Multicenter Evaluation of Five Assays for Myoglobin Determination

Martina Zaninotto,1 Franca Pagani,2 Sara Altinier,1 Paolo Amboni,3 Roberto Bonora,2 Alberto Dolci,4 Patrizia Pergolini,5 Atrialdo Vernocchi,3 Mario Plebani,1 and Mauro Panteghini2* for the Italian Society of Clinical Biochemistry and Clinical Molecular Biology–Italian Society of Laboratory Medicine Working Group on Markers of Myocardial Damage

Background: Lacking assay standardization, different myoglobin methods may produce results that differ significantly.

Methods: A multicenter study was carried out to compare the analytical performance of five commercially available assays for myoglobin measurement. Linearity, imprecision, interferences, and method comparison were studied according to NCCLS guidelines, whereas reference values were determined following IFCC recommendations.

Results: The BNA and Opus showed relatively high imprecision (all but one total CV > 7.4%). Other assays showed lower CVs, but they varied among laboratories, particularly at a normal myoglobin concentration (Access, 6.0–11%; Hitachi, 3.8–5.8%; Stratus, 3.4–6.5%). Results were lower in anticoagulated samples on the Access, in heparin and citrate samples on the Stratus, and in citrate samples on the BNA and Opus, and increased in heparin and EDTA samples on the Hitachi. Use of separator gel produced results significantly lower (P < 0.001) on the Hitachi and higher (P = 0.016) on the Opus. Bilirubin, turbidity, and hemoglobin had no effect on evaluated methods, but rheumatoid factor affected the Access. In method comparisons, high correlation coefficients (≥0.98) were obtained. The Stratus gave higher results; however, the Access and BNA gave the lowest. The following upper reference limits (μg/L) for men and women, respectively, were obtained: Access, 70 and 52; BNA, 51 and 49; Hitachi, 67 and 58; Opus, 80 and 50; and Stratus, 86 and 63.

Conclusion: The possibility of high imprecision and marked disagreement among commercial myoglobin assays should be carefully considered in clinical practice.

© 2000 American Association for Clinical Chemistry

Myoglobin, a low-molecular weight protein present in the cytosol of cardiac and skeletal muscle, is an old biochemical marker with new perspectives (1). The first report on the clinical usefulness of serum myoglobin in the diagnosis of acute myocardial infarction, using a complement fixation technique for the protein measurement, was published in 1975 (2). Since then, much technical progress have been made, and methods requiring several hours have been replaced by assays with rapid turnaround times and better analytical performance (1). The availability of simple and rapid methods for the measurement of myoglobin has greatly enhanced its clinical utilization in emergency situations for the early diagnosis and management of myocardial infarction. As underlined recently (3), myoglobin currently is the marker that most effectively fits the role of an early marker of myocardial necrosis, but because assay standardization and reference materials are lacking, the diffusion of different analytical procedures and the rush by commercial manufacturers to develop and market new methods for myoglobin determination may produce data on patient samples that differ significantly when measured by different assays. In this report, we present the results of a multicenter study carried out to

1 Dipartimento di Medicina di Laboratorio, Università-Ospedale, 35100 Padova, Italy.
2 Laboratorio Analisi Chimico Cliniche 1, Azienda Ospedaliera ‘Spedali Civili’, 25125 Brescia, Italy.
3 Laboratorio Analisi Chimico Cliniche, Ospedali Riuniti, 24100 Bergamo, Italy.
4 Laboratorio Analisi Chimico Cliniche, Casa di Cura S. Maria, 21053 Castellanza VA, Italy.
5 Laboratorio Analisi Chimico Cliniche, Azienda Ospedaliera ‘Maggiore della Carità’, 28100 Novara, Italy.


*Author for correspondence. Fax 39-030-3995369; e-mail panteghi@osp.unibs.it.

Received March 7, 2000; accepted July 5, 2000.
evaluate and compare the analytical performance of five commercially available assays to measure myoglobin in serum.

