Article Type: Review Article

Subheading: Obstetrics

Accuracy of on-site tests to detect anemia during prenatal care

Soha Sobhy [1,*], Ewelina Rogozinska [1,2], Khalid S Khan [1,2]

[1] Women's Health Research Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[2] Multidisciplinary Evidence Synthesis Hub, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

* Correspondence

Soha Sobhy, Women's Health Research Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

Email: soha.sobhy@qmul.ac.uk

Keywords: Anemia; On-site tests; Point-of-care tests; Pregnancy; Prenatal care; Screening

Synopsis: The copper sulfate, Sahli method, and HemoCue tests were found to be accurate tools for the detection of anemia during pregnancy.

Abstract

Background: Anemia is a substantial contributor to poor pregnancy outcomes in low- and middle-income countries. Access to laboratory facilities is limited; therefore, on-site testing warrants attention.

Objectives: To determine the accuracy of on-site tests to detect anemia in pregnancy.

Search strategy: MEDLINE, Embase, Scopus, CINAHL, and Web of Science were searched from inception until March 2016, with no language restrictions, using the terms “pregnancy,” “an(a)emia,” and “h(a)emoglobin.”

Selection criteria: Studies that evaluated the diagnostic accuracy of on-site hemoglobin tests versus laboratory-based reference tests during pregnancy were included.

Data collection and analysis: Study characteristics and true positive, true negative, false positive, and false negative rates were extracted. Sensitivity, specificity, likelihood ratios, and post-test probabilities were calculated. Anemia was defined as a hemoglobin level of less than 110 g/L.

Main results: Ten studies (4239 participants) were assessed. Copper sulfate provided 97% sensitivity (95% confidence interval [CI] 88%–100%) and 71% specificity (95% CI 55%–85%); the Sahli method provided 86% sensitivity (95% CI 75%–94%) and 83% specificity (95% CI 68%–93%); and HemoCue provided 85% sensitivity (95% CI 79%–90%) and 80% specificity (95% CI 76%–83%).

Conclusions: Some on-site tests are accurate and should be made widely available to improve detection of anemia in pregnancy.

1 INTRODUCTION

Maternal anemia—defined as a low blood hemoglobin concentration—remains a serious global health problem. This condition is detected among 42% of all women during pregnancy; Africa and Southeast Asia are the most affected geographic regions [1,2]. An estimated 56 million pregnant women have anemia; furthermore, the rate of this condition has declined by only 5% since 1995 [1].

Anemia is the most frequent indirect cause of adverse maternal outcomes and mortality, contributing to up to 50% of both [3]. The risks of macerated late fetal death, fresh late fetal death, and early neonatal death are consistently increased among mothers with severe anemia [4]. A link has also been found between anemia and low birth weight, small for gestational age, and preterm birth [2,5]. Furthermore, women with anemia are less likely to overcome the adverse effects of excessive blood loss and are more susceptible to infection, fatigue, and depression than are women without this disorder [6].

Testing for anemia is done infrequently and with low accuracy in low-income and middle-income countries (LIMICs), especially in rural locations. In this setting, anemia can go undetected because diagnosis is often based on symptoms and physical examination rather than on objective testing [7]. Laboratory measurement of cyanmethemoglobin by spectrometry is the gold standard, but use of this test is hampered by high cost and low availability [8]. Consequently, simple, safe, accurate, and low-cost hemoglobin assessment tools were introduced to address this issue [9]. On-site testing to detect anemia is a potential option to increase accessibility. Methods currently in use include HemoCue (HemoCue, Ängelholm, Sweden), WHO

hemoglobin color scale (HCS), and copper sulfate; however, their diagnostic accuracy is not well characterized.

Systematic reviews have been published on the accuracy of HCS, HemoCue, and non-invasive methods but they did not include studies exclusively conducted in a prenatal care setting [10,11]. The threshold used to diagnose anemia in pregnancy differs from that of a nonpregnant population [12]. Previous reviews have included a mixed population, which is a key deficiency because disease spectrum variation has an impact on accuracy [13,14]. Reviews have also used HemoCue as a reference test [11] because it is more widely available than the gold-standard laboratory test in low-income countries. Nonetheless, measurement of hemoglobin concentration by HemoCue provides lower precision than that of automated analyzers, and its precision also varies by sample type (i.e. whether the sample is taken from a capillary or vein) [15–17].

