Differential Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia: A Diagnostic Test Accuracy Meta-Analysis

Mina Jahangiri 1, Fakher Rahim 2,3, Amal Saki Malehi 1
Seyed Mohammad Sadegh Pezeshki 2, Mina Ebrahimi 2

1. Department of Biostatistics and Epidemiology, Faculty of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
2. Thalassemia and Hemoglobinopathy Research Center, Research Institute of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
3. Metabolomics and Genomics Research Center, Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran

KEYWORDS
- Iron deficiency anemia
- Thalassemia
- Systematic review
- Meta-analysis
- Diagnostic test accuracy (DTA)

ABSTRACT
We evaluated the most common indices to compare their sensitivity and specificity to introduce the most sensitive and specific index. We systematically searched five international indexing databases up to Dec 2018. For each index, we measured the diagnostic odds ratio (DOR), as well as summary ROC (SROC) curve which was used to compare the performance of each index. Deeks’ tests of all discriminant indices indicated that there is no potential publication bias. The area under curves (AUCs) of all discriminant indices indicate overall good differential performance. The M/H ratio index was more sensitive and specific compared to other studied indices. In this meta-analysis, the M/H ratio index was more potential to discriminate iron deficiency anemia (IDA) from thalassemia trait. However, we cannot use this index alone to achieve the final diagnosis. The capability of this index to discriminate IDA from thalassemia trait must be used alongside with the common laboratory procedure to ensure the final differentiation.

Introduction
Anemia is a common hematologic disorder and its lack of attention is associated with several physical and neurological complications [1]. On the other hand, its diagnosis and treatment are not very costly. The most common type of anemia is the microcytic anemia, of which two most common forms include Iron deficiency anemia (IDA) and β-thalassemia minor [2]. IDA is a common complication all across the world. Despite its high prevalence, there is still some uncertainty in diagnosis steps [3]. Thalassemia is the most common monogenic disorder which is highly prevalent in the Mediterranean, Middle East, Indian peninsula and southeast of Asia. Nowadays, in the wake of extensive immigration in the past decades, thalassemia can be seen in almost everywhere [4]. Both IDA and thalassemia trait are among microcytic hypochromic anemias. In this type of anemia, globin chain synthesis disorders and heme
2 Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia….

synthesis disorders cause microcytic and hypochromic red blood cells (RBCs), respectively [5]. Discriminating IDA and thalassemia trait in patients is a major challenge due to the resemblance of the anemic view. Evaluation of different hemoglobin levels by hemoglobin electrophoresis alongside with routine complete blood count (CBC) test and iron panel tests (such as ferritin level) are currently the main tools to discriminate IDA from thalassemia trait [6]. Despite their great utility, all routine hematologic tests such as HbA2, examination of a peripheral blood film, serum ferritin, iron, TIBC, and transferrin saturation are either not available in all clinical setups, or these are relatively time-consuming and expensive techniques. Consequently, using these indices will save time and reduce diagnostic expenses [7, 8].

Due to the importance of discriminating these two complications in patient management and considering the financial limitations specifically in countries with high prevalence of thalassemia, mathematical indices, which are simpler and less complicated solutions, have been used to achieve a differential diagnosis. Despite the introduction of a considerable number of these indices in the past years, none of them are fully sensitive and specific to discriminate IDA from thalassemia trait. In this meta-analysis, we will evaluate the most common indices to compare their sensitivity and specificity in order to introduce the most sensitive and specific index.

### Table 1. Discriminant indices for differentiation between thalassemia trait and iron deficiency anemia with their cut off values and references of publications based on these discriminant indices.

