

## Role of point of care Hb diagnostic devices in getting the right picture of anemia control: Tangi Rural Anemia Diagnostic Accuracy Study

Preetam B. Mahajan<sup>a\*</sup> and Somnath Mukherjee<sup>b</sup>

<sup>a</sup>Department of Community Medicine & Family Medicine, All India Institute of Medical Sciences, Bhubaneswar, Odisha, India; <sup>b</sup>Department of Transfusion Medicine, All India Institute of Medical Sciences, Bhubaneswar, Odisha, India

### ABSTRACT

**Objectives:** To study the accuracy and reliability of two point of care devices (Hemocue 201 [HC<sub>201</sub>]) and Hemoglobin Colour Scale (HCS) with reference to automated laboratory analyser Sysmex XP 100 (Lab<sub>SP</sub>) and determine factors influencing their performances. To discuss certain design issues while ascertaining these parameters for its judicious use in both clinical context and disease burden studies.

**Methods:** Reliability and accuracy statistics were calculated for four population subgroups that were selected randomly using cluster sampling in a rural community of eastern India. Appropriate measures were taken to reduce biases in the study. Bland Altman Plot was used to determine Bias and ROC curve analysis was used to suggest new cut-offs for HCS method.

**Results:** True prevalence varied across subgroups ranging from 12.56% in adolescent boys to 40.71% in adult women. Sensitivity for HC<sub>201</sub> was lowest among boys (80.39%) and highest among adult females (92.82%), while specificity was highest among adult males (86.94%) and lowest among adult females (75.00%). The variation across subgroups was due to differences in distribution of underlying Hb values. HC<sub>201</sub> has potential for use in clinical practice as well as disease burden estimation study. HCS was not suitable for use in clinical setting as the bias (−4.1 g/dl) was unacceptable. However, ROC analysis suggested certain cut-offs for different age groups and can be employed in poor resource settings for disease burden estimation study.

**Conclusions:** HC<sub>201</sub> is better than HCS as per the study. However, accuracy parameters are likely to vary depending on the distribution of underlying Hb distribution. Thus, same machines can demonstrate different accuracy parameters in different settings. There is need for designing studies that could help estimate these parameters each time on a subsample which would go a long way in efficiently using technology be it for guiding clinical decisions or public health actions.

### ARTICLE HISTORY

Received 29 March 2018  
Accepted 7 June 2018

### KEYWORDS

Anemia; diagnostic accuracy study; point of care diagnostic device; Hemocue 201; hemoglobin colour scale



### Introduction

Anemia is an invisible public health problem of global concern. In the year 2011, iron-deficiency anemia (IDA) alone was rated as third major cause of disability worldwide, and resulted into loss of 42.2 million disability-adjusted life years (DALYs) [1]. In India, as per the current global burden of disease study estimates, IDA is the topmost contributor of years lived with disability, accounting for 11% of all disability, its effects being most severe among women. In terms of DALYs, it showed an increase in both rank (currently six) and health loss caused, suggesting a need for renewed policy attention to this cause [2].

It is high time focus is shifted from preventive to cause specific curative action in anemia control. In its revised strategy, Ministry of Health and Family Welfare, Government of India, lays emphasis on treatment of IDA besides supplementation measures to control it [3]. To identify patients that would require treatment, clinical methods like palmer and

conjunctival pallor are being recommended. However, evidence suggests that pallor is only useful in severe anemia and its sensitivity is unacceptable in mild, moderate anemia [4–10]. This would result in missed cases and delayed treatment. One way, is to replace these clinical methods with simple, cost-effective, user-friendly point of care diagnostic devices (POCDs) that have acceptable test accuracy parameters, and margin of error within  $\pm 1$  g/dl of true hemoglobin value (as determined by reference tests).

Further, from public health point of view, it is also important to continuously monitor the trend of anemia prevalence, which would help the policy makers to adopt evidence-based public health action. In such a scenario, accurate estimates of disease burden cannot be overemphasized given the resource constraints in developing nations like ours. Currently POCDs are increasingly being used in community surveys to provide state wise estimates of anemia prevalence in different patient subgroups in India [11–13].

**CONTACT** Preetam B. Mahajan  [preetam3879@gmail.com](mailto:preetam3879@gmail.com)  Department of Community Medicine, Pondicherry Institute of Medical Sciences, Puducherry 605014, India

\*Department of Community Medicine, Pondicherry Institute of Medical Sciences, Puducherry 605014, India

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

[www.tandfonline.com/ijda](http://www.tandfonline.com/ijda)

In this article, we present the accuracy and reliability parameters of POCDs like Hemocue 201 (HC<sub>201</sub>) and Hemoglobin colour scale (HCS) with reference to automated laboratory analyser sysmex XP 100 (Lab<sub>SXP</sub>) [14]. The intended use of these POCDs could include screening, diagnosis, or surveillance. We discuss the determinants of accuracy parameters and certain design issues while ascertaining these parameters for its judicious use in both clinical context and disease burden studies.

## Patient and methods

The institute has a field practice area in Tangi block, Khorda district, in the state of Odisha (eastern India) that has a population of 0.16 million and has six sectors each with a primary health centre. The study was carried out in its largest sector Bhusandapur that had a total population of approximately 56,000. It was decided to carry out a community-based cross-sectional study to determine anemia prevalence (using Hemoglobin values) among apparently healthy males and apparently healthy nonpregnant females each, in two age groups, that is, 10–19 years and 20–60 years, who were willing to participate in the study, provide informed consent, and not suffering from any major illness at the time of the study. These same patients were subjected to testing on selected POCDs to obtain accuracy parameters and study spectrum effects.

Assuming anemia prevalence of 75–78% among females, 24% among males, while accounting for 7–10% nonresponse, and design effect of 1.5 [15], the sample size estimate for each sex group between 10–19 years and 20–60 years, respectively, would be 450 each. The sample size was powered to answer the diagnostic accuracy question reported in this article. All included subjects were tested using HC<sub>201</sub>, and HCS and compared with Lab<sub>SXP</sub>.

Cluster sampling design was chosen to select 30 villages (cluster) using probability proportionate to size methodology and 15 subjects were chosen randomly from the respective age and sex group amounting to 60 people from each village. A bottle was spun in the centre of the village to select a random direction. Last house in that direction was selected as starting point and every subsequent house along the way, as investigators moved backwards towards the centre was visited to recruit eligible participants from each stratum. If more than one person was found eligible in a given household, KISH grid was used to select the person randomly. Study was carried out over 9 months between May 2017 and Feb 2018.