Therefore, the aim of the current systematic review was to determine the accuracy of on-site tests to detect anemia among pregnant women in LIMICs.

2 MATERIALS AND METHODS

A systematic review of on-site testing for maternal anemia was conducted in line with current recommendations. The protocol was registered with the PROSPERO international prospective register of systematic reviews (CRD42015029172), and methods of data reporting adhered to the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines [18].

2.1 Literature search

The MEDLINE, Embase, Scopus, CINAHL, and Web of Science databases were searched for relevant studies from inception until March 21, 2016. No language restrictions were imposed on the literature search. Medical subject headings, text words, and word variants for "pregnancy" were combined with the terms "an(a)emia" and "h(a)emoglobin," with exclusion (“NOT”) of words relating to the fetus such as “cord,” “neonatal,” and “placenta” (Appendix S1). The reference lists of the included studies and relevant reviews were searched for additional eligible studies.

2.2 Study selection

A two-stage process was used to select studies for inclusion in the present systematic review. First, the titles and abstracts of all citations were screened for potential relevance. Second, SS and ER independently examined the full text of the retrieved papers using prespecified inclusion criteria; discrepancies were resolved by discussion with KSK. The inclusion criteria were recruitment of pregnant women attending prenatal care, comparison of any on-site test for hemoglobin concentration with a laboratory-based reference test, and provision of diagnostic accuracy data. Case–control studies, prevalence studies, and studies that enrolled non-pregnant populations were excluded.

2.3 Data extraction and quality assessment

Data extraction and assessment of study quality were performed by SS and ER, with discrepancies resolved following input from KSK. Data were extracted using a standardised pre-piloted form. Information on the type of test, setting, sample (venous or capillary blood), and the type of practitioner who conducted the tests was

included. Data were extracted for different thresholds of hemoglobin concentration as per WHO standards [12]. Anemia in pregnancy was defined as a hemoglobin concentration below 110 g/L at sea level and categorized as mild (100–109 g/L), moderate (70–99 g/L), or severe (<70 g/L). True positive, true negative, false positive, and false negative rates were extracted for the reference and index tests in a 2×2 contingency table. The parameters required to assess accuracy were recalculated in cases for which the sensitivity, specificity, positive predictive values, and negative predictive values were presented.

Study quality was assessed independently by SS and ER using the QUADAS-2 tool [14]. Four domains were evaluated: patient selection, use of the index test, implementation of the reference standard, and timing of the index test(s) and reference standard ("flow and timing"). Risk of bias (low, high, or unclear) and concerns about applicability of the findings were assessed for these domains. A high-quality study was one that had consecutive or randomized patient selection, had specific inclusion and exclusion criteria, and used recognized index and reference (gold-standard) tests that were performed with an appropriate delay between them.

2.4 Data analysis

The data were analyzed using STATA version 12.1 (Stata, College Station, TX, USA). True positive, false positive, true negative, and false negative rates were obtained from the publication or, if necessary, calculated from the reported estimates. The data were plotted in the receiver operating characteristic (ROC) space. When a sufficient number of studies was available, the diagnostic accuracy parameters were

pooled using a bivariate random-effect hierarchical model [19]. If fewer than four studies reported accuracy of a given test, the sensitivity, specificity, and likelihood ratio were pooled using a univariate model [20]. The 95% confidence interval (CI) was calculated for all estimates to indicate precision of test accuracy. Assuming a pre-test probability of 42% for a hemoglobin concentration of less than 110g/L [2,21], post-test probabilities for a negative and positive test result were calculated using an online calculator [22]. A pretest probability of 8% and 1% was used for moderate and severe anemia, respectively [21].

3 RESULTS

The selection process for the 10 studies included in the present systematic review is outlined in Figure 1. As shown in Table 1, these studies included 4239 pregnant women from seven LIMICs: India (n=2), South Africa (n=2), Sri Lanka (n=2), Benin (n=1), Kenya (n=1), Malawi (n=1), and Pakistan (n=1). Six index tests were identified and compared with a laboratory reference test. Five studies reported data on clinical examination [25–27,29,30], four on HCS [25,28–30], and three on the copper sulfate test [23,24,31]. HemoCue [25], the Sahli method [31], and a non-invasive hemoglobin sensor (NBM 2000) [32] were each reported in one study. The studies were published between 1996 and 2016. The prevalence of anemia (hemoglobin <110 g/L [12]) across the 10 studies was 23%–77%.