<table>
<thead>
<tr>
<th>Discriminant Index</th>
<th>Formula</th>
<th>Cut off value</th>
<th>Number of studies</th>
<th>Number of patients</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>England &amp; Fraser (E&amp;F) [31]</td>
<td>MCV− RBC−5xHB−3.4</td>
<td>&lt; 0 &gt; 0</td>
<td>81</td>
<td>26091</td>
<td>[10, 12, 29, 32-90]</td>
</tr>
<tr>
<td>Green &amp; King (G&amp;K) [91]</td>
<td>(MCV×RDW)/(100×HB)</td>
<td>&lt; 65 &gt; 65</td>
<td>53</td>
<td>18508</td>
<td>[10, 12, 13, 15-17, 19-24, 26-30, 32, 34, 46-48, 55, 58, 59, 63, 68, 69, 71-83, 85-90, 92-96]</td>
</tr>
<tr>
<td>Bessman (RDW) [106]</td>
<td>RDW</td>
<td>&lt; 14 &gt; 14</td>
<td>52</td>
<td>15208</td>
<td>[10, 12, 14, 16, 17, 19-21, 26, 28, 39, 41, 44-48, 50-52, 56-59, 61, 63, 68, 69, 71-77, 80, 82, 83, 86, 88, 94, 98, 100, 104, 105, 107-111]</td>
</tr>
<tr>
<td>Jayabose (RDWI) [99]</td>
<td>(MCV × RDW)/RBC</td>
<td>&lt; 220 &gt; 220</td>
<td>34</td>
<td>13916</td>
<td>[10, 12, 13, 15-17, 19-21, 26-30, 68, 71-73, 75-83, 85-88, 92, 96, 112]</td>
</tr>
<tr>
<td>Ricerca [98]</td>
<td>RDW/RBC</td>
<td>&lt; 4.4 &gt; 4.4</td>
<td>30</td>
<td>13639</td>
<td>[10, 12, 13, 15-17, 19-23, 26-29, 32, 59, 71, 72, 74, 75, 78-80, 82, 83, 85-89]</td>
</tr>
<tr>
<td>Shine &amp; Lal (S&amp;L) [113]</td>
<td>MCV^2 × MCH × 0.01</td>
<td>&lt; 15 &gt; 15</td>
<td>56</td>
<td>21685</td>
<td>[10-17, 19-23, 25-30, 32, 33, 39, 40, 43, 45-48, 50, 55, 60, 61, 65-68, 70-72, 74-76, 78-80, 82, 84-87, 89, 92, 93]</td>
</tr>
<tr>
<td>M/H Ratio [32]</td>
<td>Microcytic RBC % / hypochromic RBC %</td>
<td>&lt; 3.7 &gt; 3.7</td>
<td>15</td>
<td>3091</td>
<td>[22, 32-34, 59, 63, 89, 90, 103, 115-119]</td>
</tr>
</tbody>
</table>

### Material and Methods

This systematic review and meta-analysis was designed according to the PRISMA guideline [120]. The search strategy in this study was based on the search strategy used by Hoffmann et al. [112].

#### Literature search

To find relevant published papers, we searched six international indexing databases, including PubMed/Medline, ISI web of science, Cochrane central, ProQuest, Embase, and Scopus up to Dec 2018 systematically. We included the following search terms in the analyses: “microcytic”, “iron deficiency”, “anaemia” or “anaemia” or “iron-deficiency” and/or “thalassemia” AND “distinguish or differentiate or discriminant”, and all possible combinations. We did not restrict ourselves regarding the language of the reports and occasionally used the help of a translator.
Inclusion criteria

Two authors (FR and MJ) independently did the initial screening (title and abstract) for inclusion. We identified the original publications of all discriminant indices and searched for the publications citing them. Finally, we perused the literature reference lists in the publications found above for references not yet covered. Other publication types such as the letter, opinion, systematic review or meta-analysis etc. were excluded.

Quality assessment

The quality of published papers was assessed independently by two authors (FR and ASM) using two 12-item and 3-item tools called the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) [121], and the Standards for Reporting Studies of Diagnostic Accuracy (STARD), respectively [122].

Statistical analysis

We extracted TP (True Positive), FP (False Positive), FN (False Negative) and TN (True Negative) values from diagnostic 2-by-2 tables of published studies for calculating and pooling accuracy measures such as summary sensitivity (True Positive Rate), specificity (True Negative Rate), positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR) and area under curve (AUC). The sensitivity and specificity are defined as the probability of positive/negative test results in the subjects with/without the disease, respectively. In this study, the sensitivity and specificity present probability of thalassemia trait/iron deficiency anemia test results in the subjects with/without thalassemia trait, respectively.

PLR

PLR indicates that how many times more likely positive (thalassemia trait) test results in the subjects with the disease (subjects with thalassemia trait) versus the subjects without the disease (subjects with iron deficiency anemia). Also, NPR describes that how many times more likely negative (iron deficiency anemia) test results in the subjects with the disease (subjects with thalassemia trait) versus the subjects without the disease (subjects with iron deficiency anemia). PLR > 10 and NLR < 0.1 indicate a good differential diagnosis [123]. DOR (PLR/NLR) is equal to the ratio of the odds of positivity (thalassemia trait) in the subjects with disease (thalassemia trait) relative to the odds positivity (thalassemia trait) in the subjects without the disease (iron deficiency anemia) and discriminant index with higher DOR value show better diagnostic differential performance [124]. SROC (summary receiver operating characteristic) curve provides AUC as a global summary of discriminant index accuracy in discriminating between thalassemia trait and iron deficiency anemia and is drawn for each discriminant index. AUC represents an overall performance measure of diagnostic classification and good diagnostic tests have AUC > 0.70 [123].

