Discriminant indices for distinguishing thalassemia and iron deficiency in patients with microcytic anemia: a meta-analysis

DOl 10.1515/cclm-2015-0179
Received February 19, 2015; accepted April 1, 2015; previously published online May 12, 2015

Abstract

Background: More than 40 mathematical indices have been proposed in the hematological literature for discriminating between iron deficiency anemia and thalassemia trait in subjects with microcytic red blood cells (RBCs). None of these discriminant indices is 100% sensitive and specific and also the ranking of the discriminant indices is not consistent. Therefore, we decided to conduct the first meta-analysis of the most frequently used discriminant indices.

Methods: An extensive literature search yielded 99 articles dealing with 12 indices that were investigated five or more times. For each discriminant index we calculated the diagnostic odds ratio (DOR) and summary ROC analysis was done for comparing the performance of the indices.

Results: The ratio of microcytic to hypochromic RBCs (M/H ratio) showed the best performance, DOR=100.8. This was significantly higher than that of all other indices investigated. The RBC index scored second (DOR=47.0), closely followed by the Sirdah index (DOR=46.7) and the Ehsani index (DOR=44.7). Subsequently, there was a group of four indices with intermediate and three with lower DOR. The lowest performance (DOR=6.8) was found for the RDW (Bessman index). Overall, the indices performed better for adults than for children.

Conclusions: The M/H ratio outperformed all other discriminant indices for discriminating between iron deficiency anemia and thalassemia trait. Although its sensitivity and specificity are not high enough for making a definitive diagnosis, it is certainly of value for identifying those subjects with microcytic RBC in whom diagnostic tests for confirming thalassemia are indicated.

Keywords: discriminant index; iron deficiency anemia; microcytic erythrocytes; thalassemia.

Introduction

Microcytic anemia is commonly the consequence of iron deficiency anemia (IDA), of thalassemia trait or a combination of these. IDA is a very frequent finding, not only in developing countries due to deficient nutritional status, but also in the western world, where women of childbearing age are often diagnosed with IDA due to intermittent blood loss in combination with insufficient iron intake [1]. Thalassemia traditionally has a high prevalence in the Mediterranean area, countries in the Middle East, the Arabic peninsula and Southeast Asia, but nowadays population migration has spread thalassemia genes over nearly the entire globe. Differentiating mild or moderate IDA from thalassemia trait can be a diagnostic dilemma, as both conditions share many characteristics. Obviously a correct diagnosis in patients with microcytic anemia is important: it can provide an indication for supplementing iron to IDA patients, for avoiding unnecessary iron therapy in thalassemia carriers and of course also for preventing severe and lethal forms of thalassemia syndromes in the framework of premarital counseling in high-prevalence areas.

Apart from the basic complete blood count, laboratory tests like ferritin, hemoglobin analysis (HbA2 and abnormal Hb) and DNA analysis are the key diagnostic parameters for IDA, β- and α-thalassemia, respectively.
However, areas where thalassemia is endemic often have low health care resources and these assays may not be generally available. Therefore, several simple screening indices have been developed for differentiating between thalassemia trait and IDA [2–8] and more recently these were supplemented with other supposedly better performing indices [9–14] (Table 1). It is widely agreed that none of these indices is 100% sensitive or 100% specific. Even more complex approaches including combinations of different simple indices, multivariate discriminant analysis or artificial neural network computing are unable to reach absolute sensitivity and specificity [15–23]. It is somewhat surprising that comparative studies of these screening indices do not show a consistent picture: discriminant indices that are superior in one study may perform less well in another study. The reasons for these discrepancies are not clear; possibly regional differences in thalassemia genotypes and analytical factors play a major role. Moreover, most published studies were comprised of small numbers of patients, up to a few hundred patients only, and this may also contribute to explaining the variable outcomes. In order to overcome these numerical limitations, we undertook a study using meta-analysis and composite ROC analysis for comparing the diagnostic performance of the various discriminant indices.

Materials and methods

Literature search

For finding relevant literature we used PubMed (http://www.ncbi.nlm.nih.gov/pubmed), Scopus (www.scopus.com) and ProQuest (www.proquest.com), which together cover practically all biomedical journals and many other publications in the field. First, we used the combination search “microcytic and iron deficiency and/or thalassemia” and filtered using the terms “distinguish or differentiate or discriminant”. Second, we identified the original publications of all discriminant indices and searched for publications citing them. Finally, we perused the literature reference lists in the publications found above for references not yet covered. We did not restrict ourselves regarding language of the reports and occasionally used the help of a translator.