**Materials and Methods**

The five participating laboratories were located in a narrow geographic area in northern Italy because it was necessary to distribute and transport in real time the frozen samples from one center to another to carry out every phase of the evaluation using the same samples in different centers. Five assays/platforms from three different manufacturers were evaluated, each one in two different laboratories: (a) Access (Beckman Coulter, Brea, CA; laboratories 2 and 3); (b) BNA II (Dade Behring, Deerfield, IL; laboratories 1 and 5); (c) Hitachi 911 (Roche Diagnostics, Mannheim, Germany; laboratories 1 and 4); (d) Opus (Dade Behring; laboratories 2 and 5); and (e) Stratus II (Dade Behring; laboratories 3 and 4). The study was carried out using the manufacturers’ recommendations in the preparation and storage of reagents, calibrators, and controls and in the frequency of system calibration. Unless otherwise stated, serum was used as sample. After centrifugation and separation, serum specimens were stored at −20 °C until being assayed.

**Detection Limit**

The zero calibrator of each assay was tested 10 times on the corresponding instrument in the same analytical run. The limit of detection was defined as the concentration of myoglobin corresponding to a signal 3 SD above the obtained mean value. This experiment was not carried out on the BNA because the lowest calibration curve concentration (~26 μg/L) in this instrument is obtained automatically by serial dilutions of the calibration material and the analyzer is set to report results lower than this point as <26 μg/L.

**Linearity**

According to the NCCLS EP6-P protocol (4), human serum pools with myoglobin concentrations ~30% higher than the upper limit of the calibration curve of each assay were used to prepare four dilutions (3:4, 1:2, 1:4, 1:5), using the manufacturer’s recommended diluent. Each dilution was tested four times, and all results were obtained in the same analytical run. The criteria of NCCLS document EP6-P were used to evaluate assay linearity (4).

**Precision**

Three human serum pools with normal (50 μg/L), borderline (100 μg/L), and high (300 μg/L) myoglobin concentrations were prepared, aliquoted, frozen, and sent to all participating laboratories. Assay imprecision was estimated using the analysis of variance method described in NCCLS EP5-A (5), with two replicates per specimen per analytical run and one analytical run per day for 20 days. During the experimental period, a manufacturer’s control sample with a myoglobin concentration of ~100 μg/L was assayed to validate each run, and a single lot of reagents was used.

**Interference Studies**

*Sample type.* Fifty-one volunteers were selected, and informed consent was obtained. A blood sample was drawn from each volunteer, using vacuum collection tubes (Becton Dickinson) to obtain serum in tubes with separator gel, lithium heparinate plasma, sodium citrate plasma, and EDTA-K3 plasma. The results for these samples were compared with the paired serum values, and the significance of the differences was evaluated (Wilcoxon rank-sum test).

*Endogenous interferents.* According to the NCCLS EP7-P guideline (6), albumin (final concentration, 100 g/L), bilirubin (final concentration, 300 mg/L), hemoglobin (final concentration, 10 g/L), and Intralipid 20% solution were added (one part of interferent and 19 parts of pool) to two human serum pools with normal (60 μg/L) and high (110 μg/L) myoglobin concentrations. The results for these “test” solutions were compared with those of “control” solutions obtained from the same two pools diluted in the same way with the manufacturers’ recommended diluents instead of the interfering substances. Each sample (test and control) was assayed 10 times in random order to minimize any bias attributable to analytical drift. Taking into account the analytical imprecision of various methods, we considered as clinically significant a difference >10% between test and control mean results (6). To test interference by rheumatoid factor (RF), samples (n = 10) with high RF concentrations [48–3020 kIU/L, when determined with a rate-nephelometric latex method (BNA); previously established upper reference limit, 20 kIU/L] were evaluated in duplicate, and the mean results from different instruments were compared.

**Method Comparison**

Forty samples (20 with myoglobin concentrations of 30–50 μg/L, 16 with 50–200 μg/L, and 4 with 200–400 μg/L) were tested on each system in duplicate, 8 per day on 5 different days, using the manufacturers’ control samples to validate the analytical runs (7). To minimize the effects of carryover and analytical drift on the averages of the duplicates within the run, the two aliquots of each sample were analyzed in reverse order in the run (7). Because none of the evaluated assays was considered the “reference” procedure, we used the Deming regression to calculate analytical correlation. Difference plots were also obtained (8).