The study participants were simultaneously tested on HCS and HC<sub>201</sub> (Index test) using capillary blood in that sequence to avoid HCS interpretation being affected by the Hemoglobin (Hb) results displayed on HC<sub>201</sub> equipment. Capillary blood was collected following wicking method and using special lancets recommended for use with HC<sub>201</sub>. Later 5 ml of venous blood was drawn, within 5 min of performing the index test, and transferred in EDTA vial to central laboratory maintaining the cold chain (between 2 and 8 °C) during transit, for testing on a three-part differential instrument

SYSMEX XP 100 autoanalyzer (reference test) available in the institute at the time of the study and determine the true Hb value. The average time lag between collection of blood sample from the field and its testing in the laboratory (reference test) was about 4 h. There were no adverse events from performing either the index or the reference tests. The output on reference test was automatic and, thus, did not interfere with fact that assessor had access to index test results. We, thus, ensured that none of the ratings got influenced by performance on any of the tests. A single trained Laboratory technician was utilised for testing all the participants to eliminate interrater variability.

## Data analysis

Individuals were classified as anaemic based on WHO cut-offs for Hb for age and sex [16]. Same cut-offs were applied to values on index and reference tests to differentiate anaemic from nonanaemic individuals. We calculated accuracy statistics like sensitivity, specificity, Likelihood ratios positive and negative (LR+, LR–) using standard definitions. We plotted receiver operating characteristics (ROC) curve to determine the area under curve (AUC) for both the POCDs and also establish a suitable cut-off for the HCS method.

Reliability of POCDs was determined by studying the degree of homogeneity between the Hb measurements by utilizing the intraclass correlation coefficient (ICC) for agreement and consistency. Bland–Altman plots accommodating the double measurements within subjects were used to capture bias and 95% limits of agreement.

All statistical analyses were performed using XLSTAT, Microsoft excel version 2016, and Epi-info version 7.2.2.6. The ICCs were interpreted as follows: <0, poor; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.

Any missing data was removed at the time of the analysis. Whenever there was any error during measurement on HC<sub>201</sub>, the procedure was repeated one more time after trouble shooting the cause of error following standard guidelines. If error persisted, results were declared inconclusive. Every attempt was made to identifying the closest match while using HCS.

## Quality control

The institute has a standard operating procedure for maintaining quality while processing samples on SYSMEX XP100. These guidelines were followed right from the time sample was collected, until it was processed. Six months internal quality control data is being used to calculate routine imprecision and updated annually. Stable controls are procured from commercial sources and data is plotted on control charts (L.J. charts). Laboratory uses 2 levels of controls at least once a day. For Hemocue 201+, the cuvettes are stored as per recommendations and standard techniques are used for periodic cleaning of the cuvette holder and optical sensor to ensure error free results.

### Ethical considerations

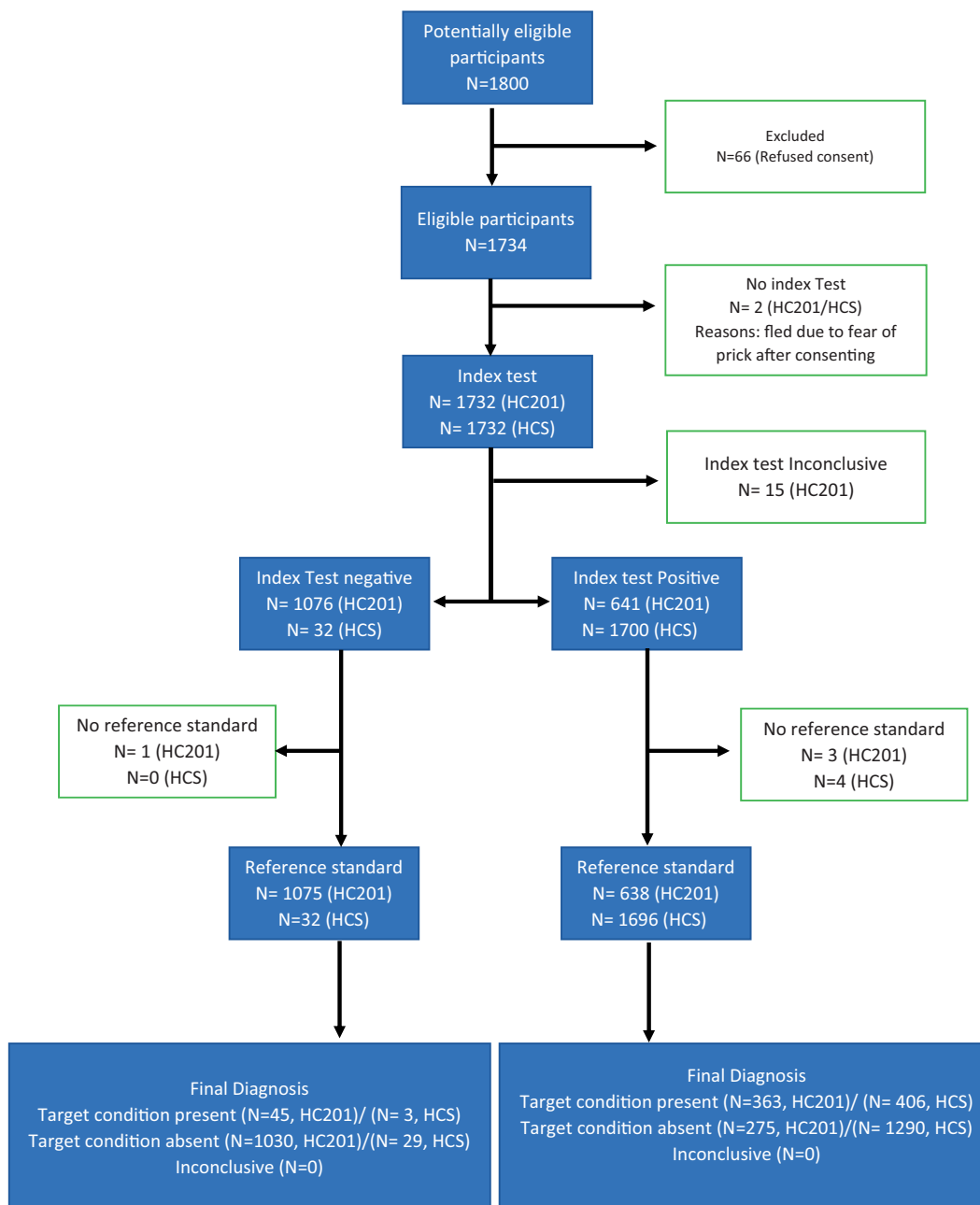
Study was approved by the Institute ethics committee (Project code-IMF/27/2016). Informed consent was obtained and confidentiality was maintained in the Department of Community and Family Medicine. We observed the ethical principles suggested by Indian Council of Medical Research [17] throughout the conduct of the study.