Studies that proposed a new discriminant index without validation in an independent patient cohort were disregarded. From studies that reported separate learning and validation sets, we only included the validation set in our analysis. Some studies had to be omitted because the results reported were insufficient for deriving sensitivity and specificity data. In order to obtain sufficient power in the statistical analysis, we included only discriminant indices that had been investigated by five or more studies.

Statistical analysis

The primary outcome in the analyses was the performance of different markers in the differential diagnosis of microcytic anemia as quantified in terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values. These values were retrieved from the publications found and entered into a database. In studies with sensitivities and specificities of 100% [i.e., studies with at least a zero in any of the cells (TP, TN, FP, FN)], 0.5 subjects were added to the all four cells, for the purpose of the analysis [24].

Pooled positive likelihood ratio (PLR) and negative likelihood ratio (NLR) with their 95% confidence intervals

Table 1: Discriminant indices for distinguishing thalassemia trait from iron deficiency anemia in patients with microcytic RBC (cut-off values transformed into generally used units: Hb in g/dL; RBC in 10^12/L; MCV in fL; MCH in pg; RDW in %).

<table>
<thead>
<tr>
<th>Discriminant index</th>
<th>Calculation</th>
<th>Cut-off value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>England and Fraser (E&amp;F)</td>
<td>MCV – RBC – (5 Hb) – 3.4</td>
<td>0</td>
<td>[2]</td>
</tr>
<tr>
<td>RBC</td>
<td>RBC</td>
<td>5.0</td>
<td>[2]</td>
</tr>
<tr>
<td>Mentzer</td>
<td>MCV / RBC</td>
<td>13</td>
<td>[3]</td>
</tr>
<tr>
<td>Srivastava</td>
<td>MCH / RBC</td>
<td>3.8</td>
<td>[4]</td>
</tr>
<tr>
<td>Shine and Lal (S&amp;L)</td>
<td>MCV^2 / MCH</td>
<td>1.53</td>
<td>[5]</td>
</tr>
<tr>
<td>Bessman</td>
<td>RDW</td>
<td>15</td>
<td>[6]</td>
</tr>
<tr>
<td>Ricerca</td>
<td>RDW / RBC</td>
<td>4.4</td>
<td>[7]</td>
</tr>
<tr>
<td>Green and King (G&amp;K)</td>
<td>MCV^2 / RDW / 100 Hb</td>
<td>65</td>
<td>[8]</td>
</tr>
<tr>
<td>Jayabose (RDW index)</td>
<td>MCV / (RBC×RDW)</td>
<td>220</td>
<td>[9]</td>
</tr>
<tr>
<td>Sirdah</td>
<td>MCV – RBC – (3 Hb)</td>
<td>27.0</td>
<td>[11]</td>
</tr>
<tr>
<td>M/H ratio</td>
<td>Microcytic RBC %/hypochromic RBC %</td>
<td>3.7</td>
<td>[12]</td>
</tr>
<tr>
<td>Ehsani</td>
<td>MCV – (10 RBC)</td>
<td>15</td>
<td>[13]</td>
</tr>
</tbody>
</table>
(95% CI) were calculated using random effects models [25]. Summarized sensitivities and specificities were also computed to assess the clinical effectiveness, even though both screening parameters are not considered appropriate for meta-analyses [26].

As an accuracy measure we calculated the diagnostic odds ratio (DOR), an indicator of test accuracy that comprises a combination of sensitivity and specificity and is independent of disease prevalence, making it very appropriate for comparing different studies [27, 28]. The higher the DOR value, the better discriminatory test performance is present.

Summary receiver operating characteristic (SROC) curves were used to summarize overall test performance and to calculate the area under the SROC curve (AUC), which is quite robust to heterogeneity [29]. An AUC value <0.75 means that the test shows deficiencies in its diagnostic accuracy.

Regarding the assessment of heterogeneity, we used the inconsistency index ($I^2$) [30], which quantifies the proportion of the total variation across studies caused by heterogeneity rather than chance, indicating heterogeneity at an $I^2$ value >30%. We also conducted sensitivity analyses according to the region of origin, patient age (adults or children) and type of analyzer.

A bivariate generalized linear mixed-effects regression model [28] was used to test the robustness of these meta-analytical summaries and to compare these results among the different discriminant formulas evaluated. This bivariate approach accounts for potential between-study heterogeneity and incorporates the possible correlation between the sensitivity and the FP rate.