**Reference Values**

Two hundred eight apparently healthy subjects (103 women and 105 men; ages 13–87 years) were studied. In addition to the general a priori exclusion criteria suggested by IFCC documents on the theory of reference values, creatine kinase activity in serum, a widely ac-
cepted biochemical marker of muscle tissue, was used as an additional criterion of selection to exclude individuals who exercised or followed a physical training program (9). In particular, 32 subjects (from the 240 initially selected) with creatine kinase activities >150 U/L (measured at 37 °C according to the IFCC recommended method) were excluded to minimize the possible influence of excessive physical activity on myoglobin concentrations. Samples on each platform were subdivided into four different analytical runs, carried out in different days using the same calibration curve. The data obtained were validated by assaying the manufacturers’ control sera at the beginning, in the middle, and at the end of each run. Reference limits were calculated using nonparametric determination of percentiles (10).

**Results**

**DETERMINATION OF THE DETECTION LIMIT**

All methods/instruments but one (Hitachi in laboratory 4) gave experimental detection limits lower than values declared by manufacturers. In particular, the following limits of detection were obtained: Access, 5.0 and 3.3 g/L (declared, 3.0 g/L); Hitachi, 1.6 and 4.6 g/L (declared, 3.0 g/L); Opus, 0.2 and 0.4 g/L (declared, 1.0 µg/L); and Stratus, 1.2 and 0.4 µg/L (declared, 1.2 µg/L).

**LINEARITY**

The results of the linearity studies are shown in Table 1. Because of the wide calibration range, experiments on the

**Table 1. Results of linearity studies.**

<table>
<thead>
<tr>
<th>Assay/Lab</th>
<th>Calibration range, µg/L</th>
<th>Evaluated range, µg/L</th>
<th>Equality of variance across dilutions (C value)a</th>
<th>Linearity (G value)b</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>10–4300</td>
<td>660–2700</td>
<td>0.853</td>
<td>1.12</td>
<td>0.9964</td>
</tr>
<tr>
<td>2nd</td>
<td>10–4300</td>
<td>120–390</td>
<td>0.045</td>
<td>0.44</td>
<td>0.9986</td>
</tr>
<tr>
<td>Lab 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>10–4300</td>
<td>720–2700</td>
<td>0.934</td>
<td>0.11</td>
<td>0.9976</td>
</tr>
<tr>
<td>2nd</td>
<td>10–4300</td>
<td>100–370</td>
<td>0.436</td>
<td>2.65</td>
<td>0.9994</td>
</tr>
<tr>
<td>BNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab 1</td>
<td>25–400</td>
<td>75–300</td>
<td>0.855</td>
<td>1.26</td>
<td>0.9988</td>
</tr>
<tr>
<td>Lab 5</td>
<td>25–400</td>
<td>75–300</td>
<td>0.800</td>
<td>0.11</td>
<td>0.9979</td>
</tr>
<tr>
<td>Hitachi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab 1</td>
<td>3–600</td>
<td>160–570</td>
<td>0.753</td>
<td>8.36</td>
<td>0.9997</td>
</tr>
<tr>
<td>Lab 4</td>
<td>3–600</td>
<td>160–570</td>
<td>0.522</td>
<td>23.31</td>
<td>0.9998</td>
</tr>
<tr>
<td>Opus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab 2</td>
<td>1–500</td>
<td>125–450</td>
<td>0.655</td>
<td>3.85</td>
<td>0.9899</td>
</tr>
<tr>
<td>Lab 5</td>
<td>1–500</td>
<td>125–450</td>
<td>0.787</td>
<td>0.37</td>
<td>0.9949</td>
</tr>
<tr>
<td>Stratus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab 3</td>
<td>1–1000</td>
<td>180–570</td>
<td>0.273</td>
<td>3.46</td>
<td>0.9966</td>
</tr>
<tr>
<td>Lab 4</td>
<td>1–1000</td>
<td>250–870</td>
<td>0.629</td>
<td>0.11</td>
<td>0.9977</td>
</tr>
</tbody>
</table>

*a If C value is >0.684, the hypothesis of equal variances across dilutions is rejected with P <0.05.