Quality assessment indicated that five (50%) of the studies had low risk of bias and low concern regarding applicability across all domains (Table 2). One (10%) study had a high risk of bias regarding patient selection. Three (30%) studies had an unclear risk of bias because the recruitment process was not adequately described.

One (10%) study had an unclear risk of bias regarding flow and timing. All studies had low risk of bias for applicability.

Accuracy of the on-site tests is outlined in Table 3. Seven studies provided diagnostic accuracy data for mild anemia (hemoglobin <110 g/L) [25,27–32]. Sensitivity ranged from 97% (copper sulfate) to 34% (NBM 2000). By contrast, specificity ranged from 92% (NBM 2000) to 62% (clinical assessment). Three studies also provided diagnostic accuracy data for mild anemia defined as hemoglobin <100 g/L [23–25]. Copper sulfate had 96% sensitivity (95% CI 89%–100%) and 89% specificity (95% CI 72%–99%); HemoCue had 94% sensitivity (95% CI 90%–97%) and 79% specificity (95% CI 75%–83%); HCS had 82% sensitivity (95% CI 76%–86%) and 45% specificity (95% CI 41%–50%); and clinical assessment had 73% sensitivity (95% CI 70%–76%) and 80% specificity (95% CI 77%–84%).

A general overview of the ROC data is shown in Figure 2, whereas ROC data for clinical assessment and HCS are shown in Figures S1 and S2, respectively.

Table S1 shows the data for moderate anemia (hemoglobin <80 g/L). One study [25] provided diagnostic accuracy data for three tests: clinical assessment (644 women), HCS (641 women), and HemoCue (671 women). Another study [31] provided data for copper sulfate (100 women) and the Sahli method (100 women) with a cutoff hemoglobin level of less than 88 g/L. HemoCue had 97% sensitivity (95% CI 85%–99%) and 94% specificity (95% CI 92%–96%); the Sahli method had 83% sensitivity (95% CI 85%–98%) and 93% specificity (95% CI 85%–98%); HCS had 82% sensitivity (95% CI 61%–93%) and 76% specificity (95% CI 73%–79%); copper

sulfate had 75% sensitivity (95% CI 55%–89%) and 94% specificity (95% CI 86%–98%); and clinical assessment had 62% sensitivity (95% CI 44%–77%) and 76% specificity (95% CI 72%–78%).

Data for severe anemia are also presented in Table S1. One study [25] defined severe anemia as a hemoglobin level of less than 60 g/L and provided accuracy data for three tests: clinical assessment (644 women), HCS (641 women), and HemoCue (671 women). Another study [31] defined severe anemia as a hemoglobin level of less than 66 g/L and provided accuracy data for copper sulfate (100 women) and the Sahli method (100 women). HemoCue had 83% sensitivity (95% CI 44%–97%) and 99% specificity (95% CI 98%–100%); the Sahli method had 75% sensitivity (95% CI 35%–97%) and 93% specificity (95% CI 86%–98%); copper sulfate had 63% sensitivity (95% CI 24%–91%) and 93% specificity (95% CI 86%–98%); HCS had 50% sensitivity (95% CI 15%–85%) and 98% specificity (95% CI 97%–99%); and clinical assessment had 50% sensitivity (95% CI 15%–85%) and 75% specificity (95% CI 71%–78%).

4 DISCUSSION

The present review assessed the accuracy of on-site tests to detect anemia in pregnancy. Copper sulfate, the Sahli method, and HemoCue had the highest sensitivities (≥85%). Furthermore, each of these methods performed better than did clinical assessment alone. Strengths of the present review included use of a systematic protocol. The analysis included studies of predominantly high quality that covered a wide range of tests and enrolled women with a range of anemia severity, which allowed generalization across the spectrum of the disease.

Cyanmethemoglobin spectrophotometry is the global reference standard for detecting hemoglobin concentration [8]. A limitation of the present review was that it included studies that used any laboratory test as a reference standard, such as automated cell counters. Although these machines are calibrated against the reference standard, this calibration could have affected the accuracy of the test.

A univariate model was used to pool sensitivity and specificity estimates when fewer than four studies were available to compare the accuracy of the identified tests. Unlike a bivariate model, this approach did not account for the correlation between variables; however, estimates obtained using both models are reported to be similar [33]. Other limitations include the fact that only a small number of women had severe anemia; therefore, the accuracy of the data at this level might be unreliable.