The heterogeneity across studies was assessed using inconsistency index ($I^2$ statistics) and Cochrane Q test (Chi-square). An $I^2$ statistic > 50% or $P < 0.1$ for Q test show substantial or significant heterogeneity among studies [125, 126]. If the heterogeneity among the studies was not substantial or significant, pooled summary accuracy would measure with their 95% confidence interval computed based on the fixed-effect model otherwise, the random-effect model is used [127, 128]. The bivariate generalized linear mixed model was also fitted to compare sensitivity and false positive rate of discriminant indices. By using a random effects model, this bivariate regression model incorporates any correlation between sensitivity and false positive rate of different discriminant indices [129]. Publication bias was assessed using Deeks' funnel plot and Deeks' test and a $P < 0.05$ indicates the presence of publication bias [130]. Data analysis was performed using two software: Meta-Disc version 1.4 (XI Cochrane Colloquium, Barcelona, Spain) [131] and R 3.0.3.

Results

A systematic literature search returned 102 articles reported the usage of 12 different indices that were investigated five or more times (Figure 1). The SROC curves with confidence region and AUCs of all discriminant indices have been measured (Supplementary).

Deeks's tests of all discriminant indices indicated that there is no potential publication bias (Table 2) and Deeks' funnel plots of all discriminant indices did not show any evidence of asymmetry (Deeks' funnel plots are not shown). SROC curves with confidence region are shown in Figure 2 and AUCs of all
Discriminant indices indicate overall good differential performance (Table 2). Consequently, the M/H ratio index has higher AUC in comparison with other indices. Bivariate regression model using Bessman (red cell distribution width = RDW) index as the reference category indicated that sensitivity of all indices is significantly higher than the sensitivity of Bessman (RDW) index ($P < 0.05$).

But the false positive rate of Bessman (RDW) index is significantly lower than Ricerra and Shine & Lal indices ($P < 0.001$) and also false positive rate of Bessman (RDW) index is lower than Srivastava index, but there was no statistically significant difference between the false positive rate of Bessman (RDW) index and Srivastava index ($P > 0.05$) (Table 3). The false positive rate of England & Fraser and Sirdah indices were significantly lower than Bessman (RDW) index ($P < 0.05$) and false positive rate of Ehsani, Green & King, Mentzer, RBC and Jayabose (RDWI) indices were lower than Bessman (RDW) index, but there was no statistically significant difference between the false positive rate of these indices and Bessman (RDW) index ($P > 0.05$) (Table 3).
Table 2. Diagnostic performance of discriminant indices with 95% confidence interval

