

# Haemoglobin Colour Scale

## Operational Research Agenda and Study Design



Department of Essential Health Technologies  
World Health Organization  
1211 Geneva, Switzerland  
eht@who.int  
[www.who.int/eht](http://www.who.int/eht)

## **ACKNOWLEDGEMENTS**

The formulation of this guideline is the outcome of a WHO consultation on the usefulness of the Haemoglobin Colour Scale held at WHO/HQ in May 2004. The World Health Organization thanks Drs Imelda Bates and Julia Critchley of the Liverpool School of Tropical Medicine, United Kingdom, as principal authors for putting it together.

### **© World Health Organization 2004**

All rights reserved. Publications of the World Health Organization can be obtained from Marketing and Dissemination, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (email: [bookorders@who.int](mailto:bookorders@who.int)). Requests for permission to reproduce or translate WHO publications – whether for sale or for noncommercial distribution – should be addressed to Publications, at the above address (email: [permissions@who.int](mailto:permissions@who.int)).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

The World Health Organization does not warrant that the information contained in this publication is complete and correct and shall not be liable for any damages incurred as a result of its use.

Printed by the WHO Document Production Services, Geneva, Switzerland.

## **1. BACKGROUND**

### ***Global burden of anaemia***

Anaemia affects over half of all young children [1] and pregnant women in developing countries and is one of the commonest causes of preventable deaths in malaria-endemic countries [2]. Reducing the anaemia burden is therefore essential in order to achieve the United Nations Millennium Development Goals concerned with reducing maternal and childhood mortality. However, those with the greatest burden of anaemia are also those that are hardest to reach – the poor and marginalized, women and children, and those who have no access to the health system.

Anaemia is the world's second most common cause of disability [3] – even non-severe anaemia is associated with increased stillbirths, low-birth-weight babies and intellectual and physical impairment in children. Anaemia is therefore one of the most serious, but relatively neglected, health problems. It is 'neglected' because compared to other major public health problems, such as HIV/AIDS, malaria and TB, proportionately little international effort has been put into developing appropriate technologies to detect and measure anaemia, and in implementing proven preventative strategies.

### ***Use of haemoglobin estimations by major international health programmes***

In poorer countries many factors interact to produce anaemia including malaria, poor nutrition, helminth infections and haemoglobinopathies. Simple, evidence-based interventions are available to counteract or alleviate many of these factors. Such interventions include the use of insecticide treated bednets and anti-malaria therapy, iron supplements and 'de-worming' programmes.

Although almost all major health programmes, including those dealing with malaria, antenatal care, child health, nutrition and HIV/AIDS, need to address the problem of anaemia, they have differing anaemia control strategies. At a consultation meeting of relevant WHO programmes [4] it became clear that only the 'Making pregnancy safer' programme implements community based screening on a national scale because this programme has a policy of routinely screening all pregnant women for anaemia [5]. Nutrition programmes are using haemoglobin (predominantly measured by HemoCue® during specific surveys) as a proxy measure of iron deficiency and are compiling a database of haemoglobin levels. Roll Back Malaria is also using HemoCue® haemoglobin measurements to monitor the effectiveness of intervention programmes. Therefore, apart from ante-natal care programmes, almost all other major programmes need haemoglobin estimations that are more accurate than can be achieved by using the HbCS. A possible exception is the HIV/AIDS programme which is still in the process of developing its anaemia policy but may require anaemia screening prior to starting zidovudine anti-retroviral therapy [6] and as a tool to monitor adverse events. Haemoglobin estimations are an essential component of the blood transfusion process, where they are used both for screening donors and for assessing the degree of anaemia in transfusion recipients. Although 12.5g/dl is the

recommended cut-off level for blood donors, many countries actually use a level of 12g/dl. The HbCS is not accurate enough to use for a 12.5g/dl cut-off but could be useful for identifying potential donors with haemoglobins above or below 12g/dl. The HbCS is considered too inaccurate to be used to guide administration of blood transfusions [4].