Post-test probabilities for all evaluated tests were calculated assuming a 42% prevalence for anemia (hemoglobin <110g/L) [2]. After the test, among women who tested negative using the copper sulfate method, the prevalence of women who did not have anemia was reduced to 4%. Using the Sahli method or HemoCue, this probability was reduced to 11% and 12%, respectively. The HCS method performed poorly and only marginally better (reduction to 27%) than clinical examination alone (reduction to 29%). Reviews and studies have recommended the use of the HCS [11,30]; however, in the present review, this test had a high likelihood ratio for a negative test result, suggesting it might not be as clinically useful as previously reported. Additional training could increase its accuracy in a clinical setting because HCS performs well under laboratory conditions [34]. The non-invasive hemoglobin

probe NMB 2000 had low sensitivity but high specificity, which suggested that this test could be indicative of the presence of anemia in the context of screening.

In LIMICs, where resources are limited, many factors prevent the use of technologically advanced equipment, especially at a primary healthcare level [35]. Healthcare workers, therefore, need a simple, cheap, and robust device for measuring hemoglobin concentration [35]; nonetheless, this approach should not be at the expense of accuracy. Choosing the best test to use depends on factors such as cost, resources required, the setting in which it will be used, the skills of the healthcare workers, and ease of use. Therefore, the choice should probably be made at the local level. The HemoCue method is the most expensive test in terms of raw materials; however, the cost of training, procurement, and maintenance must also be considered. Taking the cost of the test in a resource-poor setting into account in combination with accuracy is also important.

A test should have high sensitivity if screening for anemia is used to decide which patients should be referred for further investigation and treatment. It is important not to miss women with severe anemia because this condition can have substantial clinical implications for both the mother and her offspring. Anemia has been linked with maternal mortality, increased risk of postpartum hemorrhage, sepsis, risk of intrauterine growth restriction, and preterm delivery [3,5]. The first step to reducing these risks would be identification of anemia so that the necessary treatment and monitoring can be initiated.

Increasing test sensitivity usually results in a concomitant reduction in specificity [36]. Using a test with high sensitivity but low specificity leads to unnecessary referrals for further investigation and so adds to costs for the health service [26]. However, clinical implications of test results also play a part. The consequence of missing a truly anemic pregnant woman has more severe clinical implications than does overtreatment. Most anemia cases that occur in pregnancy are secondary to iron deficiency; mild-to-moderate cases can be treated with iron tablets, which tend to be well tolerated [37]. For severe anemia, treatment-associated risk is increased because it might involve complex interventions such as blood transfusions. The adverse effects of these interventions can be serious (e.g. the spread of blood-borne disease) [38]. Testing is also important to monitor the effects of any intervention after the diagnosis is established, as well as for planning overall management of the affected woman, including place of delivery.

Additional research must be done on this topic: there were few studies available for inclusion in the present systematic review. It would also be interesting to evaluate novel non-invasive tests, especially those using readily available technology such as mobile phones [39]. Many studies have assessed devices for the detection of anemia; however, these investigations reported correlation coefficients that might be misleading and are less clinically useful than sensitivity and specificity. Reporting different measures can also limit the ability to perform meta-analysis. Future primary studies should be encouraged to report clinically useful accuracy variables. Furthermore, a detailed cost analysis of the accurate tests identified in the present study would be beneficial.

In conclusion, current reliance on clinical examination in LIMICs should be reduced, with the availability of accurate on-site tests increased, to help to improve the detection of anemia during pregnancy.

Author contributions

SS, ER and KSK conceived the research question, designed the protocol, and were involved in the literature search, study selection, and data extraction. ER performed the statistical analysis. SS and ER created the tables, figures, and appendices. SS ER, and KSK contributed to the both the draft and final versions of the manuscript.

Acknowledgments

The present study was supported by Ammalife (registered UK charity 1120236; funded SS), ELLY Appeal (funded SS), and WHO (funded ER). The listed charities had no influence on the development, conduct, or data reporting of the current systematic review.

Conflicts of interest

The authors have no conflicts of interest.

References

[1] New S, Wirth M: Anaemia, pregnancy, and maternal mortality: the problem with globally standardised haemoglobin cutoffs. *BJOG : an international journal of obstetrics and gynaecology* 2015;122(2): 166-169.