Publication bias was assessed visually by using a scatter plot of the inverse of the square root of the effective sample size ($1/\sqrt{\text{ESS}}$) versus the diagnostic log odds ratio, which would have a symmetric funnel shape when publication bias was absent. Formal testing for publication bias was conducted using a regression of the log DOR against $1/\sqrt{\text{ESS}}$ and weighting it according to the effective sample size, with $p<0.10$ indicating significant asymmetry [31].

All data were analyzed using the software packages Meta-DiSc (version 1.4) [32] and SAS System (9.4 release; SAS Institute, Cary, NC, USA).

### Results

We identified 147 reports in which at least one discriminant index had been evaluated. Twelve of these discriminant indices had been evaluated by five or more studies (Tables 1 and 2). In total, 99 such studies comprised 135,409 patient results, ranging from 3091 for the M/H ratio to 22,022 for the England and Fraser index (Table 2). Some 30 other discriminant formulas had been investigated in <5 studies and therefore were excluded from our current meta-analysis.

Out of these 99 studies, 36 (36%) were from Europe, 24 (26%) from the Mediterranean region, 20 (20%) from Southeast Asia, 14 (14%) were from North America and

<table>
<thead>
<tr>
<th>Discriminant index</th>
<th>Studies evaluated</th>
<th>Patients included</th>
<th>References to studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>England and Fraser (E&amp;F)</td>
<td>74</td>
<td>22,022</td>
<td>[11–14, 18, 20, 21, 33–100]</td>
</tr>
<tr>
<td>Bessman (RDW)</td>
<td>48</td>
<td>12,039</td>
<td>[7, 11, 14, 17, 20, 37, 39, 42–46, 48–50, 54–57, 59, 62, 68, 69, 71–78, 80, 81, 84, 87, 90, 94, 96, 97, 100, 107–115]</td>
</tr>
<tr>
<td>RBC</td>
<td>38</td>
<td>8704</td>
<td>[9, 13, 14, 20, 34, 36–39, 54, 56–58, 65, 67–69, 71–78, 80, 81, 83, 84, 87, 90, 94, 96, 97, 99, 100, 103, 107, 113]</td>
</tr>
<tr>
<td>Jayabose (RDWI)</td>
<td>28</td>
<td>9847</td>
<td>[11, 14, 20, 68, 71–73, 75–84, 87, 89, 90, 93–97, 99, 104, 107, 113]</td>
</tr>
<tr>
<td>Ricerca</td>
<td>25</td>
<td>9593</td>
<td>[11, 12, 14, 20, 57, 71, 72, 74, 75, 78–82, 84–87, 89, 90, 93, 94, 96, 98, 99]</td>
</tr>
<tr>
<td>M/H ratio</td>
<td>15</td>
<td>3091</td>
<td>[12, 18, 21, 57, 61, 62, 85, 98, 100, 103, 118–122]</td>
</tr>
<tr>
<td>Ehsani</td>
<td>14</td>
<td>6244</td>
<td>[11, 13, 14, 20, 80, 85, 86, 91, 92, 94, 95, 97, 99, 104]</td>
</tr>
<tr>
<td>Sirdah</td>
<td>12</td>
<td>5634</td>
<td>[11, 14, 20, 80, 85, 86, 92, 94, 95, 97, 99, 104]</td>
</tr>
</tbody>
</table>
the few remaining studies were from Latin America and Australia. Forty-one studies (41%) investigated adults, 11 (11%) included only children, 13 (13%) were focused on mixed populations of adults and children and 35 (35%) did not provide the patients’ age. With regard to the hematology analyzers used, 32 (32%) conducted the analyses with Coulter, 20 (20%) with Bayer and 18 (18%) with Sysmex; the remaining 30% were performed with other analyzers or the analyzer type was not specified.

For each discriminant index, the DOR was calculated using the data from all applicable studies (Table 3). It appeared that the M/H ratio displayed the highest DOR, namely 100.8 (95% confidence interval 39.6–256.3); this DOR was higher than the DOR of all other discriminant indices. The RBC index gave the second highest DOR (47.0; 95% CI 29.5–74.9), closely followed by the Sirdah index (46.7; 95% CI 23.4–92.9) and the Ehsani index (44.7; CI 26.8–74.7), as shown in Table 3. There were four indices with a relatively low performance (DOR <16): the Bessman (RDW), Shine and Lal, Srivastava and Ricerca indices (Table 3). There appeared to be qualitative evidence for DOR heterogeneity between studies (I 2 >70% in all indices; not shown). Considering the Bessman index (RDW) as a reference, the bivariate analysis showed results and comparisons to be very similar (see Table 4). In view of space constraints, the original data on TPs, FPs and FNs per study are not included; however, for interested readers they are available upon request.