### Results

Overall 1732 participants were subjected to index tests. HC<sub>201</sub> was inconclusive in 15 and reference tests results was not available in 4 due to withdrawal of consent for intravenous blood (Figure 1). We had 408 males and 430 females in

the age group 10–19 years, while 413 males and 481 females in the age group 20–60 years (Figure 2). It was difficult to recruit adult males as most were working. Adult females were overrecruited due to their request that could not be denied in community settings.

Hemoglobin was normally distributed in both sexes and age groups that underwent reference test (Figure 3). The mean (SD) Hb values for adolescent boys, adolescent girls, adult males, and adult females were 13.23 g/dl (1.23), 12.47 g/dl (1.28), 14.36 g/dl (1.53), and 12.07 g/dl (1.52), respectively, and the difference was statistically significant. Overall in all samples combined 3.3% subjects had Hb <10 g/dl, 19% had Hb ≥10 and <12 g/dl, 27% with Hb from 12 g/dl to <13 g/dl, 40% between 13 g/dl and <15 g/dl, rest above 15 g/dl.



**Figure 1.** STARD diagram reporting flow of participants through the study comparing HC<sub>201</sub>, HCS with SYSMEX XP100 for Hb estimation in rural population of Odisha, India 2017.

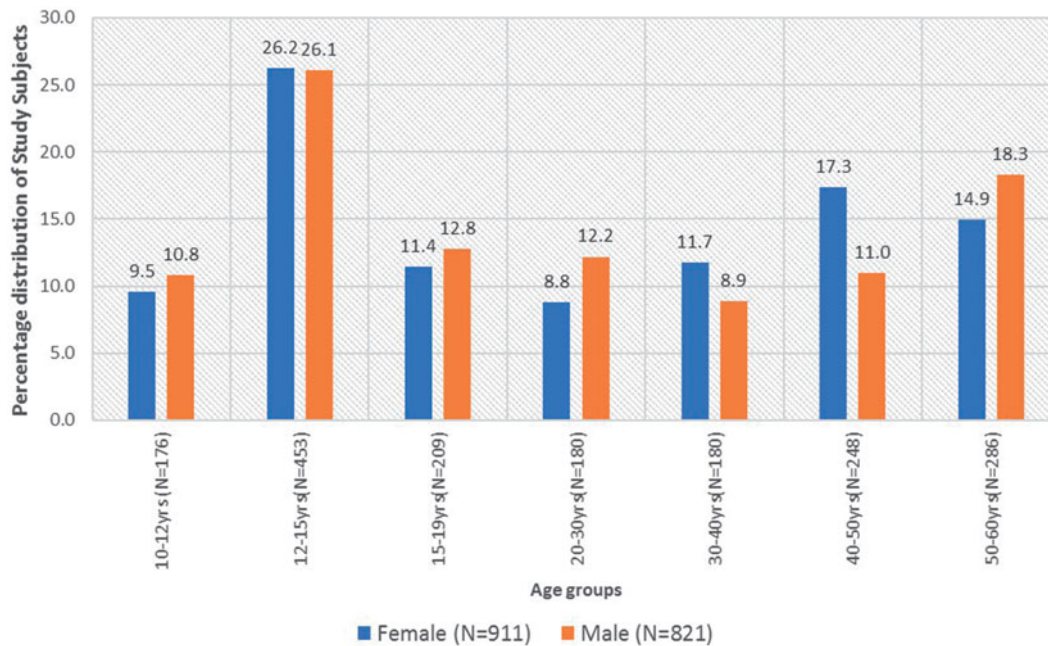


Figure 2. Age and sex distribution of subjects undergoing diagnostic accuracy testing in Tangi Rural Anemia diagnostic accuracy study, 2017.

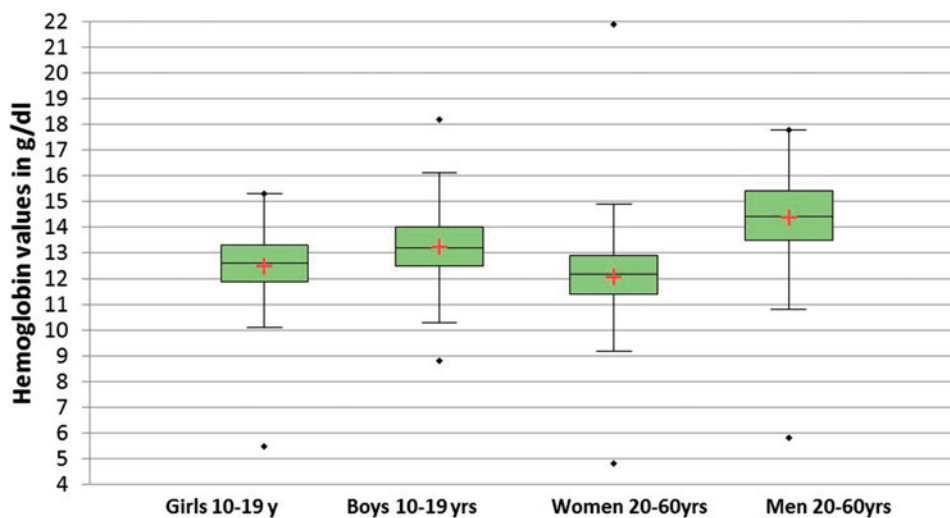


Figure 3. Box plots of population subgroups depicting distribution of true Hb values as determined by reference standard.

Tables 1 and 2 provides accuracy statistics for both the index tests. At the WHO clinical cut-off for anemia, HCS provides very poor specificity of 2.2% that could lead to massive false positive diagnosis. On the contrary HC<sub>201</sub> exhibits sensitivity and specificity of 89% and 79%, respectively, that would lead in missing 11% cases and over diagnosing 21% cases. True anemia proportion in the overall sample was 23.82%. With the calculated sensitivity and specificity of HC<sub>201</sub> and HCS for the overall sample, the estimated proportion (i.e. all test positives with index tests contributing to the prevalence calculation in this scenario) of anemia within the sample would increase by 13.4% and 74% points above the true value thus overestimating the true prevalence suggested by the reference test.