[2] Rahman MM, Abe SK, Rahman MS, Kanda M, Narita S, Bilano V, et al.: Maternal anemia and risk of adverse birth and health outcomes in low- and middle-income

countries: systematic review and meta-analysis. *The American journal of clinical nutrition* 2016;103(2): 495-504.

[3] Lumbiganon P, Laopaiboon M, Intarut N, Vogel JP, Souza JP, Gulmezoglu AM, et al.: Indirect causes of severe adverse maternal outcomes: a secondary analysis of the WHO Multicountry Survey on Maternal and Newborn Health. *BJOG : an international journal of obstetrics and gynaecology* 2014;121 Suppl 1: 32-39.

[4] Vogel JP, Souza JP, Mori R, Morisaki N, Lumbiganon P, Laopaiboon M, et al.: Maternal complications and perinatal mortality: findings of the World Health Organization Multicountry Survey on Maternal and Newborn Health. *BJOG : an international journal of obstetrics and gynaecology* 2014;121 Suppl 1: 76-88.

[5] Sukrat B, Wilasrusmee C, Siribumrungwong B, McEvoy M, Okascharoen C, Attia J, et al.: Hemoglobin concentration and pregnancy outcomes: a systematic review and meta-analysis. *Biomed Res Int* 2013;2013: 769057.

[6] Sanghvi TG HP, Wainwright E.: Maternal iron-folic acid supplementation programs: evidence of impact and implementation. *Food Nutr Bull* 2010;31(2 suppl): S100-107.

[7] 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.

[8] Davis BH, Jungerius B, International Council for the Standardization of H: International Council for Standardization in Haematology technical report 1-2009: new reference material for haemiglobincyanide for use in standardization of blood haemoglobin measurements. *Int J Lab Hematol* 2010;32(2): 139-141.

[9] Roberto Musi MLR: Point of Care Testing and Transfusion Safety in Resource Limited Settings: A Review. *Journal of Blood Disorders & Transfusion* 2015;06(02).

[10] Hiscock R KD, Simmons S: Systematic review and meta-analysis of method comparison studies of Masimo pulse co-oximeters (Radical-7™ or Pronto-7™) and HemoCue® absorption spectrometers (B-Hemoglobin or 201+) with laboratory haemoglobin estimation. *Anaesthesia and intensive care* 2015;43.(3): 341-350.

[11] Marn H, Critchley JA: 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. *The Lancet Global Health* 2016;4(4): e251-e265.

[12] WHO. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity Vitamin and Mineral Nutrition Information System. Geneva: World Health Organisation,(WHO/NMH/NHD/MNM/11.1). 2011.

[13] Mulherin SA: Spectrum Bias or Spectrum Effect? Subgroup Variation in Diagnostic Test Evaluation. *Annals of Internal Medicine* 2002;137(7): 598.

[14] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al.: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8): 529-536.

[15] Adam I, Ahmed S, Mahmoud MH, Yassin MI: Comparison of HemoCue(R) hemoglobin-meter and automated hematology analyzer in measurement of hemoglobin levels in pregnant women at Khartoum hospital, Sudan. *Diagn Pathol* 2012;7: 30.

[16] Daae LN HS, Mathisen PM, Mironska K.: A comparison between haematological parameters in 'capillary' and venous blood from healthy adults. *Scand J Clin Lab Invest* 1988;48(7): 723-726.

[17] Saxena R, Malik R: Comparison of HemoCue method with the cyanmethemoglobin method for estimation of hemoglobin. *Indian pediatrics* 2003;40(9): 917.

[18] Moher D, Liberati A, Tetzlaff J, Altman DG, Group P: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine* 2009;6(7): e1000097.

[19] Harbord RM WP: metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. . *The Stata Journal* 2009;9(2): 211-229.

[20] Deeks JJ, Altman DG: Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329(7458): 168-169.

[21] Stevens GA, Finucane MM, De-Regil LM, Paciorek CJ, Flaxman SR, Branca F, et al.: Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995–2011: a systematic analysis of population-representative data. *The Lancet Global Health* 2013;1(1): e16-e25.

[22] Easycalculation. Diagnostic Post Test Probability of Disease Calculator. https://www.easycalculation.com/statistics/post-test-probability.php. Accessed December 2016.

[23] Pistorius LR, Funk M, Pattinson RC, Howarth GR. Screening for anemia in pregnancy with copper sulfate densitometry. *Int J Gynecol Obstet* 1996;52(1): 33–36.