One of the recommendations of the consultation meeting was that WHO should clearly define the role of the HbCS in public health to ensure its safe, appropriate and cost-effective use. To achieve this goal, WHO needs to provide a framework for guiding the scope and quality of information needed in the areas of operational research and an evaluation of current usage of HbCS. It also needs to develop a strategy for disseminating information, educational tools and communication approaches about best practice relating to use of the HbCS. The target groups for this dissemination strategy will be public health stakeholders, particularly national and international policy makers and advisors, and researchers interested in pursuing projects on public health aspects of anaemia diagnosis and management.

### **Methods available for estimating haemoglobin for public health purposes**

The vast majority of anaemia diagnoses at community level are based on clinical diagnosis alone. This is well recognised to have low sensitivity and specificity, particularly for mild and even moderate anaemia [7]. Clinical indicators of anaemia are certainly not adequate for population-based surveys of anaemia [8]. It is recommended that reliable, 'relatively simple and inexpensive' methods should be used for measuring haemoglobin [8] with an implicit assumption that such methods are available at community level. Some methods use inexpensive consumables (e.g haemoglobinocyanide method) but the fact that they are technically complicated and require careful supervision and quality assurance processes is often overlooked. Other methods (e.g. HemoCue®) are technically more simple but use costly disposable cuvettes making them too expensive for widespread use [9]. The misperceptions held by many programme managers and policy makers about the technical issues around haemoglobin estimations are fuelled by a lack of communication at the clinical-laboratory interface and insufficient evidence about cost-effectiveness on which to base rational decisions about the choice of tools for measuring haemoglobin.

### **Development and use of the Haemoglobin Colour Scale**

Development of the Haemoglobin Colour Scale (HbCS) commenced in 1995 in response to a recognised need for a simple, rapid and cheap method for estimating haemoglobin that could be performed on finger prick samples, and that could be used in situations where there was no laboratory, and where anaemia diagnosis otherwise relied on clinical signs. The HbCS became commercially available in 2001. It is manufactured by Copack in Germany under a licence agreement with WHO. The basic kit (scale, instructions and 200 test strips) costs around €7.70, equivalent to approximately US\$0.01/test. However, distribution and freight costs, as well as potential government levies, significantly increase the cost to the end user. The product has been included on the UNICEF standard products list (catalogue no. 0003390) and this may reduce these costs.

The HbCS relies on comparing the colour of a drop of blood absorbed on special chromatography paper with standard colours on a laminated card displayed in increments of 2g/dl. Initial evaluations of the performance of the HbCS concentrated on comparing its

accuracy with standard laboratory techniques. As these methods measure haemoglobin levels to within 0.1g/dl it was not surprising that the results, and their interpretation, varied widely between studies. There have been very few studies designed to test the original hypothesis that drove development of the HbCS; namely, 'is the HbCS a useful tool for improving the detection and management of anaemia in resource-poor settings where there is no laboratory?' A recent systematic review of all the published studies (N=11) that assessed performance of the HbCS demonstrated reasonable sensitivity but sometimes low specificity for the HbCS [10]. Conclusions were limited as the design of the studies and the results were very heterogeneous. Some were not carried out using routine health services and none assessed the impact of using the HbCS on clinical outcomes.

In contrast to the lack of clear evidence of its benefit compared to clinical diagnosis, the HbCS is undoubtedly cheaper than other manual techniques for estimating haemoglobin. In a study to determine the comparative costs of all resources needed to support various haemoglobin methods in routine use in a district hospital in Malawi, the HbCS cost \$0.12/test compared to \$0.35/test for colorimetric methods and \$0.75/test for HemoCue® (Table 1) [9].

**Table 1. Projected costs of resources for annual haemoglobin workload (6161 tests) in Ntcheu District Hospital, using different methods, in US dollars**