The performance of each discriminant index is graphically illustrated in SROC plots, which also show the

<table>
<thead>
<tr>
<th>Discriminant index</th>
<th>DOR (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M/H ratio</td>
<td>100.8 (39.6–256.3)</td>
<td>6.8 (4.8–9.8)</td>
<td>0.07 (0.03–0.2)</td>
<td>0.92 (0.87–0.98)</td>
<td>0.86 (0.81–0.91)</td>
<td>0.956</td>
</tr>
<tr>
<td>RBC</td>
<td>47.0 (29.5–74.9)</td>
<td>8.1 (5.8–11.4)</td>
<td>0.17 (0.13–0.22)</td>
<td>0.85 (0.80–0.88)</td>
<td>0.90 (0.86–0.93)</td>
<td>0.923</td>
</tr>
<tr>
<td>Sirdah</td>
<td>46.7 (23.4–92.9)</td>
<td>8.6 (4.8–15.5)</td>
<td>0.18 (0.12–0.27)</td>
<td>0.83 (0.75–0.89)</td>
<td>0.90 (0.83–0.95)</td>
<td>0.903</td>
</tr>
<tr>
<td>Ehsani</td>
<td>44.7 (26.8–74.7)</td>
<td>5.1 (3.7–7.0)</td>
<td>0.11 (0.10–0.18)</td>
<td>0.91 (0.85–0.96)</td>
<td>0.82 (0.76–0.87)</td>
<td>0.925</td>
</tr>
<tr>
<td>England and Fraser (E&amp;F)</td>
<td>34.7 (25.0–48.2)</td>
<td>9.5 (7.2–12.6)</td>
<td>0.27 (0.23–0.32)</td>
<td>0.75 (0.70–0.79)</td>
<td>0.92 (0.90–0.94)</td>
<td>0.887</td>
</tr>
<tr>
<td>Green and King (G&amp;K)</td>
<td>29.8 (18.5–47.8)</td>
<td>7.2 (5.2–10.0)</td>
<td>0.24 (0.2–0.3)</td>
<td>0.79 (0.73–0.83)</td>
<td>0.89 (0.85–0.92)</td>
<td>0.898</td>
</tr>
<tr>
<td>Jayabose (RDWI)</td>
<td>28.6 (17.8–45.9)</td>
<td>5.6 (4.4–7.1)</td>
<td>0.20 (0.14–0.27)</td>
<td>0.83 (0.78–0.88)</td>
<td>0.85 (0.81–0.88)</td>
<td>0.902</td>
</tr>
<tr>
<td>Mentzer</td>
<td>27.6 (20.7–36.6)</td>
<td>5.6 (4.6–6.8)</td>
<td>0.20 (0.17–0.24)</td>
<td>0.82 (0.79–0.86)</td>
<td>0.85 (0.82–0.88)</td>
<td>0.896</td>
</tr>
<tr>
<td>Shine and Lal (S&amp;L)</td>
<td>15.7 (8.8–28.0)</td>
<td>1.6 (1.3–2.0)</td>
<td>0.10 (0.07–0.16)</td>
<td>0.96 (0.93–0.97)</td>
<td>0.41 (0.27–0.56)</td>
<td>0.885</td>
</tr>
<tr>
<td>Ricerca</td>
<td>15.6 (7.9–30.9)</td>
<td>2.0 (1.4–2.7)</td>
<td>0.12 (0.07–0.22)</td>
<td>0.93 (0.88–0.97)</td>
<td>0.52 (0.36–0.67)</td>
<td>0.850</td>
</tr>
<tr>
<td>Srivastava</td>
<td>15.0 (10.9–20.6)</td>
<td>4.1 (3.3–5.1)</td>
<td>0.28 (0.23–0.34)</td>
<td>0.78 (0.72–0.82)</td>
<td>0.81 (0.77–0.85)</td>
<td>0.850</td>
</tr>
<tr>
<td>Bessman (RDW)</td>
<td>6.8 (4.0–11.7)</td>
<td>5.1 (4.2–6.2)</td>
<td>0.21 (0.17–0.27)</td>
<td>0.62 (0.61–0.63)</td>
<td>0.66 (0.65–0.68)</td>
<td>0.778</td>
</tr>
</tbody>
</table>

AUC, area under the ROC curve; NLR, negative likelihood ratio; PLR, positive likelihood ratio. The higher DOR values, the better discriminatory test performance is present. Positive and negative likelihood ratios >10 and <0.1 indicate that the test generates strong evidence to rule in or rule out a thalassemia diagnosis, respectively.