Most (79.22%) of the HC<sub>201</sub> results fell within 1 g/dl of the sysmex values, 13.54% between  $-1$  g/dl and  $-2$  g/dl, 3.85% g/dl  $> -2$  g/dl, and 3.39%  $> 1$  g/dl. Similarly, few (2.92%) of the

HCS results fell within 1 g/dl of the sysmex values, 7.88% between  $-1$  g/dl and  $-2$  g/dl, 14.94% between  $-2$  g/dl and  $-3$  g/dl, While most (73.96%)  $< -3$  g/dl and 0.29%  $> 1$  g/dl.

Intraclass correlation for agreement (ICC-A) between HC<sub>201</sub> and SYSMEX XP100 was 0.903 (CI: 0.845–0.934), while ICC for consistency (ICC-C) was 0.917 (CI: 0.909–0.924). Average bias for HC 201 was  $-0.417$  (CI:  $-0.462$ ,  $-0.371$ ) and limits of agreement between  $-2.302$  and  $1.468$  (Figure 4).

While, ICC-A between HCS and SYSMEX XP100 was 0.24 (CI:  $-0.158$ – $0.604$ ), ICC-C was 0.677 (CI: 0.645–0.706) and average bias for HCS was  $-4.121$  (CI:  $-4.203$ ,  $-4.039$ ) with limits of agreement between  $-7.521$  and  $-0.720$  (Figure 5).

ROC curve analysis details are provided in Table 3 and plot in Figure 6. Since the bias for HCS was not acceptable, we attempted to define a new cut-off for the test. It appears that perhaps at below score six subjects could be defined as having anemia, and this would have specificity of 88.85%,



**Table 1.** Accuracy parameters of index test (HC<sub>201</sub>) among study participants in Tangi Rural Anemia diagnostic accuracy study, 2017.

Statistic	All samples (N = 1713)	Adolescent boys (N = 406)	Adolescent girls (N = 430)	Adult males (N = 398)	Adult females (N = 479)
HC <sub>201</sub>	Value (95% CI)	Value (95% CI)	Value (95% CI)	Value (95% CI)	Value (95% CI)
Sensitivity	88.97% (85.52–91.84)	80.39% (66.88–90.18)	89.11% (81.35–94.44)	83.61% (71.91–91.85)	92.82% (88.25–96.02)
Specificity	78.93% (76.61–81.11)	77.46% (72.76–81.71)	75.68% (70.68–80.22)	86.94% (82.87–90.35)	75.00% (69.54–79.93)
Positive likelihood ratio	4.22 (3.78–4.71)	3.57 (2.82–4.52)	3.66 (2.99–4.49)	6.4 (4.76–8.62)	3.71 (3.02–4.56)
Negative likelihood ratio	0.14 (0.11–0.18)	0.25 (0.14–0.44)	0.14 (0.08–0.25)	0.19 (0.11–0.33)	0.1 (0.06–0.16)
Disease PREVALENCE	23.82% (21.82–25.91)	12.56% (9.50–16.18)	23.49% (19.56–27.79)	15.33% (11.93–19.25)	40.71% (36.27–45.26)
Positive predictive value	56.90% (54.17–59.58)	33.88% (28.82–39.35)	52.94% (47.88–57.94)	53.68% (46.27–60.94)	71.83% (67.49–75.79)
Negative predictive value	95.81% (94.55–96.79)	96.49% (94.02–97.96)	95.77% (92.81–97.54)	96.7% (94.32–98.10)	93.83% (90.14–96.20)
Accuracy	81.32% (79.39–83.14)	77.83% (73.47–81.78)	78.84% (74.67–82.60)	86.43% (82.67–89.64)	82.25% (78.53–85.57)

**Table 2.** Accuracy parameters for HCS at different cut-offs among study participants in Tangi Rural Anemia diagnostic accuracy study, 2017.