	HbCS (%)	Lovibond (%)	DHT meter (%)	WBA colorimeter (%)	Jenway colorimeter (%)	HCN (%)	HemoCue (%)
<i>Staff<sup>a</sup></i>	202.8 (27.0)	338.0 (36.0)	338.0 (27.5)	1013.9 (46.7)	1013.9 (46.4)	1013.9 (28.8)	338.0 (7.3)
<i>Supplies<sup>b</sup></i>	549.6 (73.0)	472.3 (50.3)	472.3 (38.4)	657.0 (30.3)	657.0 (30.1)	1587.1 (45.2)	3990.1 (86.0)
<i>Equipment</i>	0 (0)	129.1 (13.7)	183.3 (14.9)	264.7 (12.2)	278.6 (12.7)	677.3 (19.3)	74.1 (1.6)
<i>Overheads</i>	0 (0)	0 (0)	236.1 (19.2)	236.1 (10.9)	236.1 (10.8)	236.1 (6.7)	236.1 (5.1)
<b>Total</b>	<b>752.4</b>	<b>939.4</b>	<b>1229.7</b>	<b>2171.7</b>	<b>2185.6</b>	<b>3514.4</b>	<b>4638.3</b>
<b>Cost per test</b>	<b>0.12</b>	<b>0.15</b>	<b>0.20</b>	<b>0.35</b>	<b>0.35</b>	<b>0.57</b>	<b>0.75</b>

<sup>a</sup> Staff costs include the time taken to perform 6161 tests for each method; the cost of the time taken to calibrate machines was negligible.

<sup>b</sup> The costs included in supplies are those associated with performing a test, the control test and the material necessary for calibration.

Comparisons between different haemoglobin methods are not straightforward. Methods that use expensive consumables have increasing costs/test as the workload increases, whereas for methods that use minimal amounts of supplies, the cost/test will decrease with increasing workload. The HbCS is easy to use and in the same Malawian study it was rated third with a score of 17/25 for user-friendliness (Table 2).

**Table 2. User-friendliness of different haemoglobin methods**

<b>Characteristics</b> (maximum score 5 each)	<b>Hemo Cue</b>	<b>DHT meter</b>	<b>HbCS</b>	<b>Jenway colorimeter</b>	<b>Lovibond</b>	<b>WBA colorimeter</b>	<b>HCN reference method</b>
Easy to use	5	4	5	4	4	4	2
Minimal training/supervision required	5	3	5	2	5	2	1
Simple reading of results	5	5	2	5	2	3	3
Simple internal quality	5	2	0	3	0	3	3
Suitable for village use	5	5	5	1	4	1	1
<b>Total (max 25)</b>	<b>25</b>	<b>19</b>	<b>17</b>	<b>15</b>	<b>15</b>	<b>13</b>	<b>10</b>

## **2. PRINCIPLES OF FUTURE HbCS RESEARCH STUDIES**

Overall, the HbCS is more sensitive than clinical diagnosis to screen for anaemia, but specificity is a problem and the potential over-diagnosis rates may be important if proposed interventions are expensive or associated with adverse effects. There are several operational research questions that need to be addressed through prospective research projects. These relate particularly to the effectiveness of the HbCS in detecting anaemia compared to the current widely used method of clinical diagnosis alone. So far, all the HbCS studies have been concerned with diagnostic accuracy. A neglected area, but of critical importance for determining the public health role of the HbCS, is evaluation of its impact on clinical outcomes. Measuring these outcomes will be difficult, especially in the resource-poor, community settings where the HbCS needs to be tested, so it will also be important to collect intermediate process indicators.

Since most of the HbCS studies have been carried out under ideal conditions, more pragmatic studies are needed to collect concrete data on how the HbCS performs in 'real life' situations. The HbCS provides an immediate visual result that is easily interpreted, even by those who cannot read, and is therefore popular with patients and health care providers. This potential educational role of the HbCS, and the effect this may have on test requesting patterns and compliance, has not been formally examined and studies need to be designed that can evaluate this potential secondary benefit of the HbCS. Potential outcomes to measure this effect may relate to the number of individuals screened for anaemia, better follow up attendances and compliance with treatment, and spontaneous patient requests for other laboratory tests.

### **3 HbCS: OPERATIONAL RESEARCH QUESTIONS**

Further data on the diagnostic performance (sensitivity, specificity, likelihood ratios, positive predictive values and negative predictive values) of the HbCS is needed in different population groups and settings where it may be used. Sensitivity and specificity are intrinsic properties of a diagnostic test, but results from published studies have shown great heterogeneity in these estimates [10]. Precise sensitivity and specificity is therefore unclear at present. Estimates of positive and negative predictive values are also required. These are influenced by the prevalence of anaemia in the population, and provide information of immediate clinical utility (i.e. the likelihood of anaemia among those who test positive on the HbCS).