<table>
<thead>
<tr>
<th>Discriminant Index</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
<th>DOR (95% CI)</th>
<th>AUC</th>
<th>Deeks’ Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ehsani [9]</td>
<td>0.82 (0.81,0.83)</td>
<td>0.81 (0.80,0.83)</td>
<td>4.57 (3.66-5.71)</td>
<td>0.19 (0.14,0.25)</td>
<td>27.93 (19.23-40.57)</td>
<td>0.905</td>
<td>0.87</td>
</tr>
<tr>
<td>England &amp; Fraser (E&amp;F) [31]</td>
<td>0.77 (0.76,0.78)</td>
<td>0.85 (0.84,0.86)</td>
<td>7.71 (6.14-9.7)</td>
<td>0.27 (0.20,0.37)</td>
<td>33.67 (24.83-45.65)</td>
<td>0.9135</td>
<td>0.26</td>
</tr>
<tr>
<td>Green &amp; King (G&amp;K) [91]</td>
<td>0.81 (0.80,0.82)</td>
<td>0.85 (0.84,0.86)</td>
<td>6.49 (4.91-8.59)</td>
<td>0.2 (0.14,0.28)</td>
<td>35.35 (23.57-53.02)</td>
<td>0.919</td>
<td>0.82</td>
</tr>
<tr>
<td>Mentzer [97]</td>
<td>0.81 (0.81,0.82)</td>
<td>0.83 (0.82,0.84)</td>
<td>4.76 (4.12-5.50)</td>
<td>0.21 (0.18,0.26)</td>
<td>25.68 (19.95-33.08)</td>
<td>0.906</td>
<td>0.62</td>
</tr>
<tr>
<td>RBC [36]</td>
<td>0.83 (0.81,0.84)</td>
<td>0.83 (0.82,0.84)</td>
<td>6.14 (4.43-8.49)</td>
<td>0.2 (0.15,0.27)</td>
<td>35.46 (23.49,53.53)</td>
<td>0.921</td>
<td>0.58</td>
</tr>
<tr>
<td>Bessman (RDW) [106]</td>
<td>0.63 (0.62,0.64)</td>
<td>0.63 (0.62,0.65)</td>
<td>2.16 (1.76,2.65)</td>
<td>0.45 (0.33,0.62)</td>
<td>6.1 (3.63,10.27)</td>
<td>0.808</td>
<td>0.56</td>
</tr>
<tr>
<td>Jayabose (RDWI) [99]</td>
<td>0.84 (0.83,0.85)</td>
<td>0.84 (0.83,0.85)</td>
<td>5.13 (3.77,11)</td>
<td>0.2 (0.15,0.25)</td>
<td>28.91 (20.07,41.64)</td>
<td>0.912</td>
<td>0.39</td>
</tr>
<tr>
<td>Ricerca [98]</td>
<td>0.89 (0.89,0.90)</td>
<td>0.57 (0.56,0.59)</td>
<td>2.28 (1.79,2.89)</td>
<td>0.17 (0.12,0.22)</td>
<td>19.89 (14.07,28.12)</td>
<td>0.899</td>
<td>0.73</td>
</tr>
<tr>
<td>Shine &amp; Lal (S&amp;L) [113]</td>
<td>0.90 (0.90,0.91)</td>
<td>0.52 (0.51,0.53)</td>
<td>1.90 (1.55,2.32)</td>
<td>0.17 (0.13,0.23)</td>
<td>16.52 (10.40,26.24)</td>
<td>0.899</td>
<td>0.61</td>
</tr>
<tr>
<td>Sirdah [17]</td>
<td>0.79 (0.77,0.80)</td>
<td>0.86 (0.85,0.87)</td>
<td>7.71 (5.73,10.38)</td>
<td>0.25 (0.19,0.34)</td>
<td>39.30 (25.90,59.64)</td>
<td>0.933</td>
<td>0.70</td>
</tr>
<tr>
<td>Srivastava [114]</td>
<td>0.77 (0.76,0.78)</td>
<td>0.79 (0.78,0.80)</td>
<td>3.85 (2.74,5.2)</td>
<td>0.31 (0.26,0.36)</td>
<td>14.25 (10.92,18.60)</td>
<td>0.861</td>
<td>0.57</td>
</tr>
<tr>
<td>M/H Ratio [32]</td>
<td>0.92 (0.87,0.98)</td>
<td>0.86 (0.81,0.91)</td>
<td>6.8 (4.89,8.9)</td>
<td>0.07 (0.03,0.2)</td>
<td>100.8 (39.6,256.3)</td>
<td>0.956</td>
<td></td>
</tr>
</tbody>
</table>


Table 3. The result of analysis of data by bivariate generalized linear mixed model

<table>
<thead>
<tr>
<th>Discriminant Index</th>
<th>Sensitivity</th>
<th>False Positive Rate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Estimate</td>
<td>95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td></td>
<td>0.201</td>
<td>(-0.217,0.619)</td>
<td>0.346</td>
</tr>
<tr>
<td>Ehsani</td>
<td>1.566</td>
<td>(0.915,2.217)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>England &amp; Fraser (E&amp;F)</td>
<td>0.810</td>
<td>(0.289,1.331)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Green &amp; King (G&amp;K)</td>
<td>1.310</td>
<td>(0.767,1.854)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mentzer</td>
<td>1.333</td>
<td>(0.816,1.849)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>RBC</td>
<td>1.377</td>
<td>(0.809,1.945)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Jayabose (RDWI)</td>
<td>1.496</td>
<td>(0.910,2.082)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ricerca</td>
<td>2.374</td>
<td>(1.750,2.998)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Shine &amp; Lal (S&amp;L)</td>
<td>2.669</td>
<td>(2.097,3.242)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sirdah</td>
<td>1.069</td>
<td>(0.388,1.750)</td>
<td>0.002</td>
</tr>
<tr>
<td>Srivastava</td>
<td>0.916</td>
<td>(0.379,1.453)</td>
<td>0.001</td>
</tr>
<tr>
<td>M/H Ratio</td>
<td>1.8</td>
<td>(1.125)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Bessman (RDW) index is used as reference group. CI: Confidence Interval.
Fig. 2. SROC (summary receiver operating characteristic) curves with confidence region of discriminant indices.
Discussion