	HCS at clinical cut-offs Value (95% CI)	HCS <9 Value (95% CI)	HCS <7 Value (95% CI)	HCS <5 Value (95% CI)
<b>HCS (N = 1728)</b>				
Sensitivity	99.27% (97.87–99.85)	71.88% (67.26–76.19)	22.98% (18.99–27.37)	2.44% (1.18–4.45)
Specificity	2.2% (1.48–3.14)	49.89% (47.15–52.62)	88.86% (87.03–90.50)	99.39% (98.81–99.74)
Positive likelihood ratio	1.01 (1.00–1.03)	1.43 (1.32–1.56)	2.06 (1.63–2.61)	4.03 (1.60–10.15)
Negative likelihood ratio	0.33 (0.10–1.09)	0.56 (0.48–0.66)	0.87 (0.82–0.92)	0.98 (0.97–1.00)
Disease prevalence	23.67% (21.68–25.75)	23.67% (21.68–25.75)	23.67% (21.68–25.75)	23.67% (21.68–25.75)
Positive predictive value	23.94% (23.73–24.15)	30.79% (29.09–32.54)	39.00% (33.60–44.69)	55.56% (33.18–75.88)
Negative predictive value	90.63% (74.75–96.93)	85.12% (82.92–87.08)	78.82% (77.86–79.74)	76.67% (76.38–76.95)
Accuracy	25.17% (23.14–27.29)	55.09% (52.71–57.46)	73.26% (71.11–75.34)	76.45% (74.37–78.43)
<b>Adolescent boys HCS (N = 406)</b>				
Sensitivity	100% (93.02–100.00)	70.59% (56.17–82.51)	13.73% (5.70–26.26)	0.00% (0.00–6.98)
Specificity	0.56% (0.07–2.02)	42.82% (37.61–48.15)	86.20% (82.17–89.61)	98.87% (97.14–99.69)
Positive Likelihood Ratio	1.01 (1.00–1.01)	1.23 (1.01–1.51)	0.99 (0.48–2.07)	0.00
Negative Likelihood Ratio	0	0.69 (0.44–1.07)	1.00 (0.89–1.13)	1.01 (1.00–1.02)
Disease prevalence	12.56% (9.50–16.18)	12.56% (9.50–16.18)	12.56% (9.50–16.18)	12.56% (9.50–16.18)
Positive Predictive Value	12.62% (12.54–12.71)	15.06% (12.69–17.79)	12.50% (6.41–22.96)	0
Negative Predictive Value	100%	91.02% (86.69–94.03)	87.43% (86.08–88.66)	87.31% (87.19–87.44)
Accuracy	13.05% (9.93–16.73)	46.31% (41.37–51.29)	77.09% (72.69–81.09)	86.45% (82.73–89.63)
<b>Adolescent girls HCS (N = 430)</b>				
Sensitivity	99.01% (94.61–99.97)	81.19% (72.19–88.28)	34.65% (25.46–44.77)	3.96% (1.09–9.83)
Specificity	0.91% (0.19–2.64)	21.58% (17.26–26.42)	77.81% (72.93–82.19)	98.78% (96.92–99.67)
Positive Likelihood Ratio	1 (0.98–1.02)	1.04 (0.93–1.16)	1.56 (1.12–2.18)	3.26 (0.83–12.79)
Negative Likelihood Ratio	1.09 (0.11–10.32)	0.87 (0.55–1.37)	0.84 (0.72–0.98)	0.97 (0.93–1.01)
Disease prevalence	23.49% (19.56–27.79)	23.49% (19.56–27.79)	23.49% (19.56–27.79)	23.49% (19.56–27.79)
Positive Predictive Value	23.47% (23.08–23.87)	24.12% (22.17–26.18)	32.41% (25.53–40.14)	50.00% (20.30–79.70)
Negative Predictive Value	75.00% (23.98–96.61)	78.89% (70.34–85.48)	79.50% (76.89–81.89)	77.01% (76.27–77.74)
Accuracy	23.95% (19.99–28.28)	35.58% (31.05–40.31)	67.67% (63.03–72.08)	76.51% (72.21–80.44)
<b>Adult males HCS (N = 412)</b>				
Sensitivity	98.39% (91.34–99.96)	33.87% (22.33–47.01)	12.90% (5.74–23.85)	1.61% (0.04–8.66)
Specificity	3.71% (1.99–6.27)	89.43% (85.72–92.45)	100.00% (98.95–100.00)	100.00% (98.95–100.00)
Positive Likelihood Ratio	1.02 (0.98–1.06)	3.20 (2.02–5.09)	0	–
Negative Likelihood Ratio	0.43 (0.06–3.26)	0.74 (0.62–0.89)	0.87 (0.79–0.96)	0.98 (0.95–1.02)
Disease prevalence	15.05% (11.74–18.87)	15.05% (11.74–18.87)	15.05% (11.74–18.87)	15.05% (11.74–18.87)
Positive Predictive Value	15.33% (14.84–15.83)	36.21% (26.33–47.40)	100.00%	100.00%
Negative Predictive Value	92.86% (63.39–98.99)	88.42% (86.42–90.15)	86.63% (85.48–87.70)	85.16% (84.75–85.56)
Accuracy	17.96% (14.37–22.02)	81.07% (76.95–84.74)	86.89% (83.25–90.00)	85.19% (81.39–88.48)
<b>Adult females HCS (N = 480)</b>				
Sensitivity	99.49% (97.18–99.99)	79.49% (73.13–84.92)	22.56% (16.90–29.08)	2.56% (0.84–5.88)
Specificity	3.86% (1.94–6.80)	42.81% (36.99–48.78)	91.23% (87.32–94.24)	100.00% (98.71–100.00)
Positive Likelihood Ratio	1.03 (1.01–1.06)	1.39 (1.23–1.57)	2.57 (1.63–4.06)	–
Negative Likelihood Ratio	0.13 (0.02–1.02)	0.48 (0.35–0.65)	0.85 (0.78–0.92)	0.97 (0.95–1.00)
Disease prevalence	40.62% (36.20–45.17)	40.62% (36.20–45.17)	40.62% (36.20–45.17)	40.62% (36.20–45.17)
Positive Predictive Value	41.45% (40.84–42.07)	48.74% (45.67–51.82)	63.77% (52.73–73.52)	100.00%
Negative Predictive Value	91.67% (58.88–98.83)	75.31% (69.17–80.57)	63.26% (61.29–65.19)	60.00% (59.45–60.55)
Accuracy	42.71% (38.24–47.27)	57.71% (53.15–62.17)	63.33% (58.85–67.65)	60.42% (55.88–64.82)

sensitivity 22.93%, LR+ of 2.04 and LR– 0.87. However, accuracy statistics at various cut-offs are shown in Table 2.

## Discussion

POCDs in Hb estimation has utility in both clinical practice and public health surveillance. Inaccurate tests will not

only affect treatment decisions but also provide inconsistent estimates of disease burden in the community. Presence of bias can lead to incorrect estimates of accuracy, while presence of source of variation would result in differences in true diagnostic accuracy across studies, thus limiting applicability to a specific clinical or public health question.

Number of biases and sources of variations have been identified [18–21] and one way of reducing it is by following recognized standards while designing such studies [22–24]. In our study, we examined the performance of POCDs like HC<sub>201</sub> and HCS with respect to its accuracy and reliability in four different carefully selected population subgroups. We also computed these parameters in a combined sample ( $N=1713$ ; Tables 2 and 3) to demonstrate the presence of spectrum effects, further emphasizing the problems of spectrum and bias as suggested by Ransohoff and Feinstein [19]. Our patients (Figures 2 and 3) were enrolled randomly using a cross sectional study design that included a broad spectrum of Hb values distributed normally in each age and sex group and these groups were representative of similar population groups encountered in primary care settings, as well as population-based surveys. This would minimize both selection and spectrum bias. Interpretations of every test were carried out independent of each other to tackle information bias. LAB<sub>SXP</sub> values had a variation coefficient of 1.5% or less for Hb, when peripheral blood or control blood was analyzed more than 10 times continuously, minimizing possibility of classification bias due to reference test. Chances of disease progression bias were nil as samples for testing by test and reference methods were collected simultaneously, and every patient were subjected to both the tests eliminating incorporation bias or work up bias.

It is generally expected that a diagnostic device provides accurate estimates of underlying clinical trait; however, in real conditions we seldom achieve this due to disagreement between the measured value and underlying true value in addition to the reasons mentioned earlier. However, a near perfect agreement is always desirable. In Hb estimation, test methods are acceptable as long as at-least 95% of test values are within  $\pm 1$  g/dl of reference values. A plot of differences between the methods against their mean helps to examine this expectation [25]. In present study, the average bias of HC<sub>201</sub> was  $-0.417$  g/dl (Figure 4), however the bias was not consistent throughout the Hb range (Figure 7). It was  $-0.198$  g/dl,  $-0.276$  g/dl, and  $-0.467$  g/dl for true Hb values  $\leq 10$  g/dl,  $\leq 12$  g/dl,  $> 12$  g/dl, respectively. This can particularly lead to misclassification bias, if the distribution of Hb values in other study samples differ and thus limit applicability. Misclassification is more likely to result among patients with values around the clinical cut-offs used for defining anemia. The distribution of Hb values were examined in four study groups (Figure 3) and were found to be statistically different based on Kruskal–Wallis test ( $p < .0001$ ). We have thus provided group specific estimates of accuracy parameters to inform applicability in respective groups in other settings.

The estimates of accuracy parameters for HC<sub>201</sub> differed in each group (Table 1) due to variation in age/sex distribution, disease severity etc. However, since we have specified the spectrum and provided indices of accuracy in pertinent clinical subgroups, this will enhance its clinical applicability in similar patient groups. If we were to recommend use of accuracy parameters derived from the entire 1732 sample then spectrum effect (a phenomenon wherein the accuracy parameters differ in subgroups) would come into picture. The term was first introduced by Ransohoff and Feinstein

[19]. Whether spectrum effect leads to bias depends on its effects on the posttest probabilities and more importantly if this results into differences in clinical decisions [26]. A number of methods have been described that could help distinguish spectrum effects from spectrum biases that result in variability of medical diagnostic test performance [26,27].

As expected in males the specificity was higher and in females' sensitivity was higher, this being due to shift of mean values of Hb on either side of the spectrum. Except for the adult males with LR+ of 6.4, rest LR+ were comparable. LR– varied across the age and sex groups in comparison to overall sample. Ratio of LR+ in various subgroups with that of the overall sample (from Table 1) suggest marked deviation among adult males. While, ratio of subgroup specific LR– estimates with that of overall sample indicate deviation in all age and sex groups except adolescent girls.

Spectrum bias will result in clinically important differences in post-test probabilities under two circumstances: (1) presence of bias in the ratio of LR+ and prevalence of disease is low, and (2) when bias exist in ratio of LR– and prevalence of disease is high [26]. As per our findings (Table 1), such a situation exists among adult males and females (This is if we were to combine the whole sample instead of looking at subgroup specific estimates). Thus, spectrum effect is likely to result into spectrum bias in these subgroups, unless we use subgroup specific accuracy parameters in decision making. This further strengthens the argument that only those spectrum of patients' needs to be included in diagnostic accuracy studies that will be similar to the population in which the test will be used in practice [23].

While for purpose of clinical use it is always desirable to have very few false negatives and positives, we need to realize that the accuracy parameters estimated would also depend on the composition of study sample, besides measurement errors across Hb range. As a rule, for rare disease, positive test results should be cautiously interpreted among subgroups who differ from the sample used to determine accuracy parameters. The same caution applies to negative test results when the disease is more common [28].

Thus, variations in accuracy parameters are known to occur due to differences in demographic features of study populations, disease prevalence, disease severity, interobserver variability, availability of clinical information, test technology, test execution, and so forth [18]. In our study, sensitivity increased while specificity decreased with rise in disease prevalence (Table 1). Similar findings have been reported in other disease conditions too [18,29]. With increasing anemia prevalence, the spectrum of Hb values would shift leftwards increasing numbers having more severe disease. This in turn would result in rise in sensitivity. Our findings provides further evidence challenging the popular belief that sensitivity, specificity, likelihood ratios are independent of disease prevalence [30]. In fact, these parameters strongly depend on the distribution of the underlying traits relative to the diagnostic cut-point, which vary between populations or between subgroups within the same population and thus have important practical implications.

We further move our attention to how this would influence the disease burden estimates made by using POCDs

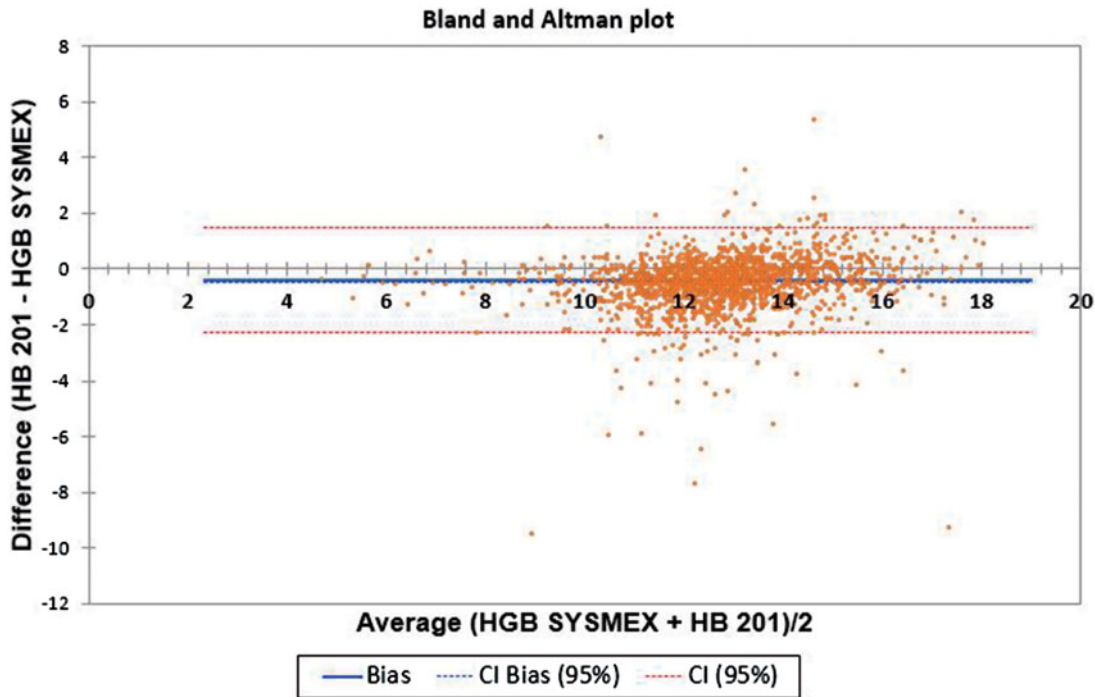


Figure 4. Bland Altman plot comparing HC201 with SYSMEX XP100 for Hb estimation in rural population of Odisha, India 2017.

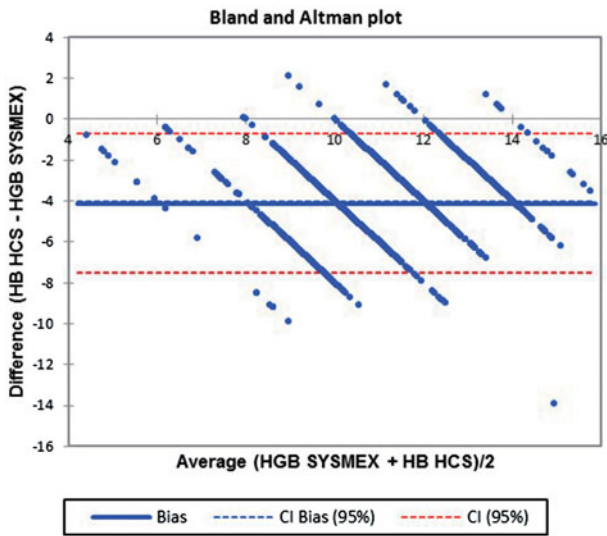


Figure 5. Bland Altman plot comparing HCS with SYSMEX XP100 for Hb estimation in rural population of Odisha, India 2017.

that have varying accuracy estimates based on the factors discussed so far. Measures of disease frequency like prevalence, incidence typically represent the number of people who were tested positive for a diagnostic test rather than those who actually had the disease. Unless the false positives are equal to false negatives the estimates are unlikely to be accurate. If the disease prevalence is lower than 50%, specificity of the test needs to be greater than sensitivity. Conversely for diseases with higher prevalence, sensitivity has to be higher to obtain accurate estimates. So, the solution lies in either using appropriate diagnostic tests for a given true prevalence or estimating the necessary test statistic and provides corrected estimates. This has been nicely illustrated by Campbell et al. [31].