Further studies are required which evaluate the HbCS against routinely practised, simple clinical diagnosis in patient groups, which may use the HbCS (children, especially those under 5 years, pregnant women, potential ARV users, monitoring ARV therapy, blood donors). These studies should also address the acceptability of the HbCS to these various groups of patients as reluctance to have any blood test is a problem for community antenatal services and may be a major barrier to implementing anything other than purely clinical anaemia diagnosis. Future research needs to incorporate measures of clinical outcomes as well as diagnostic accuracy. Depending on the patient population in which the HbCS is used, these outcome indicators could include anaemia management or birth weight. Process indicators such as compliance with iron therapy or correct allocation to anaemia severity category could also be measured. Ideally, randomised controlled trials are required, where the patient population is randomised to either anaemia diagnosis using i) the HbCS or ii) other locally available methods of diagnosis (e.g. clinical examination).

Research projects should also take account of the potential impact of wide scale use of the HbCS. These include the need to introduce finger pricks in communities where clinical diagnosis has been the norm. This creates additional problems such as reluctance to have blood tests, disposal of sharps, and the risk of needle-stick injuries. Simple quality control mechanisms need to be developed and piloted with the aim of ensuring consistency of results. This is particularly important because the test is likely to be performed by non-technical health workers.

Projects also need to be undertaken to evaluate the requirements for scaling up the availability and use of HbCS. These will include evaluations of various types of packaging (e.g. bundling HbCS kits with lancets), production costs, optimising training and supervisory networks, assessing quality assurance monitoring systems and identifying feasible distribution networks.

### **4. DESIGNING STUDIES TO EVALUATE THE HbCS**

There are now validated guidelines and standards for assessing the methodological quality of diagnostic studies [11-15], and studies evaluating the HbCS should address these quality issues. In particular, there is empirical evidence that studies which use case-control designs, used different reference tests, are not blinded, and which do not adequately describe the population or test characteristics, overestimate diagnostic performance [16]. In order to provide information that is of practical relevance, research projects should aim

to compare HbCS with current methods of anaemia diagnosis in pragmatic studies that are as close to 'real life' as possible. In brief, studies that aim to evaluate the performance of the HbCS should:

- follow Standards for Reporting of Diagnostic Accuracy (STARD) to improve the quality of reporting of studies on diagnostic accuracy (see Appendix 1)[12]
- provide clear descriptions of selection criteria, methods of testing and reference standard ('gold standard') used
- specify inclusion and exclusion criteria
- provide information about ages and pregnancy status of study population as haemoglobin values vary with age and during pregnancy
- explain withdrawals
- use a simultaneous analysis of samples
- describe the type of sample and use the same sample for all haemoglobin measurements
- be blinded (i.e. the HbCS result should be interpreted without knowledge of the reference standard measured independently of the test).
- measure haemoglobin independently of all other clinical information (e.g. patient notes)
- include a reality check for real life use
- specify the background prevalence of anaemia
- provide definitions of none, mild, moderate and severe anaemia
- state whether HbCS measured anaemia to nearest 1g/dl or 2g/dl
- provide actual haemoglobin values rather than batch results into categories (because the differing cut-offs between studies means that results cannot be compared)
- report on inter and intra-user agreement for health care providers (using kappa statistics)
- analyse data in an appropriate manner. Appropriate methods include Bland-Altman "limits of agreement", estimates of diagnostic accuracy. Inappropriate methods include correlation coefficients and t tests[17]
- provide data on the number of true positives, false positives, true negatives and false negatives, and provide estimates of precision (e.g. 95% confidence intervals) around major statistics.

This list has been adapted from validated guidelines for assessing the quality of diagnostic studies [11-14, 18] . Some items have been omitted (e.g. those dealing with 'incorporation bias' as only a single reference standard is used to verify disease status[13]. A few items specific to the HbCS have been added (e.g. measuring haemoglobin to the nearest 1 or 2g/dl). It is recognised that studies will not always be able to adhere to the complete list, for example in certain cases reference methods will not always be available locally and samples may have to be sent to a distant laboratory (simultaneous evaluation of HbCS

and reference standard results will then not be possible). Where possible and appropriate study design should be a randomised controlled trial. As a minimum, all quantitative studies should be prospective in design, use a consecutive case series of patients, and the same reference test for all patients, be blinded, and describe clearly the population under study, test characteristics, type of sample and findings[15-16].