*b If G value is >3.89, the hypothesis of a linear fit is rejected with P <0.05.

Access were repeated in a lower concentration range. In some cases, the hypothesis of equality of variances across dilutions was rejected, indicating relatively high imprecision in the results of each dilution tested four times (Access in the high concentration range; BNA and Opus in laboratory 5) or very small variances in the assays, such as the Hitachi in laboratory 1, which showed good precision. When we tested the results for lack of fit of the linear model, all assays except for the Hitachi showed acceptable linearity. In the case of the Hitachi, a very precise assay, data declared statistically nonlinear (G values, 8.36 and 23.31) were, however, clinically acceptable at visual inspection, showing the highest correlation coefficients as well (Table 1).

**PRECISION STUDIES**

The imprecision results for the various assays are shown in Table 2. The BNA and Opus showed relatively high imprecision. The other assays had lower CVs, but they varied among laboratories performing the same assay. Analytical quality specifications analysis was performed to test whether the precision of each method was satisfactory (11). According to data on the biological variation of serum myoglobin (12), the analytical goal (expressed as CV) for optimum performance is <2.8%, the goal for desirable performance is <5.6%, and the goal for minimum performance is <8.3%. On the basis of intralaboratory variations, none of the assays evaluated showed optimum performance for all three types of samples: two instruments in one laboratory (Hitachi and Stratus in laboratory 4) showed desirable performance, and three instruments in one (Access in laboratory 3 and Opus in laboratory 2) or two laboratories (BNA) exceeded the requirements for minimum performance in one or more samples. In all, two methods performed best for analytical precision: the Hitachi and the Stratus.

**INTERFERENCE STUDIES**

With the Access, results were significantly lower in all anticoagulated samples (n = 48 for each type of anticoagulant). The mean difference (95% confidence interval) for heparin samples was −9.8% (−16.0% to −3.6%; P = 0.003); the mean difference for EDTA was −13.8% (−20.1% to −7.5%; P <0.001); and the mean difference for citrate was −30.2% (−36.7% to −23.6%; P <0.001). With the Stratus, the mean difference for heparin was −6.0% (−13.0% to −1.0%; P = 0.002); and the mean difference for citrate was −18.6% (−22.0% to −15.2%; P <0.001). For citrate samples, the BNA had a mean difference of −13.1% (−25.8% to −0.4%; P = 0.046; n = 6), and the Opus had a mean difference of −12.1% (−21.7% to −2.5%; P <0.001; n = 51). However, on the Hitachi (n = 34), heparin [10.6% (5–15.5%); P <0.001] and EDTA [7.8% (4–11.4%); P <0.001] results were increased. The use of separator gel produced results significantly lower on the Hitachi [−25.8% (−31% to −20.0%); P <0.001] and higher on the Opus [6.1% (1–10.7%); P = 0.016].
All tested endogenous interferents had statistically ($P < 0.05$) significant positive or negative effects on the BNA and Hitachi assays, whereas none interfered in the Access. Considering the a priori established limit of acceptability ($6 \pm 10\%$ in comparison with control values), only bilirubin ($300 \text{ mg/L}$) on the Opus and Stratus, the two immunoassays with fluorometric detection, caused markedly negative interference (mean difference between myoglobin results in test and control pools, $-27.5\%$ and $-27.6\%$ on the Opus and $-24\%$ and $-21\%$ on the Stratus for normal and high myoglobin pools, respectively). Because purified bilirubin (mixed isomers) from bovine gallstones (Sigma), which has a high intrinsic fluorescence, was used as interferent in this experiment, we decided to reevaluate the bilirubin interference with a different experimental method, by preparing a dilution series using human serum samples with high (510 mg/L) and low (3 mg/L) bilirubin concentration. At a bilirubin concentration of 306 mg/L, mean myoglobin recoveries were 96.1\% on the Opus and 100.9\% on the Stratus, showing no influence by human serum bilirubin on these assays.