For each discriminant index, the DOR was calculated using the data from all applicable studies (Table 3). It appeared that the M/H ratio displayed the highest DOR, namely 100.8 (95% confidence interval 39.6–256.3); this DOR was higher than the DOR of all other discriminant indices. The RBC index gave the second highest DOR (470; 95% CI 29.5–74.9), closely followed by the Sirdah index (46.7; 95% CI 23.4–92.9) and the Ehsani index (44.7; CI 26.8–74.7), as shown in Table 3. There were four indices with a relatively low performance (DOR<16): the Bessman (RDW), Shine and Lal, Srivastava and Ricerca indices (Table 3). There appeared to be qualitative evidence for DOR heterogeneity between studies (I 2>70% in all indices; not shown). Considering the Bessman index (RDW) as a reference, the bivariate analysis showed results and comparisons to be very similar (see Table 4). In view of space constraints, the original data on TPs, FPs and FNs per study are not included; however, for interested readers they are available upon request.

The performance of each discriminant index is graphically illustrated in SROC plots, which also show the

<table>
<thead>
<tr>
<th>Discriminant index</th>
<th>Sensitivity (95% CI)</th>
<th>p-Value</th>
<th>False positive rate (95% CI)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.6 (0.2–0.9)</td>
<td>&lt;0.001</td>
<td>−1.2 (−1.5 to −0.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M/H ratio</td>
<td>1.8 (1.1–2.5)</td>
<td>&lt;0.001</td>
<td>−0.6 (−1.3 to 0.1)</td>
<td>0.10</td>
</tr>
<tr>
<td>RBC</td>
<td>1.1 (0.6–1.6)</td>
<td>&lt;0.001</td>
<td>−0.9 (−1.4 to −0.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sirdah</td>
<td>1.0 (0.3–1.8)</td>
<td>&lt;0.001</td>
<td>−1.0 (−1.8 to −0.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ehsani</td>
<td>1.7 (1.0–2.4)</td>
<td>&lt;0.001</td>
<td>−0.3 (−1.1 to 0.4)</td>
<td>0.30</td>
</tr>
<tr>
<td>England and Fraser (E&amp;F)</td>
<td>0.5 (0.1–0.9)</td>
<td>&lt;0.001</td>
<td>−1.1 (−1.6 to −0.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Green and King (G&amp;K)</td>
<td>0.7 (0.2–1.2)</td>
<td>&lt;0.001</td>
<td>−0.8 (−1.3 to −0.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Jayabose (RDWI)</td>
<td>1.0 (0.5–1.5)</td>
<td>&lt;0.001</td>
<td>−0.5 (−0.9 to −0.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mentzer</td>
<td>1.0 (0.5–1.4)</td>
<td>&lt;0.001</td>
<td>−0.5 (−1.1 to 0.0)</td>
<td>0.10</td>
</tr>
<tr>
<td>Shine and Lal (S&amp;L)</td>
<td>2.2 (1.7–2.7)</td>
<td>&lt;0.001</td>
<td>1.5 (1.0–2.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ricerca</td>
<td>1.8 (1.2–2.4)</td>
<td>&lt;0.001</td>
<td>1.0 (0.5–1.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Srivastava</td>
<td>0.6 (0.2–1.1)</td>
<td>&lt;0.001</td>
<td>−0.2 (−0.7 to 0.2)</td>
<td>0.30</td>
</tr>
<tr>
<td>Bessman (RDW)</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>

Reference: the Bessman (RDW) was considered as the reference group.
dispersion of the data from the individual studies reporting on that specific index (Figure 1 and Table 3). The M/H ratio showed the highest AUC value (0.956), again indicating the best diagnostic performance.