[24] Wilkinson D, Sach ME: Cost effective on-site screening for anaemia in pregnancy in primary care clinics. *South African Medical Journal* 1997;87(4): 463-465.

[25] 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. *Bulletin of the World Health Organization* 1999;77(1): 15-21.

[26] Shulman CE, Levene M, Morison L, Dorman E, Peshu N, Marsh K: Screening for severe anaemia in pregnancy in Kenya, using pallor examination and self-reported morbidity. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2001;95(3): 250-255.

[27] Fourn L, Salami L: [Diagnostic value of tegument pallor in anemia in pregnant women in Benin]. *Sante publique (Vandoeuvre-les-Nancy, France)* 2004;16(1): 123-132.

[28] Prathapan S LG, 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): 245-250.

[29] 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. *The Ceylon medical journal* 2012;57(3): 101-106.

[30] Khan AA, Fatmi Z, Kadir MM: Accuracy and Use of WHO Hemoglobin Color Scale for Diagnosis of Anemia Among Pregnant Women by Health Care Providers in Periurban Settings in Karachi, Pakistan. *Asia-Pacific journal of public health / Asia-Pacific Academic Consortium for Public Health* 2015;27(6): 610-619.

[31] Agnihotri M, Ambad R, Rahule AS: Study of Evaluation of Sensitivity and Specificity of Simple Screening Methods for Assessment of Anaemia in Pregnant Women. *Journal of Contemporary Medicine and Dentistry* 2015;3(1): 62-66.

[32] Ahankari AS, Fogarty AW, Tata LJ, Dixit JV, Myles PR: Assessment of a non-invasive haemoglobin sensor NBM 200 among pregnant women in rural India. *BMJ Innovations* 2016;2(2): 70-77.

[33] Simel DL, Bossuyt PM: Differences between univariate and bivariate models for summarizing diagnostic accuracy may not be large. *J Clin Epidemiol* 2009;62(12): 1292-1300.

[34] Lewis SM, Stott GJ, Wynn KJ: An inexpensive and reliable new haemoglobin colour scale for assessing anaemia. *J Clin Pathol* 1998;51(1): 21-24.

[35] PATH. Anemia detection in health services: Guidelines for program managers. . In: health P-Pfati, ed.: USAID. 1996.

[36] R.Parikh AM, S. Parikh, G. Chandra Sekhar, R. Thomas: Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008;56(1): 45-50.

[37] Pena-Rosas JP, De-Regil LM, Garcia-Casal MN, Dowswell T: Daily oral iron supplementation during pregnancy. *The Cochrane database of systematic reviews* 2015;7: CD004736.

[38] Maxwell MJ, Wilson MJA: Complications of blood transfusion. *Continuing Education in Anaesthesia, Critical Care & Pain* 2006;6(6): 225-229.

[39] Africare. Flash of insight: mobile app tests for anemia with camera flash. https://www.africare.org/flash-of-insight-mobile-app-tests-for-anemia-with-camera-flash/. Published 2015. Accessed August 2016.

Figure legends

Figure 1 Study selection process.

Figure 2 Receiver operating characteristic curve of on-site test accuracy to detect mild anemia during pregnancy. Mild anemia was defined as a Hb level of either less than 110 g/L or less than 100 g/L. Abbreviations: Hb, hemoglobin; HCS, hemoglobin color scale.

Supporting information legends

Appendix S1 Search history for the systematic review. The same search was used in all databases.

Figure S1 Receiver operating characteristic curve of the accuracy of clinical signs to detect mild anemia (hemoglobin level <110 g/L) during pregnancy.

Figure S2 Receiver operating characteristic curve of the accuracy of hemoglobin color scale to detect mild anemia (hemoglobin level <110 g/L) during pregnancy.

Table S1 Accuracy of on-site tests used to diagnose anemia in pregnancy.

Table 1 Characteristics of accuracy studies for tests to detect anemia during pregnancy.