Table 3. ROC analysis parameters during comparison of HCS and HC<sub>201</sub> for Hb estimation in rural population of Odisha, India 2017.

POCDs	Proposed cut-off with maximum accuracy	Accuracy at proposed cut-off	AUC	Proportion of anemia cases in subsample undergone paired testing
Hemocue 201 (n = 1713)	11.4	0.852	0.887 (0.865–0.908)	23.82%
Hemoglobin Colour Scale (HCS) (n = 1713)	6	0.762	0.64 (0.612–0.669)	23.82%

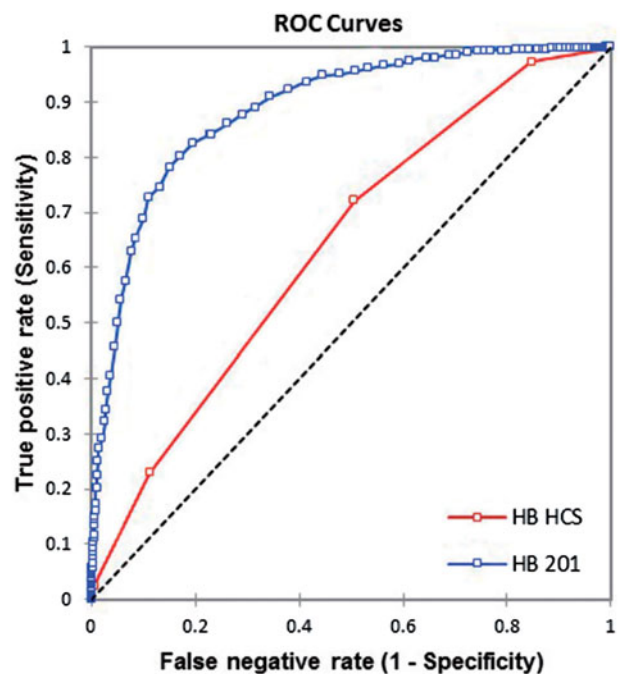
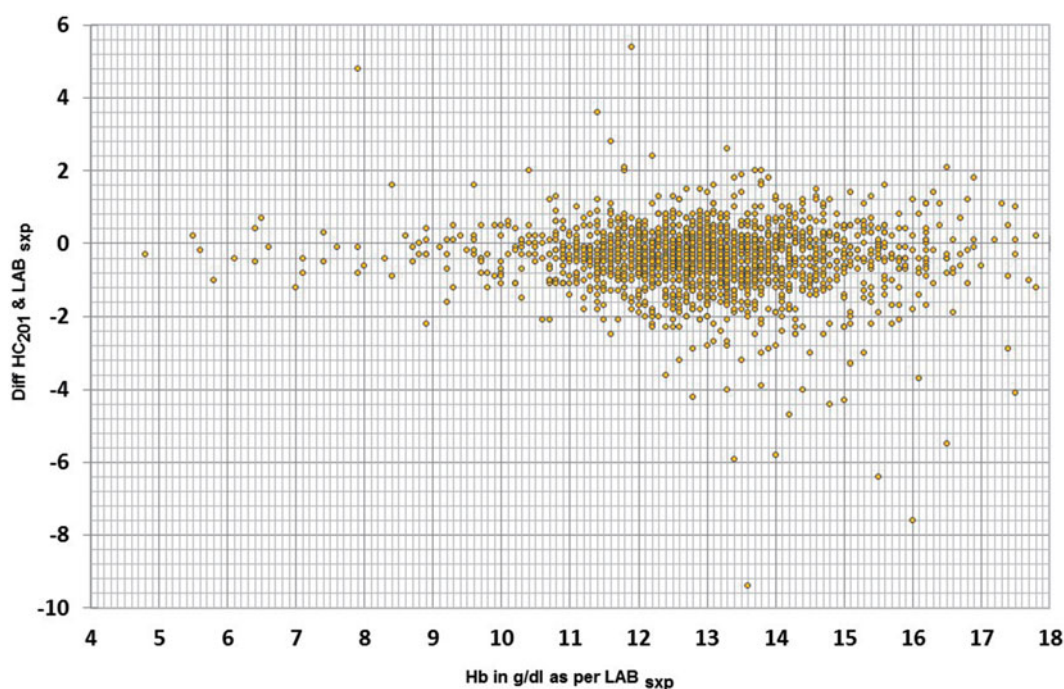


Figure 6. ROC curve comparing HCS with HC<sub>201</sub> for Hb estimation in rural population of Odisha, India 2017.



**Figure 7.** Scatter of difference of Hb values estimated by HC201 and Reference test against true Hb values provided by the reference test.

With regards to HCS we would suggest a different approach. The values on the scale are expected to represent the true underlying Hb levels. But the bias obtained in the study was not clinically acceptable. So, we decided to run a ROC analysis instead and suggest a cut-point to distinguish people with and without disease amenable to further testing or estimating disease burden as discussed earlier. Based on the findings in Table 2, it does not seem reasonable to use HCS at any cut-off in diagnosis and treatment of patients, though cut-off of  $<7$  for all the age and sex groups except adult males in whom cut-off  $<9$  could be used for disease burden estimation and appropriate corrections could be used to get the true picture. Choice of HC<sub>201</sub> or HCS for disease burden study would depend on availability of resources, though we would like to favour HC<sub>201</sub> considering the degree of bias, and accuracy parameters. HC<sub>201</sub> requires continuous supplies of cuvettes. These are often seen as limiting factors due to escalating operational costs of using POCDs. The prices of these cuvettes need to be regulated or subsidized for use in public health care system. Further studies are required to establish the cost-effectiveness of these POCDs and guide policy makers in their efforts to control and prevent anemia.