### ***Populations to be included in HbCS studies***

The HbCS has a potential role to play in both preventative and curative health services. The major potential users of the HbCS that should be addressed in any research programme are:

- Ante-natal clinics - for routine anaemia screening and detection of those that need specific treatment
- Under 5s clinics – detection of mild, moderate and severe anaemia to guide management strategies
- ARV programmes – to screen for pre-treatment anaemia as a guide to appropriate therapy and to monitor therapy or disease related exacerbations of anaemia
- Blood transfusion service – to detect and exclude anaemic donors
- Malaria and nutrition programmes – to measure success of interventions (though it is more likely that HemoCue® or equivalent would be used in these circumstances).

### ***Conflicts of interest***

Researchers with positive or negative interests in the development or sale of the HbCS should clearly state their interests and any potential effect that this may have on the interpretation of the results.

### ***Quality control of studies***

This should be ensured through:

- Use of external monitors (e.g. scientific committee)
- Approval of study by relevant ethical committees
- Listing of bias and effect modifiers with an explanation of how these will be controlled
- Final report of a standard eligible for publication in peer-reviewed journal.

### ***Sample size and randomisation***

Wherever possible, trials of the HbCS and its effects on clinical outcomes should include randomised comparisons. Future studies, whether randomised or not, should have sufficient power to detect clinically important differences between the diagnostic tests in clinical outcomes (e.g. a 250g difference in birth weight).

**Table 3. Endpoints, indicators and data collection methods**

Endpoints	Indicators	Data collection method
Correct estimation of haemoglobin	Measures of diagnostic accuracy (e.g. sens, spec, PPV, NPV, LR)*	Prospective studies of diagnostic accuracy
Correct categorisation of anaemia	Clients correctly allocated to anaemia category	Independent assessment of number of clients allocated to correct anaemia category
Correct selection of non-anaemic blood donors	Donors selected correctly	Compare ability of HbCS and reference method to select non-anaemic donors
Appropriate management instigated	Increase in Hb levels	Clinical audit Independent observations
Newborn birth weight	Birth weight	Systematic labour ward data collection
User preference	Health service users continue to use HbCS	Semi-structured interviews Questionnaire
Client preference	Clients prefer HbCS to other methods of anaemia diagnosis	Semi-structured interviews Questionnaire
Improved adherence to therapy	Compliance assessment	Pill counts Questionnaire Follow up attendances
Acceptance of blood tests	Client accepts and/or requests other laboratory tests	Semi-structured interviews
Maintenance of high quality results	Quality monitoring results	Internal quality control checks External quality assurance assessments
Optimal training and supervision	Adherence to standard operating procedures	Independent observation External quality assessment
HbCS is cost-effective	Reduction in severe anaemia and low birth weight babies	Comparative economic analysis

\*Sens = sensitivity, spec = specificity, PPV = positive predictive value, NPV = negative predictive value, LR = likelihood ratio