Among the evaluated assays, only the Access produced results higher (1.3- to 4.3-fold) than the manufacturer’s recommended cutoff (70 $\mu$g/L) in samples with high concentrations of RF (>$600$ kIU/L), whereas the other four myoglobin assays gave results for all samples below the cutoffs recommended by each of the manufacturers. However, there appears to be little relationship between RF concentrations and the degree of assay interference.

**METHOD COMPARISON**

The Deming correlations for different assays are summarized in Table 3. Although good correlation coefficients were observed, there were significant proportional and constant biases in these data sets. The Stratus assay typically produced the highest results, whereas the Access and the BNA gave the lowest. All correlations involving the Hitachi showed a significant positive constant bias (intercepts, 9.9 –17.3 $\mu$g/L). Fig. 1 shows the differences between methods, expressed as a percentage of the average, plotted against the method average. It is apparent from Fig. 1 that there was a systematic bias in the correlations involving the Stratus and the Hitachi and that for the latter there was also a clear relationship between the method difference and the myoglobin concentration.

**REFERENCE VALUES**

Because the preliminary evaluation of findings showed a significant sex-related difference in myoglobin concentrations in the reference population, with $P$ values ranging from

| Table 3. Deming regression equations for myoglobin assay comparisons.$^{a}$ |
|-----------------------------|-----------------|----------------|-----------------|-----------------|
| $y$ | $x$ | $n$ | Slope (SD) | Intercept (SD), $\mu$g/L | $S_{yx}$ |
| Access | BNA | 29$^{a}$ | 0.87 (0.10) | 5.7 (8.4) | 5.5 |
| Access | Opus | 40 | 0.80 (0.13) | 5.5 (8.2) | 5.4 |
| Access | Stratus | 40 | 0.75 (0.12) | 0.3 (8.1) | 5.0 |
| BNA | Opus | 29$^{a}$ | 0.86 (0.16) | 4.5 (10.9) | 6.5 |
| BNA | Stratus | 29$^{a}$ | 0.82 (0.10) | $-2.1$ (10.0) | 5.8 |
| Hitachi | Access | 40 | 1.13 (0.12) | 9.9 (8.3) | 5.6 |
| Hitachi | BNA | 29$^{a}$ | 0.98 (0.05) | 13.8 (6.5) | 4.2 |
| Hitachi | Opus | 40 | 0.86 (0.12) | 17.3 (8.8) | 5.7 |
| Hitachi | Stratus | 40 | 0.82 (0.09) | 11.1 (6.8) | 4.2 |
| Opus | Stratus | 40 | 0.88 (0.12) | $-4.1$ (11.0) | 6.3 |

$a$ Correlation coefficient is 0.99 for all comparisons except Opus/Stratus (0.98).

$b$ Myoglobin concentrations in 11 patients’ samples were below the limit of detection for BNA assay.
Fig. 1. Plots of the differences among the evaluated methods for myoglobin determination.

Percentage of difference (y axis) = \((\text{method } A - \text{method } B) / \text{average of the two methods}\) \times 100. If the 95% confidence interval (CI) of the mean difference includes 0, there is no evidence of systematic bias between the compared assays.
from 0.032 for the Hitachi to <0.0001 for the Access, Opus, and Stratus, the calculation for the reference limits for these assays was performed separately for men and women. In particular, the reference limits [95% central range of the distribution (and the 99th percentile value) in men and women, respectively] were estimated to be 5.1–69.5 μg/L (78.5 μg/L) and 3.3–51.5 μg/L (81.2 μg/L) for the Access; 9.6–67.0 μg/L (81.4 μg/L) and 8.5–57.7 μg/L (88.9 μg/L) for the Hitachi; 13.8–79.8 μg/L (93.4 μg/L) and 11.0–49.9 μg/L (83.1 μg/L) for the Opus; and 11.1–85.5 μg/L (93.3 μg/L) and 13.4–62.7 μg/L (112.4 μg/L) for the Stratus (although values below the lower limit have no clinical significance). Results obtained with the BNA assay showed no statistically significant sex-related difference (P = 0.151) because >75% of the myoglobin values from the reference individuals were lower than the detection limit of the assay. The upper reference limits for the BNA were 51.4 and 49.4 μg/L for men and women, respectively. The corresponding 99th percentile values were 74.7 and 82.0 μg/L. No age-related trends (r =0.26) in myoglobin values were detected for any method.