The strength of our study is that we have assembled a community-based sample to derive subgroups specific accuracy estimates. We have discussed the importance of taking into consideration the disease prevalence and population composition before applying the findings to people either in clinical settings or population studies. The revised cut-offs proposed for HCS for population-based surveys require further testing in different setting. Thus, being aware of reasons for variation in accuracy estimates for same machines in different settings, and the need for designing studies that could help estimate these parameters each time on a subsample as outlined by Campbell et al. [31] could go a long way in

efficiently using technology be it for guiding clinical decisions or public health actions.

## Transparency

### Declaration of funding

Intramural Research Grant No: IMF/27/2016 from AIIMS Bhubaneswar.

### Declaration of financial/other relationships

The author have no financial or other relationships to declare. JME peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

### Author contributions

PBM designed the study, collected data, analyzed and wrote the first draft and approved it. SM collected the data, analyzed it and revised the manuscript and accepted the final version.

### Acknowledgments

We would like to thank our Institute (AIIMS hubaneswar, Institute of National Importance under MoH&FW, Govt of India) for providing intramural grant (No: IMF/27/2016) to meet the cost of all reagents and consumables required for the conduct of the study as well other resources, such as printing and travel facilities to study site throughout the study duration. PBM was working as Associate Professor, Department of Community Medicine and Family Medicine at AIIMS Bhubaneswar at the time of submission of this manuscript.

## References

- [1] Mathers C, Stevens G, Ho J. WHO methods and data sources for global burden of disease estimates 2000–2011 [Internet]. Glob Heal. Estim Tech Pap. 2013;4:2013.



- [2] Indian Council of Medical research, Public Health Foundation of India, and Institute for Health Metrics and Evaluation. India: Health of the Nation's States- The India State-level Disease Burden Initiative. New Delhi, India: ICMR, PHFI, and IHME; 2017.
- [3] Guidelines for Control of Iron Deficiency Anaemia [Internet]. 2013 [cited 2018 Feb 23]. Available from: [http://www.pbnrhm.org/docs/iron\\_plus\\_guidelines.pdf](http://www.pbnrhm.org/docs/iron_plus_guidelines.pdf)
- [4] Butt Z, Ashfaq U, Sherazi SFH, et al. Diagnostic accuracy of "pallor" for detecting mild and severe anaemia in hospitalized patients. *J Pak Med Assoc.* 2010;60:762–765.
- [5] Kalantri A, Karambelkar M, Joshi R, et al. Accuracy and reliability of pallor for detecting anaemia: a hospital-based diagnostic accuracy study. *PLoS One.* 2010;5:e8545.
- [6] Chalco JP, Huicho L, Alamo C, et al. Accuracy of clinical pallor in the diagnosis of anaemia in children: a meta-analysis. *BMC Pediatr.* 2005;5:46.
- [7] Meda N, Dao Y, Touré B, et al. Assessing severe maternal anemia and its consequences: the value of a simple examination of the coloration of palpebral conjunctiva. *Sante.* 9:12–17.
- [8] Luby SP, Kazembe PN, Redd SC, et al. Using clinical signs to diagnose anaemia in African children. *Bull. World Health Organ.* 1995;73:477–482.
- [9] Kalter HD, Burnham G, Kolstad PR, et al. Evaluation of clinical signs to diagnose anaemia in Uganda and Bangladesh, in areas with and without malaria. *Bull World Health Organ.* 1997;75 (Suppl 1):103–111.
- [10] Stoltzfus RJ, Edward-Raj A, Dreyfuss ML, et al. Clinical pallor is useful to detect severe anemia in populations where anemia is prevalent and severe. *J Nutr.* 1999;129:1675–1681.
- [11] District Level Household & Facility Survey .... [Internet]. [cited 2018 Mar 4]. Available from: <http://rchiips.org/DLHS-4.html>
- [12] National Family Health Survey [Internet]. [cited 2018 Mar 4]. Available from: <http://rchiips.org/NFHS/index.shtml>
- [13] Annual Health Survey: Clinical, Anthropometric & Bio-chemical (CAB) Survey | Open Government Data (OGD) Platform India [Internet]. [cited 2018 Mar 4]. Available from: <https://data.gov.in/catalog/annual-health-survey-clinical-anthropometric-bio-chemical-cab-survey>
- [14] XP Series | Sysmex India – Hematology Instrumentation [Internet]. [cited 2018 Mar 9]. Available from: <http://www.sysmex.co.in/products/diagnostics/hematology/xp-series/>
- [15] Sharman A. Anemia testing in population-based surveys general information and guidelines for country monitors and program managers [Internet]. 2000; [cited 2018 Feb 24]. Available from: <https://dhsprogram.com/pubs/pdf/od22/od22.pdf>
- [16] WHO. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and mineral nutrition information system. Geneva: World Health Organization; [Internet]. 2011; (WHO/NMH/NHD/MNM/11.1) [cited 2018 Feb 24]. Available from: <http://www.who.int/vmnis/indicators/haemoglobin.pdf>
- [17] Mathur R, editor. National ethical guidelines for biomedical and health research involving human participants. New Delhi: Director-General Indian Council of Medical Research; 2017; [cited 2018 Mar 20]. Available from: [http://icmr.nic.in/guidelines/ICMR\\_Ethical\\_Guidelines\\_2017.pdf](http://icmr.nic.in/guidelines/ICMR_Ethical_Guidelines_2017.pdf)
- [18] Whiting PF, Rutjes AWS, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol.* 2013;66:1093–1104.
- [19] Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med.* 1978;299:926–930.
- [20] Schmidt RL, Factor RE. Understanding sources of bias in diagnostic accuracy studies. *Arch Pathol Lab Med.* 2013;137:558–565.
- [21] Roeber L. Types of bias in studies of diagnostic test accuracy. *Evid Based Med Pract.* 2016;22:e113.
- [22] Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
- [23] Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA.* 1995;274:645–651..
- [24] Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529–536.
- [25] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
- [26] Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med.* 2004;23:125–135.
- [27] Elie C, Coste J. A methodological framework to distinguish spectrum effects from spectrum biases and to assess diagnostic and screening test accuracy for patient populations: application to the Papanicolaou cervical cancer smear test. *BMC Med Res Methodol.* 2008;8:2018–2027.
- [28] Baron JA. Uncertainty in bayes. *Med Decis Mak.* 1994;14:46–51.
- [29] Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol.* 2009;62:5–12.
- [30] Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med.* 1997;16:981–991
- [31] Campbell H, Biloglav Z, Rudan I. Reducing bias from test misclassification in burden of disease studies: use of test to actual positive ratio—new test parameter. *Croat Med J.* 2008;49:402–414.