In a sub-analysis we sought to identify covariates that might be of influence on the diagnostic performance of the discriminant indices. Overall, the indices appeared to perform much better in adults than in children: DOR 33.6 (95% CI 27.7–40.7) for adults as compared with DOR 11.7 (95% CI 7.8–17.7) for children. Studies that had enrolled both children and adults, but did not report separate results, had intermediate diagnostic performance. Also the geographical region showed important differences: the highest overall DOR (53.1; 95% CI 41.1–68.6) was found in European populations, whereas the lowest DOR was obtained in studies from Southeast Asia (9.3; 95% CI 7.5–11.7), as illustrated in Table 5. The make of the hematology analyzer used had only minor influence: the older Coulter analyzers showed somewhat lower diagnostic performance, whereas the current Beckman-Coulter generation and the analyzers of all other manufacturers scored rather similar (not shown).

Finally, it appeared that potential publication bias was present (p<0.001), as judged by the degree of asymmetry in the funnel plots (not shown).

Discussion

Differentiating IDA from thalassemia carrier status is a frequent issue in medical practice, in particular in subjects with mild or moderate IDA and in regions where thalassemia is common. It is not possible to distinguish both conditions using simple routine blood counts, as they are both associated with microcytic and hypochromic erythrocytes. However, in thalassemia RBC do tend to be more microcytic, whereas iron deficient RBC are often more hypochromic [59, 118]. These differences have been exploited by developing simple mathematical formulas for emphasizing the differences in RBC indices as a tool for distinguishing IDA from thalassemia trait [2–8]. However, the discriminative power of these simple indices never reached maximum diagnostic performance. The large number of discriminant indices described in the literature reflects that researchers were continuously stimulated devising new and supposedly better indices for applying in their local patient population. In the last decade, multiple studies have been published which compared different discriminant indices in the same patient cohort, aimed at identifying the index with the best overall performance. Yet, no single index emerged as the best and it became evident that even the performance ranking of the indices was different across the various investigations. Therefore we carried out a meta-analysis in order to find the discriminant indices with the highest overall performance. To our best knowledge this is the first time that this subject was investigated using a meta-analysis.

Any meta-analysis has inherent limitations and also in our study we faced various methodological issues. Most importantly, the designs of the studies investigated were far from homogeneous: there was huge variation in patient selection criteria, in types of thalassemia included, in geographical origin of the patients, in type of hematology analyzer used and in cut-off value for the respective discriminant indices. As each of these factors may play a role in the diagnostic utility of the indices, we will discuss them in more detail below.

Patient selection

Most discriminant indices were designed for distinguishing IDA and thalassemia in subjects with microcytic RBC. These two conditions explain the vast majority of microcytic RBC, but other diseases may be associated with microcytosis, too. For example, patients with anemia of chronic disease (ACD), although most often normocytic, may occasionally have microcytic anemia and many studies did not report whether ACD was an exclusion criterion for patient selection. However, some studies classified patients with ACD separately and the results of these studies indicate that ACD patients are more similar to IDA than to thalassemia carriers [56, 110]. Therefore, misclassification of ACD as thalassemia can be considered unlikely.

Patient age

Overall, the indices performed better in adults than in children. However, when assessed in more detail it appeared that the older indices (England and Fraser, Mentzer, Green and King) evidently performed better in adults. In contrast, some newer discriminant indices (Jayabose, Sirdah and Ehsani) had a much better performance in children. The Srivastava and RBC indices appeared to be equally powerful in adults as in children. The M/H ratio has until now only been investigated in adult populations, so it remains to be seen how this discriminant index performs in children.
Figure 1: Summary receiver operator characteristics (SROC) plots of the 12 discriminant indices. In each plot, the dots represent the individual studies, with their sensitivity and specificity coordinates. The small circle is the summarized sensitivity and specificity, with the 95% confidence interval around it.
Thalassemia types

Virtually all studies included carriers of β-thalassemia and some studies recruited both α- and β-thalassemia carriers. Few studies comprised only carriers of α-thalassemia [78, 87, 123] or α- and δβ-thalassemia [124]. The overall picture that emerges from these studies is that all discriminant indices perform better in β- than in α-thalassemia, even if only microcytic α-thalassemia carriers were included [35, 38, 61, 71, 98].

Some investigators included also subjects with other types of hemoglobinopathy, like HbE [116, 117, 125], HbO-Arab [126] and HbS, both sickle cell thalassemia and sickle cell disease [48, 94]. Unfortunately the numbers reported are too small for making a solid conclusion as to the utility of the discriminant indices in these conditions.