Microcytic anemia is one of the most common forms of anemia that physicians deal with during daily practice, and are often faced with the challenge of its managing. The differentiation of the two most common forms of microcytic anemia which include IDA and β-thalassemia minor is very important because, in addition to high prevalence, it is sometimes difficult to differentiate them; therefore patients suffer from loss of time and money. Today, screening of thalassemia minor has a special place in countries dealing with a huge number of families at risk of a severe form of the disease. Identifying minor thalassemia is also important because the treatment of these patients with iron leads to increase in mean corpuscular volume (MCV) levels. The main method of anemia diagnosis is a reduction in serum iron level, saturated transferrin, and ferritin levels with an increase in iron binding capacity. Although, the method of thalassemia minor diagnosis is measuring HbA2 levels by means of hemoglobin electrophoresis test [132], in some cases of thalassemia minor, HbA2 levels do not rise and on the other hand, measurement of HbA2 levels or iron storages in the body are considered as time-consuming and costly methods. Moreover, there is a controversy regarding whether HbA2 levels provide a valid and reliable measure of diagnosis of β-thalassemia trait in people with iron deficiency or not [133].

Nowadays, other low-cost and rapid calculation methods based on the measurement of the red blood cell indices are very common and have been used as first-line diagnostic methods so far. The sensitivity and specificity of these methods have been assessed in various studies and their validity has been proven [134-139]. Since 1973, there has always been an attempt to introduce newer and more efficient discriminating indices between IDA and thalassemia minor. The calculations of these indices are sometimes so easy and sometimes challenging. The dissimilarity of age and sex groups and the difference in serum hemoglobin and ferritin thresholds in detecting IDA or thalassemia could lead to categorizing patients with various severities of iron deficiency as an IDA group [140].

IDA and thalassemia trait RBC indices resemblance in CBC test makes it hard to only lean on CBC test in discriminating them. On the other hand, performing hemoglobin electrophoresis and more importantly, the interpretation of its results are not applicable in every thalassemia endemic region. These are some of the reasons why the usage of mathematical indices is beneficial and useful in discriminating IDA from thalassemia trait even with their lack of full sensitivity and specificity [112]. Moreover, it is necessary to have in mind that previous studies evaluating these indices may have not completely considered some of the effective parameters which cause a major limitation in the study of mathematical indices. For instance, how to introduce cutoff values, study design, patient inclusion criteria and thalassemia genotype of patients are all among affective parameters which are required to be considered in order to reach a definite outcome in the analysis. Among introduced indices, it seems that later ones are more reliable in discriminating IDA from thalassemia trait because of their higher sensitivity and specificity.

Conclusion and clinical application

In this meta-analysis, we showed that the M/H ratio index is more sensitive and specific compared to our other studied indices. Therefore, it has more potential to discriminate IDA from thalassemia trait. However, we still cannot use this index alone to achieve a final diagnosis. The capability of this index to discriminate IDA from thalassemia trait must be used alongside with standard guidelines for identification of these entities to ensure the final differentiation. Ultimately, the detection of IDA from thalassemia trait using these indices is a valuable diagnostic approach. Although these tests have relatively high sensitivity and specificity, the result should be interpreted based on the combination of clinical and laboratory ratifications. Further studies regarding the measurement of the biomarkers for screening and fast detection are suggested.

Conflict of Interest

Authors declared no conflict of interest.

Acknowledgments

We are grateful to all colleagues who help us in the study.
8 Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia....

References
24. Nishad AAN, Pathmeswaran A, Wickremasinghe A, Premawardhena A. The Thal-index with the BTT prediction. exe to discriminate β-thalassaemia traits from other microcytic anaemias. 2012.
41. Ghionni H, Miotti T, Camandona V. Routine erythrocyte measurements and differentiation of
Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia....


63. Liu T-C, Seong P-S, Lin T-K. The erythrocyte cell hemoglobin distribution width segregates thalassemia traits from other nonthalassemic


12 Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia….


14 Diagnosis of Microcytic Anemia, Thalassemia or Iron Deficiency Anemia….


126. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. 2005.


How to Cite This Article: