	(0·38,0·84)
										1·00	(0·99,1·00)	0·84	(0·80,0·87)
Barduagni 2003	School-children, mean age 8.4 years (6-11 years)	150	Qena Governorate, Upper Egypt	Prospective diagnostic accuracy study comparing HCS and Sahli´s haemoglobin-meter	HemoCue (capillary)	1 nurse	unclear	A: <12	0·17	0·88	(0·70,0·98)	Not assessed	
										0·49	(0·40,0·58)		
								SA: <7	No cases	No cases			
Montresor 2003	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	HemoCue (capillary)	13 HCW at dispensaries (HCS); different 8 HCW (pallor signs)	2 days	A: <11	0·83	0·95	(0·94,0·96)	0·41	(0·39,0·44)

				study, (only field data used for this review)		HCW (pallor signs)						
									0·14	(0·10,0·19)	0·76	(0·71,0·81)
								SA: <7	0·2	0·82	(0·78,0·87)	0·78	(0·72,0·82)
										0·86	(0·84,0·88)	0·65	(0·62,0·68)
Gies 2003	Pregnant women	403	Urban health centre in Assawa, southern Ethiopia, Rift valley, 1700m 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	A: <11	0·15	0·44	(0·31,0·57)	0·44	(0·32,0·58)
										0·87	(0·83,0·90)	0·79	(0·74,0·83)
								SA: <7	0·003	No data		No data	
Lindblade 2006p	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)	4·5h (1·5h explaining the study and 3h practicing the HCS on 5 specimen with known Hb level)	A: <11	0·52	0·60	(0·55,0·66)	0·67	(0·61,0·72)
										0·94	(0·90,0·96)	0·55	(0·49,0·60)
								SA: <7	0·025	0·44	(0·20,0.70)	1·00	(0·79,1·00)
										1·00	(0·99,1·00)	0·45	(0·41,0·49)
Lindblade 2006c	Children (2-24m)	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)	4·5h (1·5h explaining the study and 3h practicing the HCS on 5 specimen with known Hb level)	A: <11	0·74	0·79	(0·75,0·84)	0·64	(0·59,0·69)
										0·85	(0·77,0·91)	0·59	(0·49,0·68)
								SA: <7	0·10	0·63	(0·47,0·77)	0·88	(0·75,0·96)
										0·97	(0·95,0·99)	0·45	(0·40,0·50)

van Rheenen 2007	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	A: at birth: <12·5g/dl 2 months: <9·5g/dl 4 months: <10·4g/dl	0·12	0·40	(0·23,0·59)	Not assessed
										0·96	(0·92,0·98)	
								SA: No data				
Sinha 2008	Children (6-35 months)	772	67 villages of 3 health centers (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	A: <11	0.80	0.90	(0.87,0.92)	0.67	(0.63,0.70)
										0.97	(0.93,0.99)	0.98	(0.94,1.00)
								SA: <7	0.013	0.00	(0.00,0.31)	0.00	(0.00,0.31)
										1.00	(1.00,1.00)	1.00	(1.00,1.00)
Rusmawatinigtyas 2009	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	A: 11·5	0·12	0·93	(0·68,1·00)	Not assessed	
										1·00	(0·97,1·00)		
								SA: no data					
Bala 2011	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	A: <11	0·70	0·83	(0·74,0·90)	0·91	(0·83,0·96)
										0·33	(0·19,0·50)	0·13	(0·04,0·27)
								SA: <7	0·016	1·00	(0·16,1·00)	1·00	(0·16,1·00)
										0·98	(0·94,1·00)	0·98	(0·93,1·00)
Prathapan 2011	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	A: <11	0·21	0·62	(0·38,0·82)	Not assessed	
										0·86	(0·77,0·93)		
								SA: no data					

								SA: no data					
Chathurani 2012	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	A: <11	0·16	0·50	(0·26,0·74)	0·19	(0·11,0·29)
										0·76	(0·67,0·84)	0·88	(0·85,0·91)
								SA: <7	No cases				
Aldridge 2012	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	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 lab technician,1 psychiatric nurse)	1 hour	A: <11	0.71	0·33	(0·29,0·36)	0·58	(0·54,0·62)
										0·87	(0·83,0·91)	0·55	(0·48,0·61)
								SA: <5	0·0067	0·14	(0·00,0·58)	0·00	(0·00,0·46)
										1·00	(0·99,1·00)	1·00	(0·99,1·00)
Shah 2014	Women of reproductive age (15-45 years)	501	8 villages of Jhagadia block located in Guajarat, 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	1/2 day	A: <12	0·71	0·96	(0·94,0·98)	Not assessed	
										0·22	(0·15,0·29)		
								SA: <7	0·024	0·83	(0·52,0·98)		
										0·99	(0·98,1·00)		

Abbreviations: 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 3

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

☺ = low ☹ = high ? = unclear	RISK OF BIAS							APPLICABILITY CONCERNS				
	Patient Selection		Index Test			Reference Test	Flow And Timing	Patient Selection	Index Test			Reference Test
	Randomi-zation or consecutive cases	No in-appropriate exclusions	WHO certified HCS[1]	Blinding HCS vs. reference test	Reliability of HCS readings	Test likely to correctly diagnose anaemia	Concurrent sampling of HCS and reference test	Included patients match the target population	HCS operator matches review objective	Training intensity at least 1 hour, at most 1 day	WHO according cut-off (11 g/dl)	Test allows assessment of HCS accuracy
van den Broek 1999	?	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺
Montresor 2000	?	☺	☺	☺	☺	☺	☺	☺	☹	☺	☺	☺
Barduagni 2003	☺	☺	☺	☺	?	☺	☺	☺	☺	?	☹	☺
Montresor 2003	?	☺	☺	☺	☺	☺	☺	☺	☺	☹	☺	☺
Gies 2003	☺	☺	☺	☺	☺	☺	☺	☺	☺	☹	☺	☺
Lindblade 2006c	☺	☺	☺	?	☺	☺	☺	☺	☺	☺	☺	☺
Lindblade 2006p	☺	☺	☺	?	☺	☺	☺	☺	☺	☺	☺	☺
van Rheenen 2007	☺	☹	☺	☺	?	☺	☺	☹	☹	?	☹	☺
Sinha 2008	☺	☺	☺	☺	?	☹	☺	☺	☹	?	☺	☹
Rusmawatiningtyas 2009	☺	?	☺	☺	☺	☺	☺	☺	☹	☹	☹	☺
Bala 2012	☺	☺	☹[1]	☺	☺	☹	☺	☺	☺	?	☺	☹
Prathapan 2011	☺	☹	?	☺	?	☺	☺	☺	☺	?	☺	☺
Chathurani 2012	☹	☹	?	☺	?	☺	☹	☺	☺	?	☺	☺
Aldridge 2012	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺
Shah 2014	☺	☺	☺	☺	?	☺	☺	☺	☺	☺	☹	☺

[1]The Haemoglobin Colour Scale from the Indian manufacturer Kruise Path is not certified by the original German manufacturer Copack GmbH, see appendix 2 of the online supplement for more details

Table 4: Sensitivity analysis of pooled estimates for HCS/Clinical assessment accuracy