Thalassemia with concomitant iron deficiency

The presence of IDA in a thalassemia carrier is by no means a rare finding: many studies included such patients and they almost unanimously demonstrate that discriminant indices identify these patients as most likely having IDA [118]. Although diagnostically incorrect, from a clinical perspective this is not problematic, as such patients need iron supplementation anyway. An indication for underlying thalassemia trait can only be obtained once the IDA component of the microcytic anemia is successfully resolved [127].

Geographical origin

It has been reported that the RBC indices MCV, MCH and MCHC show remarkably small differences over the globe [128], enabling using them for internal quality control purposes. As many of the discriminant indices are based on these and other basic RBC parameters, one might expect that the indices would perform similarly in different areas of the world. However, our analysis surprisingly yielded indications for considerable differences between different geographical regions (Table 5). Overall, the indices performed best in European countries, but with notable differences: e.g., the Mentzer and Shine and Lal indices scored poorer than in other regions, while Green and King, Ricerca, Jayabose, Sirdah and Ehsani indices were superior in European studies. With the limitations of a meta-analysis explained above, one could roughly state that in a Mediterranean population the Mentzer, Shine and Lal and Sirdah indices would be preferred; in South-east Asia the Srivastava index, whereas in Chinese populations the Ricerca index and Bessman index (RDW) can be expected to perform better. Anyway, our investigation has made clear that the performance of any index seems to depend on the regional population in which it is applied.

Table 5: Overall diagnostic performance of the discriminant indices by geographical region, arranged in order of diagnostic odds ratio (DOR) with 95% confidence intervals (95% CI).

<table>
<thead>
<tr>
<th>Region</th>
<th>DOR (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>53.1 (41.1–68.6)</td>
<td>6.6 (5.4–8.2)</td>
<td>0.15 (0.12–0.19)</td>
</tr>
<tr>
<td>Australia</td>
<td>26.2 (13.2–52.2)</td>
<td>4.7 (3.3–6.7)</td>
<td>0.22 (0.12–0.34)</td>
</tr>
<tr>
<td>China, Hong-Kong, Malaysia, Singapore</td>
<td>19.6 (12.1–31.7)</td>
<td>4.2 (2.0–8.5)</td>
<td>0.23 (0.12–0.45)</td>
</tr>
<tr>
<td>Central and South America</td>
<td>19.6 (9.9–38.5)</td>
<td>5.8 (2.5–13.4)</td>
<td>0.32 (0.24–0.42)</td>
</tr>
<tr>
<td>Eastern Mediterranean and Middle East</td>
<td>17.7 (14.3–21.8)</td>
<td>3.9 (3.2–4.7)</td>
<td>0.26 (0.22–0.30)</td>
</tr>
<tr>
<td>North America</td>
<td>12.9 (8.8–18.9)</td>
<td>3.9 (3.0–5.1)</td>
<td>0.35 (0.28–0.44)</td>
</tr>
<tr>
<td>Southeast Asia (India, Bangladesh, Sri Lanka, Thailand)</td>
<td>9.3 (7.5–11.7)</td>
<td>2.7 (2.3–3.2)</td>
<td>0.32 (0.28–0.38)</td>
</tr>
</tbody>
</table>

NLR, negative likelihood ratio; PLR, positive likelihood ratio.

Type of hematology analyzer

As the reports investigated in our meta-analysis span a period of four decades, it was inevitable that a wide variety of hematology analyzers was used for performing the studies investigated here. Despite this, the influence of analyzer type on the diagnostic outcome was relatively small. Studies using older Coulter analyzers, which were abundant in clinical laboratories in the 1970s and 1980s, demonstrated somewhat lower performance than the analyzers in use during the last 20 years. The question arises whether this is due to the analyzer itself or to the lower degree of inter-laboratory standardization and quality control in those earlier years. Studies reported in the 21st century, where a high level of analytical standardization of basic RBC parameters is achieved, show a high degree of homogeneity and therefore we believe that the influence
of analyzer type on the performance of the discriminant indices is limited, if not negligible. The only situation where one might expect more heterogeneity is for those discriminant indices that incorporate RDW, because this parameter is not well standardized and shows considerable differences between different analyzers [129, 130]. This factor may explain the moderate to low diagnostic performance of RDW-containing indices, as shown in Table 3.