**Discussion**

Over the past decade, there have been significant changes in the biochemical approach for the evaluation of patients with chest pain and suspected myocardial injury. Because of the rapid progress in biochemical knowledge and technological improvements, more sensitive and specific tests for early diagnosis, triage, and risk stratification of patients with acute coronary syndrome have been developed (13). Recommendations have been published for the optimal use of these new markers, suggesting as best compromise for diagnosis of myocardial infarction the association of an early marker (which increases in blood within 6 h after symptom onset) and a cardiac-specific marker, i.e., cardiac troponin (3,14). As recently confirmed in a systematic review of the literature (15), myoglobin is at present the commercially available marker that most effectively fits the role of an early marker. For this reason, practically all of the diagnostic companies offering a cardiac marker panel on various heterogeneous or homogeneous systems have implemented assays for this protein. Of course, it would be in the best interests of the medical community and of in vitro diagnostic industries that results from different assays resemble each other. The data obtained in our study show that the assays correlate well, but they do not agree in standardization. Furthermore, the differences in performance of the same method/instruments in different laboratories underline that current routine analysis of myoglobin may also lack congruency among laboratories using the same platform. Although some of the evaluated methods have recently been replaced by new efficient procedures, the results obtained in a national External Quality Assessment Scheme for cardiac markers, organized by our Working Group, confirm that for myoglobin, as well as for troponin I, the results of some commercial analyzers are significantly different from each other (data not shown). Biases observed in patients’ samples and the corresponding differences in the upper reference limits for different myoglobin assays arise directly from a standardization problem mainly attributable to the lack of an accepted reference material for myoglobin and of a high-quality reference method; a dedicated effort by international scientific bodies, such as the IFCC Committee on Standardization of Markers of Cardiac Damage, is therefore needed to achieve harmonization of myoglobin results from different assays (16).

In this multicenter study, we have also shown that some commercial assays for myoglobin determination do not meet the desired target for imprecision, which is derived directly from data on biological variation (11). Although the overall intralaboratory variation was <11% in this study, this variation may be clinically unacceptable, and an improvement in the precision of measurements is needed if myoglobin is to be offered on a routine clinical basis (12). Great attention was also given to the possible influence of different types of anticoagulants on the measured myoglobin values and to potential interferences by most important endogenous substances. National and international committees agree that plasma is the specimen of choice for cardiac marker analysis (3,14). However, we have demonstrated that different myoglobin assays can show significant anticoagulant interferences, which precludes the use of the anticoagulant to eliminate the extra time needed for clotting and to reduce the overall preanalytical time.

Interference from RF, which can mimic the measured protein, has been reported recently in commercial immunoassays for cardiac markers (17–19). We found several false positives in the Access myoglobin assay attributable to RF. In agreement with previously published findings (17), there was no concordance between RF concentrations and the degree of interference. RFs are a group of heterogeneous immunoglobulins whose common property is their ability to react with the Fc portion of the IgG molecule. This heterogeneity probably accounts for the lack of uniform interference in immunoassays (17). Interestingly, in a recent study, measurement of cardiac troponin I on the same platform was unaffected by RF, which may reflect a difference in diluent composition or in sample dilution in the two Access assays (20).

In conclusion, the results of this study stress the importance of achieving standardization of myoglobin measurements, also in terms of maximum allowable imprecision and lack of interference. Until then, the possibility of high imprecision and significant disagreement among commercial myoglobin assays should be carefully considered in clinical practice.
The manufacturers of the myoglobin assay kits examined here generously provided us with materials, information, and advice, for which we are grateful.

References