Subgroup of studies		Anaemia		Severe anaemia	
		HCS	Clinical signs	HCS	Clinical signs
All studies	No. of participants (studies)	7245 (15)	6413 (10)	6663 (9)	5476 (8)
	Prevalence median (range)	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)	0·78 (0·65,0·87)	0·52 (0·36,0·67)	0·54 (0·36,0·71)	0·43 (0·09,0·85)
	Specificity (95% CI)	0·79 (0·61,0·90)	0·75 (0·56,0·88)	0·995 (0·98,0·999)	0·93 (0·56,0·99)
	PV+	0·84	0·83	0·73	0·11
	PV-	0·72	0·40	0·99	0·99
All studies without multiple HCS testing[1]	No. of participants (studies)	5813 (13)	6413 (10)	4547 (7)	5476 (8)
	Prevalence median (range)	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)	0·80 (0·68,0·88)	0·52 (0·36,0·67)	0·57 (0·36,0·76)	0·45 (0·12,0·83)
	Specificity (95% CI)	0·80 (0·59,0·91)	0·75 (0·56,0·88)	0·996 (0·95,0·999)	0·92 (0·62,0·99)
	PV+	0·81	0·83	0·79	0·11
	PV-	0·79	0·40	0·99	0·99
Studies without high risk of bias (including multiple HCS testing)[2]	No. of participants (studies)	4322 (8)	4977 (7)	3646 (5)	4575 (6)
	Prevalence median (range)	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)	0·84 (0·70,0·92)	0·46 (0·34,0·58)	0·68 (0·55,0·79)	0·62 (0·22,0·90)
	Specificity (95% CI)	0·76 (0·43,0·93)	0·74 (0·61,0·83)	0·99 (0·95,0·998)	0·80 (0·46,0·95)
	PV+	0·85	0·81	0·72	0·08
	PV-	0·74	0·36	0·99	0·99
All comparative studies (HCS vs. clinical signs)[3]	No. of participants (studies)	6680 (10)	6413 (10)	6162 (8)	5476 (8)
	Prevalence median (range)	0·70 (0·15-0·83)	0·70 (0·15-0·83)	0·02 (0·006-0·2)	0·02 (0·006-0·2)
	Sensitivity (95% CI)	0·75 (0·60,0·85)	0·52 (0·36,0·67)	0·50 (0·30,0·70)	0·45 (0·12,0·83)
	Specificity (95% CI)	0·76 (0·56,0·89)	0·75 (0·56,0·88)	0·995 (0·97,0·999)	0·92 (0·62,0·99)
	PV+	0·88	0·83	0·68	0·11
	PV-	0·57	0·40	0·99	0·99
All studies with common threshold for anaemia (<11g/dl) and severe anaemia (Hb < 7 g/dl)[4]	No. of participants (studies)	6781 (11)	6413 (10)	4547 (7)	4045 (6)
	Prevalence median (range)	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)	0·74 (0·60,0·84)	0·52 (0·36,0·67)	0·57 (0·36,0·76)	0·54 (0·16,0·88)
	Specificity (95% CI)	0·77 (0·59,0·89)	0·75 (0·56,0·88)	0·996 (0·95,0·999)	0·91 (0·44,0·99)
	PV+	0·88	0·83	0·79	0·16
	PV-	0·56	0·40	0·99	0·99

[1]

PV+	0·88	0·83	0·79	0·16
PV-	0·56	0·40	0·99	0·99

[1]Aldridge (2012) and van den Broek (1999) 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 in the manuscript since it only includes statistically unbiased studies.

[2]van Rheenen 2007, Sinha 2008, Bala 2012, Prataphan 2011 and Chathurani 2012 were excluded for high risk of bias. See appendix 2 of the online supplement for details.

[3]Barduagni 2003, van Rheenen 2007, Rusmawatinigtyas 2009, Prataphan 2011 and Shah 2014 did not assess anaemia by clinical signs

[4]The following studies used thresholds different from the WHO recommendations in school-age children and pregnant women for the diagnosis of anaemia (<11 g/dl): Barduagni 2003 (<12 g/dl), van Rheenen (different age-specific thresholds for new-born), Rusmawatinigtyas 2009 (<11.5 g/dl), and severe anaemia (<7 g/dl): van den Broek (<6 g/dl), Aldridge 2012 (5 g/dl). Shah 2014 also tested non-pregnant women (<12 g/dl).

Table 5: meta-regression analysis of the effect of covariates on HCS accuracy

Potential sources of heterogeneity	Sub-groups	Sensitivity	Specificity	Evidence of statistical difference (p-value)
Prevalence of anaemia	Very high (≥40%)	0·79	0·69	0·3068
	Low to moderate (<40%)	0·68	0·90	
Population	children	0·83	0·93	0·3153
	women	0·66	0·68	
Training[1]	High (≥ 1/2day)	0·78	0·74	0·8091
	Low (<1/2day)	0·74	0·83	
Reference test	Point-of-care test[2]	0·78	0·76	0·5897
	Laboratory test[3]	0·66	0·88	
Blood sample[4]	same	0.79	0.84	0.2721
	different	0.66	0.67	

[1]we performed the meta-regression analysis for training under the assumption that those studies without information on training had less than half day (low) of training

[2]point-of-care tests included: HemoCue, Filter Paper Cyanmethaemoglobin method and Sahli´s hemometer

[3]laboratory tests were performed in clinical laboratories, included: Electronic coulter counter, Hematology Analyzer (HmX), Spectrometry method and laboratory based Cyanmethaemoglobin Method

[4]blood samples for the HCS and the reference test had either the same origin (capillary, venous or umbilical chord) or different sources (e.g. capillary vs. venous). In those cases where it was unclear whether the origin was the same, we assumed that the sources of the blood sample were different.