## 5. REFERENCES

1. Schellenberg D et al. The silent burden of anaemia in Tanzanian children: a community-based study. *Bulletin of the WHO* 2003; 81: 581-590.
2. McDermott JM et al. Prospective assessment of mortality among a cohort of pregnant women in rural Malawi. *American Journal of Tropical Medicine* 1996;55(1):66-70.
3. Murray CJL LA. The global burden of disease. Cambridge MA ed. Harvard University Press; 1996.
4. Review of the Haemoglobin Colour Scale. Report of an informal consultation. 10 May 2004. WHO/EHT/04.12. WHO, Geneva, Switzerland.
5. Care of mother and baby at the health centre: A practical guide. Introduction and Table 2.3 Prenatal care. WHO/MSM/94.2-Rev 2
6. Scaling up antiretroviral therapy in resource limited settings: guidelines for a public health approach. 2003 revision. WHO report 1/4/2002. Table E. P24.
7. Shulman CE, Levene M, Morison L, Dorman E, Peshu N, and 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: 250-55.
8. Iron deficiency anaemia. Association, prevention and control. A guide for programme managers. Published by UNICEF, United Nations University and WHO, p34-35 WHO/NHD/01.3.
9. Lara AM, Mundy C, Kandulu J, Chisuwo L, and Bates I. Choosing a haemoglobin method for district hospitals in Malawi. *Journal of Clinical Pathology* 2004 [In press].
10. Haemoglobin Colour Scale –Review of studies. EHT, WHO, Geneva. May 2004.
11. Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, and Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine* 120, 667-76. 1994.
12. Bossuyt, PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, and de Vet HCW. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 326(7379), 41-44. 2003.
13. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, and Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 3, 1-13. 2003.
14. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 323, 157-62. 2001.
15. Whiting, P, Rutjes, AWS, Reitsma JB, Glas AS, Bossuyt PMM, and Kleijnen J. Sources of Variation and Bias in Studies of Diagnostic Accuracy. *Annals of Internal Medicine* 140, 189-202. 2004.
16. Lijmer, Jeroen G., Mol, Ben Willem, Heisterkamp, Siem, Bonsel, Gouke J., Prins, Martin H., van der Meulen, Jan H. P., and Bossuyt, Patrick M. M. Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests. *Journal of the American Medical Association* 282(11), 1061-1066. 1999.
17. White SA, Van Den Broek NR. Methods for Assessing Reliability and Validity for a Measurement Tool: a Case Study and Critique Using the Who Haemoglobin Colour Scale. *Statistics in Medicine* 2004;23:1603-19.
18. Deeks JJ. Systematic reviews of evaluations of diagnostic tests. In Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care. Meta-analysis in Context*, London: BMJ Publishing Group, 2001.

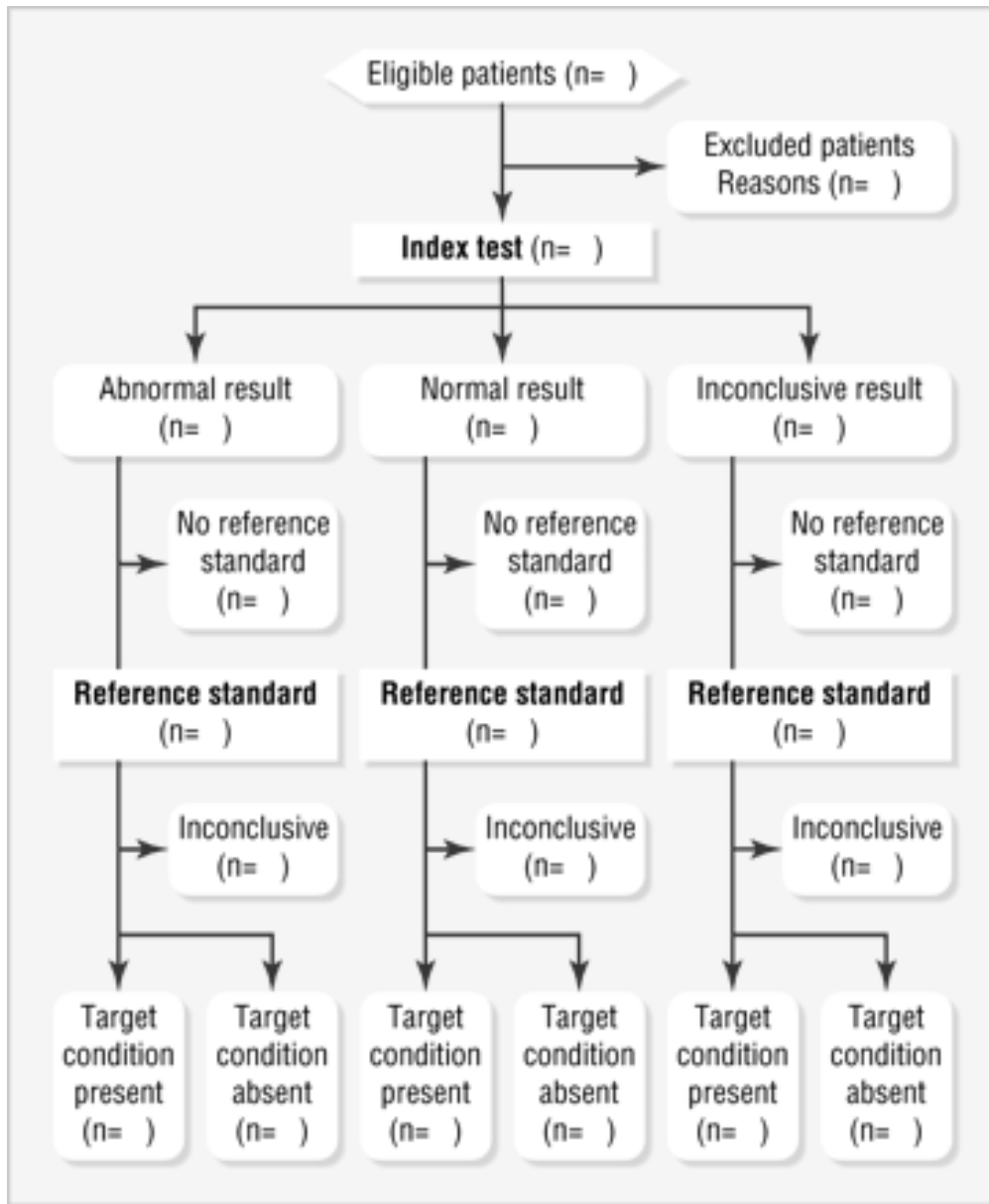
## Appendix 1. STARD checklist for reporting diagnostic accuracy studies