Cut-off values

One of the complications of our meta-analysis is that many authors have not used the original cut-off values of the discriminant indices, but applied an alternative cut-off. For example, Mentzer originally published his index with 13 as the cut-off value [3]. Other authors, however, used values between 13 and 14 [20, 39, 48, 62, 70], between 14 and 15 [7, 34, 38, 43, 55, 58, 59, 65, 82, 87, 93, 94, 96, 101], 15.5 [57], 17 [44, 61, 92] or even as high as 20 [102], without proper validation. Therefore the effect of a modified cut-off value on an index’s performance is difficult to judge and may require further investigations. However, our present analysis has shown that cut-off values are not the most important contributors to the performance of a discriminant index.

Conclusions

This meta-analysis has demonstrated high variation in the performance of discriminant indices for distinguishing thalassemia trait from IDA. In general, the newer indices seem to be able to make this distinction better than the more traditional formulas. We have also shown that age (adult or child) and geographical region, but not the type of hematology analyzer, are important factors determining the diagnostic utility of the discriminant indices. We have objectively shown the superiority of the M/H ratio over other discriminant indices. We have also shown that the newer discrimination indices are important factors that age (adult or child) and geographical region, but not the type of hematology analyzer, are important factors determining the diagnostic utility of the discriminant indices. We have objectively shown the superiority of the M/H ratio over other discriminant indices. Notwithstanding its high performance, even the M/H ratio cannot be used for making a final diagnosis of thalassemia trait. Its value lies in screening of microcytic individuals in order to select those in whom additional laboratory investigations are warranted for confirming the presence of thalassemia.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Financial support: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References


91. Nishad AA, Pathmeswaran A, Wakimramasinghe AR, Premawardhena A. The Thal-index with the BTT prediction.exe to discriminate ß-thalassaemia traits from other microcytic anaemias. Thalass Rep 2012;2:e1–2.


112. Lima CS, Reis AR, Grotto HZ, Saad ST, Costa FF. Comparison of red cell distribution width and a red cell discriminant function incorporating volume dispersion for distinguishing iron deficiency from beta thalassaemia trait in patients with microcyto-


Bionotes

**Johannes J.M.L. Hoffmann**

Johannes J.M.L. Hoffmann has worked in clinical chemistry since 1976 when he became a trainee. Once certified as a Specialist in Laboratory Medicine he started working as the head of the hematology laboratory in a large tertiary-care teaching hospital in the Netherlands, where he was later also appointed director of the Department of Clinical Laboratories. In 1992 he obtained his PhD in Medical Sciences at Leiden University, the Netherlands, on a thesis in the field of fibrinolysis. Since 2008 he has been responsible for scientific affairs in hematology with Abbott Diagnostics in Europe. His scientific work comprises over 100 papers in peer-reviewed journals, mainly focused on general hematology, flow cytometry, coagulation and fibrinolysis. He is also author and co-author of several books on laboratory medicine and hematology. He is a member of the Editorial Board of *Clinical Chemistry and Laboratory Medicine* and acts as a reviewer for various other journals. In addition, he is a member of international committees and working groups on standardization in laboratory hematology.

**Eloísa Urrechaga**

Eloísa Urrechaga started her career in Clinical Chemistry in 1984 as a trainee in Hospital General de Asturias (Oviedo, Spain); she became head of laboratory in a community hospital, Hospital de Jarrio (Asturias, Spain) in 1990. From 2001 until present she has worked as a consultant for clinical chemistry in the Hospital Galdakao-Uxansolo (Bizkaia, Spain), responsible for the Hematology Laboratory. In 2010 she obtained her PhD in Medical Sciences at Zaragoza University (Zaragoza, Spain) on a thesis in the field of erythropoiesis. She is an expert of the Spanish Science and Innovation Ministry and the Agency for Assessment of new technologies for research project validation. Member of the WHO Guideline Development Group – Ferritin International Micronutrient Malnutrition Prevention and Control, World Health Organization and a member of the Analytical Hematology and Trace Elements Commissions of the Spanish Society of Clinical Chemistry and Molecular Pathology. She is particularly involved in the development of new hematological parameters applied to the study of anemia and erythropoiesis, as well as glycohemoglobin. Recently, she published papers in the area of automation, biological parameters in functional iron deficiency, β-thalassemia, anemia of chronic diseases and other erythropoietic disorders. She serves as an editorial board member and scientific reviewer in different journals, including *Clinical Chemistry and Laboratory Medicine*.

**Urko Aguirre**

Urko Aguirre is a mathematician and holds a Master in Computer Sciences at the University of the Basque Country (UPV/EHU). He is currently working at the Clinical Research Unit of the Hospital Galdakao-Uxansolo (Bizkaia, Spain). His main research field is focused on predictive modeling.