Study Selection Flow Diagram

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

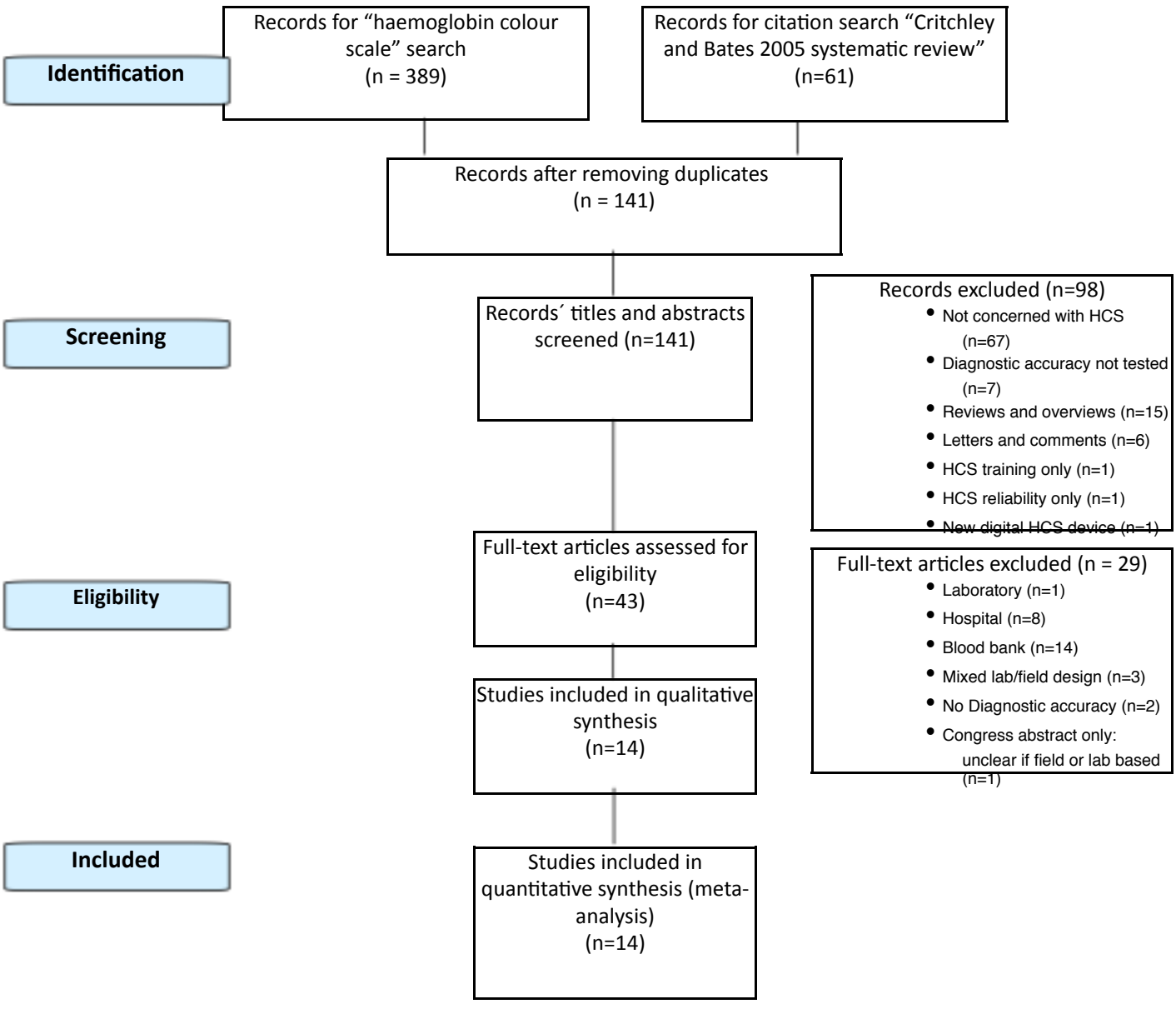


Figure 2

Click here to download high resolution image

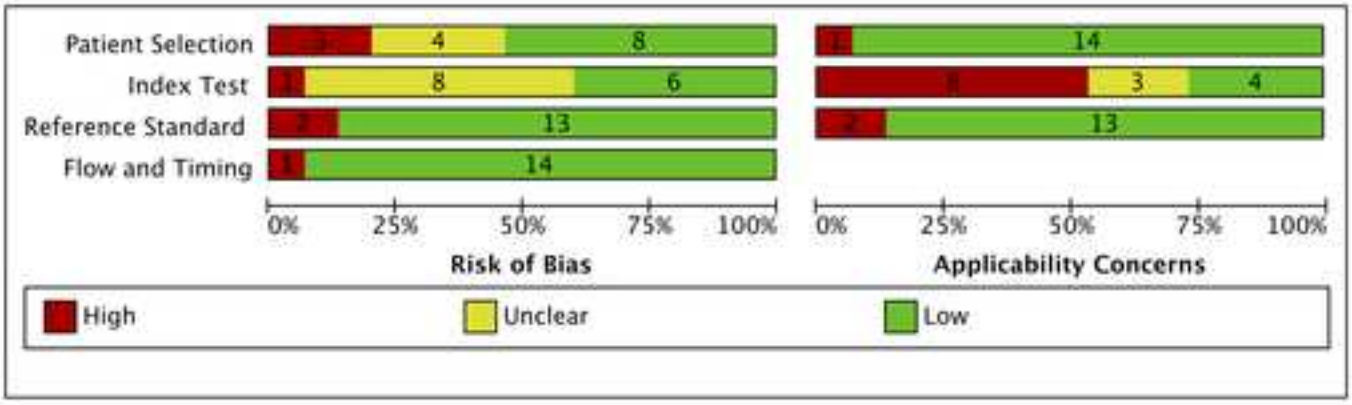
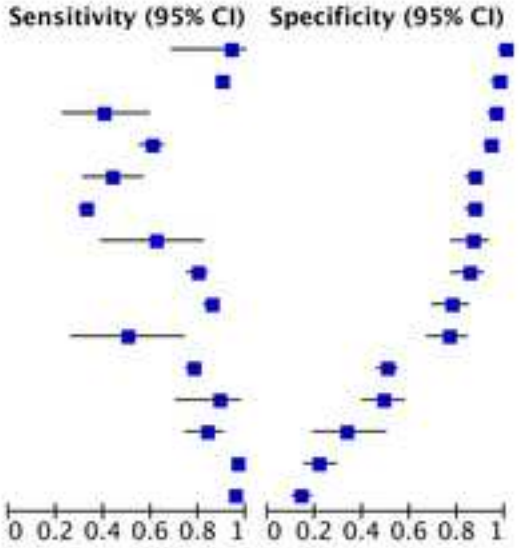