Section and topic	Item	Description
Title, abstract, and keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading "sensitivity and specificity")
Introduction	2	State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups
Methods:		
Participants	3	Describe the study population: the inclusion and exclusion criteria and the settings and locations where the data were collected
	4	Describe participant recruitment: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?
	5	Describe participant sampling: was this a consecutive series of participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected
	6	Describe data collection: was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	Describe the reference standard and its rationale
	8	Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for index tests or reference standard, or both
	9	Describe definition of and rationale for the units, cut-off points, or categories of the results of the index tests and the reference standard
	10	Describe the number, training, and expertise of the persons executing and reading the index tests and the reference standard
	11	Were the readers of the index tests and the reference standard blind (masked) to the results of the other test? Describe any other clinical information available to the readers.
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy and the statistical methods used to quantify uncertainty (eg 95% confidence intervals)
	13	Describe methods for calculating test reproducibility, if done
Results:		
Participants	14	Report when study was done, including beginning and ending dates of recruitment
	15	Report clinical and demographic characteristics (eg age, sex, spectrum of presenting symptoms, comorbidity, current treatments, and recruitment centre)

- |              |    |  |
|--------------|----|--|
|              | 16 | Report how many participants satisfying the criteria for inclusion did or did not undergo the index tests or the reference standard, or both; describe why participants failed to receive either test (a flow diagram is strongly recommended)                   |
| Test results | 17 | Report time interval from index tests to reference standard, and any treatment administered between  |
|              | 18 | Report distribution of severity of disease (define criteria) in those with the target condition and other diagnoses in participants without the target condition   |
|              | 19 | Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, report the distribution of the test results by the results of the reference standard |
|              | 20 | Report any adverse events from performing the index test or the reference standard   |
| Estimates    | 21 | Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)  |
|              | 22 | Report how indeterminate results, missing responses, and outliers of index tests were handled  |
|              | 23 | Report estimates of variability of diagnostic accuracy between readers, centres, or subgroups of participants, if done   |
|              | 24 | Report estimates of test reproducibility, if done  |
| Discussion   | 25 | Discuss the clinical applicability of the study findings   |

Reproduced from: Bossuyt, Patrick M, Reitsma, Johannes B, Bruns, David E, Gatsonis, Constantine A, Glasziou, Paul P, Irwig, Les M, Lijmer, Jeroen G, Moher, David, Rennie, Drummond, and de Vet, Henrica C W. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 326(7379), 41-44. 2003.

**Appendix 2. Flow diagram for a Study on Diagnostic Accuracy**



Reproduced from: Bossuyt, Patrick M, Reitsma, Johannes B, Bruns, David E, Gatsonis, Constantine A, Glasziou, Paul P, Irwig, Les M, Lijmer, Jeroen G, Moher, David, Rennie, Drummond, and de Vet, Henrica C W. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 326(7379), 41-44. 2003.