Study	Country and setting	No. of pregnant women included	Index test	Reference test
Pistorius et al. 1996 [23]	Pretoria, South Africa Public hospital	100	$CuSO_4$	Coulter counter
Wilkinson and Sach 1997 [24]	Kwala Zulu, South Africa One mobile clinic team serving 14 clinic points across the district	449	$CuSO_4$	Sysmex analyzer
van den Broek et al. 1999 [25]	Malawi Three rural hospitals and two health centers	644	Clinical signs, HCS, and HemoCue	Coulter counter
Shulman et al. 2001 [26]	Mombasa, Kenya District hospital	1787	Clinical signs	Coulter counter
Fourn and Salami 2004 [27]	Benin Rural maternity clinic	480	Clinical signs	Spectrophotometry laboratory test
Prathapan et al. 2011 [28]	Colombo district, Sri Lanka Field prenatal clinics in 11 of 13 Ministry of Health areas in this district	101	HCS	Spectrophotometry Laboratory test
Chathurani et al. 2012 [29]	Anuradhapura district, Sri Lanka Ministry of Health field clinics in this district	115	Clinical signs and HCS	Cyanmethemoglobin method
Khan et al. 2015 [30]	Karachi, Pakistan Community-based prenatal clinics in the towns of Gadap, Bin Qasim, Kemari, and New Karachi	194	Clinical signs and HCS	Calorimeteric hemoglobinometry
Agnihotri et al. 2015 [31]	India Prenatal clinic in the Obstetrics and Gynecology department Jawaharlal Nehru Medical College	100	Sahli method and $CuSO_4$	Cyanmethemoglobin method
Ahankari et al. 2016 [32]	India Villages in Tuljapur and Lohara blocks of Osmanabad district, Maharashtra (n=33).	269	Non-invasive hemoglobin sensor (NBM 2000)	Sysmex analyzer XP-100

Abbreviations: $CuSO_4$, copper sulfate; HCS, hemoglobin color scale.

Table 2 Risk of bias of accuracy studies for tests to detect anemia during pregnancy.[a]

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Pistorius et al. 1996 [23]	Low	Low	Low	Low	Low	Low	Low
Wilkinson and Sach 1997 [24]	Low	Low	Low	Low	Low	Low	Low
van den Broek et al. 1999 [25]	Unclear	Low	Low	Low	Low	Low	Low
Shulman et al. 2001 [26]	Low	Low	Low	Low	Low	Low	Low
Fourn and Salami 2004 [27]	Low	Low	Low	Low	Low	Low	Low
Chathurani et al. 2012 [29]	High	High	Low	Low	Low	Low	Low
Prapathan et al. 2011 [28]	Low	High	Low	Low	Low	Low	Low
Khan et al. 2015 [30]	Low	Low	Low	Low	Low	Low	Low
Agnihotri et al. 2015 [31]	Unclear	Low	Low	Unclear	Low	Low	Low
Ahankari et al. 2016 [32]	Unclear	Low	Low	Low	Low	Low	Low

[a] Risk of bias was measured using the QUADAS-2 quality assessment tool [14] and defined as low, high, or unclear.

Table 3 Accuracy of on-site tests to diagnose mild anemia during pregnancy.[a,b]

Index test	Prevalence, %	No. of studies	No. of women	Sensitivity, %	Specificity, %	Likelihood ratio (95% CI)		Post-test probability, %	
						Positive test result	Negative test result	Positive test result	Negative test result
Clinical	47 (27–72)	4	1853	56 (19–92)	62 (30–93)	1.7 (1.0–2.9)	0.57 (0.33–1.03)	55	29
HCS	42 (17–67)	4	1051	67 (56–76)	67 (48–82)	2.0 (1.3–3.1)	0.50 (0.40–0.62)	59	27
$CuSO_4$	58 (48–67)	1	100	97 (88–100)	71 (55–85)	3.4 (2.1–5.5)	0.05 (0.01–0.19)	71	3.5
Sahli method	58 (48–67)	1	100	86 (75–94)	83 (68–93)	5.2 (2.6–10.3)	0.17 (0.09–0.32)	79	11
Non-invasive hemoglobin sensor (NBM 2000)	77 (72–82)	1	269	34 (27–41)	92 (82–97)	4.1 (1.7–9.7)	0.72 (0.64–0.82)	75	34
HemoCue	23 (20–27)	1	671	85 (79–90)	80 (76–83)	4.3 (3.5–5.1)	0.18 (0.12–0.27)	76	12

Abbreviations: HCS, hemoglobin color scale; $CuSO_4$, copper sulfate.

[a] The pretest probability for mild anemia (hemoglobin <110 g/L) was 42%.

[b] Values in parentheses are 95% confidence intervals.