Figure 3

Click here to download high resolution image

HCS for the diagnosis of anaemia

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Rusmawatiningtyas 2009	14	0	1	109	11.5	0.93 [0.68, 1.00]	1.00 [0.97, 1.00]
Sinha 2008	555	4	65	148	11.0	0.90 [0.87, 0.92]	0.97 [0.93, 0.99]
van Rheenen 2007	12	9	18	211		0.40 [0.23, 0.59]	0.96 [0.92, 0.98]
Lindblade 2006p	203	19	134	287	11.0	0.60 [0.55, 0.66]	0.94 [0.90, 0.96]
Gies 2003	27	44	35	297	11.0	0.44 [0.31, 0.57]	0.87 [0.83, 0.90]
Aldridge 2012	243	40	500	267	11.0	0.33 [0.29, 0.36]	0.87 [0.83, 0.91]
Prathapan 2011	13	11	8	69	11.0	0.62 [0.38, 0.82]	0.86 [0.77, 0.93]
Lindblade 2006c	259	17	67	95	11.0	0.79 [0.75, 0.84]	0.85 [0.77, 0.91]
Montresor 2000	358	26	62	89	11.0	0.85 [0.81, 0.88]	0.77 [0.69, 0.85]
Chathurani 2012	9	23	9	74	11.0	0.50 [0.26, 0.74]	0.76 [0.67, 0.84]
van den Broek 1999	479	223	138	226	11.0	0.78 [0.74, 0.81]	0.50 [0.46, 0.55]
Barduagni 2003	23	63	3	60	12.0	0.88 [0.70, 0.98]	0.49 [0.40, 0.58]
Bala 2012	75	26	15	13	11.0	0.83 [0.74, 0.90]	0.33 [0.19, 0.50]
Shah 2014	344	112	14	31	12.0	0.96 [0.94, 0.98]	0.22 [0.15, 0.29]
Montresor 2003	1205	224	63	37	11.0	0.95 [0.94, 0.96]	0.14 [0.10, 0.19]



Clinical signs for the diagnosis of anaemia

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Sinha 2008	414	3	206	149	11.0	0.67 [0.63, 0.70]	0.98 [0.94, 1.00]
Montresor 2000	92	9	328	106	11.0	0.22 [0.18, 0.26]	0.92 [0.86, 0.96]
Chathurani 2012	16	52	69	398	11.0	0.19 [0.11, 0.29]	0.88 [0.85, 0.91]
van den Broek 1999	126	42	252	223	11.0	0.33 [0.29, 0.38]	0.84 [0.79, 0.88]
Gies 2003	27	71	34	270	11.0	0.44 [0.32, 0.58]	0.79 [0.74, 0.83]
Montresor 2003	523	62	745	199	11.0	0.41 [0.39, 0.44]	0.76 [0.71, 0.81]
Lindblade 2006c	209	46	117	66	11.0	0.64 [0.59, 0.69]	0.59 [0.49, 0.68]
Aldridge 2012	325	104	234	125	11.0	0.58 [0.54, 0.62]	0.55 [0.48, 0.61]
Lindblade 2006p	224	139	112	167	11.0	0.67 [0.61, 0.72]	0.55 [0.49, 0.60]
Bala 2012	82	34	8	5	11.0	0.91 [0.83, 0.96]	0.13 [0.04, 0.27]

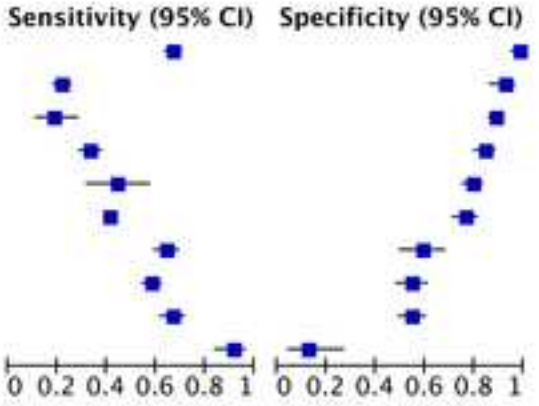
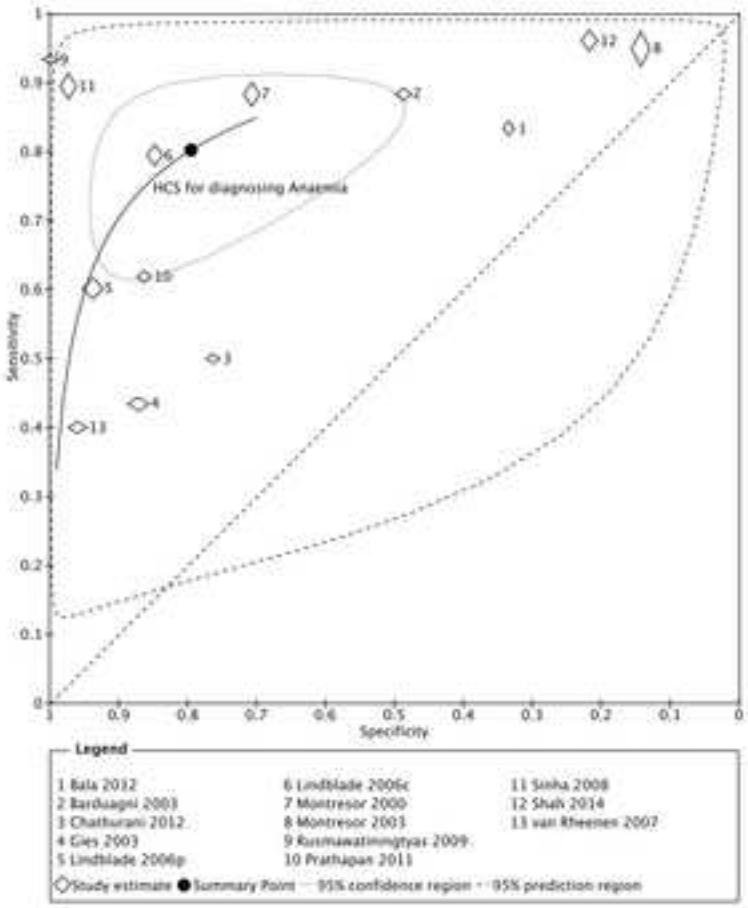


Figure 4 a b

Click here to download high resolution image



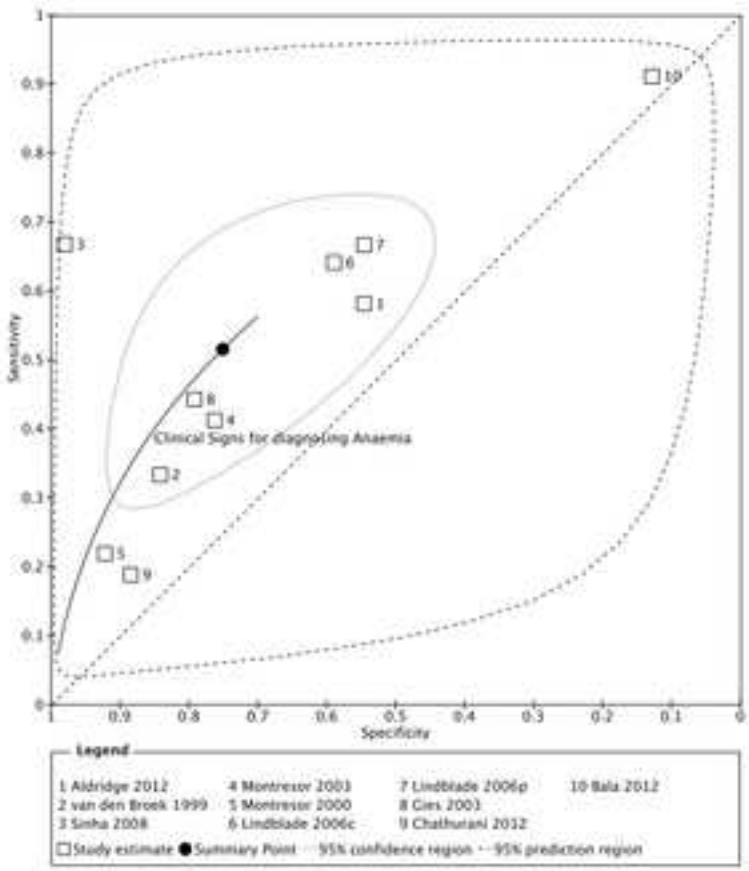
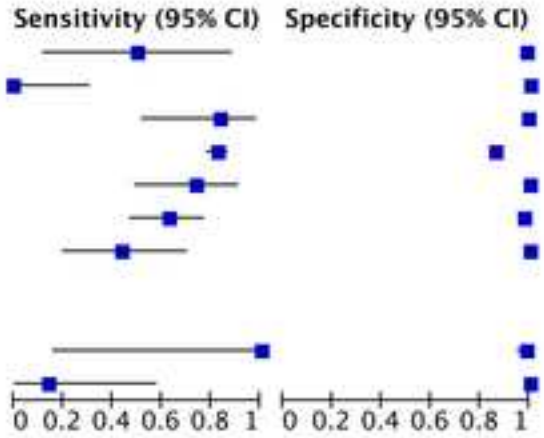


Figure 5 a b

Click here to download high resolution image

HCS for the diagnosis of severe anaemia

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
van den Broek 1999	3	16	3	1044	6.0	0.50 [0.12, 0.88]	0.98 [0.98, 0.99]
Sinha 2008	0	0	10	762	7.0	0.00 [0.00, 0.31]	1.00 [1.00, 1.00]
Shah 2014	10	4	2	485	7.0	0.83 [0.52, 0.98]	0.99 [0.98, 1.00]
Montresor 2003	253	172	54	1050	7.0	0.82 [0.78, 0.87]	0.86 [0.84, 0.88]
Montresor 2000	14	1	5	515	7.0	0.74 [0.49, 0.91]	1.00 [0.99, 1.00]
Lindblade 2006c	27	10	16	385	7.0	0.63 [0.47, 0.77]	0.97 [0.95, 0.99]
Lindblade 2006p	7	1	9	626	7.0	0.44 [0.20, 0.70]	1.00 [0.99, 1.00]
Chathurani 2012	0	0	0	0	7.0	Not estimable	Not estimable
Barduagni 2003	0	0	0	0	7.0	Not estimable	Not estimable
Bala 2012	2	2	0	125	7.0	1.00 [0.16, 1.00]	0.98 [0.94, 1.00]
Aldridge 2012	1	1	6	1042	5.0	0.14 [0.00, 0.58]	1.00 [0.99, 1.00]



Clinical signs for the diagnosis of severe anaemia

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
van den Broek 1999	2	166	1	474	6.0	0.67 [0.09, 0.99]	0.74 [0.70, 0.77]
Sinha 2008	0	0	10	762	7.0	0.00 [0.00, 0.31]	1.00 [1.00, 1.00]
Montresor 2003	238	428	69	794	7.0	0.78 [0.72, 0.82]	0.65 [0.62, 0.68]
Montresor 2000	12	83	7	433	7.0	0.63 [0.38, 0.84]	0.84 [0.80, 0.87]
Lindblade 2006c	38	217	5	178	7.0	0.88 [0.75, 0.96]	0.45 [0.40, 0.50]
Lindblade 2006p	16	347	0	279	7.0	1.00 [0.79, 1.00]	0.45 [0.41, 0.49]
Bala 2012	2	3	0	124	7.0	1.00 [0.16, 1.00]	0.98 [0.93, 1.00]
Aldridge 2012	0	3	6	779	5.0	0.00 [0.00, 0.46]	1.00 [0.99, 1.00]

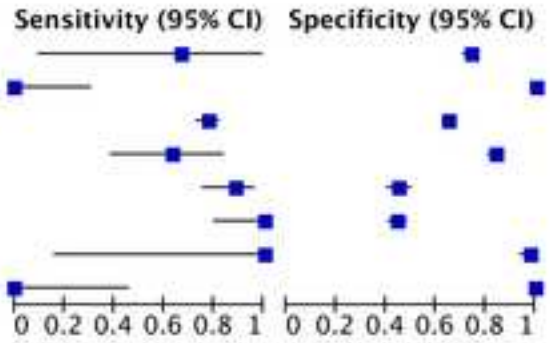


Figure 6 a b

Click here to download high resolution image