Comparison of Serum and Plasma Methylmalonic Acid Measurements in 13 Laboratories: An International Study

Christine M. Pfeiffer,* S. Jay Smith, Dayton T. Miller, and Elaine W. Gunter

Background: Detection of cobalamin deficiency is increasingly important, and methylmalonic acid (MMA) appears to be a useful marker. Information on interlaboratory variation and on methodological differences for MMA in serum and plasma is limited.

Methods: Using gas chromatography/mass spectrometry, 13 laboratories participated in a 2-day analysis of 8 serum and 11 EDTA-plasma specimens. Results were analyzed for imprecision, recovery, and differences among laboratories and methods.

Results: The mean among-laboratory imprecision (CV) was 19% and 21% for serum and plasma samples, respectively, and 9.3% and 7.8% for serum and plasma samples with added MMA, respectively. The mean within-laboratory (among-run) CV was 13% for both serum and plasma samples and 5.2% and 4.9% for serum and plasma samples with added MMA. Within-method imprecision was the same or higher than among-method imprecision. The mean among-laboratory recovery of MMA was 105% and 95% in serum and plasma, respectively. Most laboratories showed a proportional bias relative to the consensus mean of up to 15%. Two laboratories reported results that on average were almost 30% higher than the consensus mean.

Conclusions: No method differences were found, but significant among-laboratory imprecision was found in the present study. Improvements are needed to reduce the analytical imprecision of most laboratories, and attention must be focused on calibration issues. Differences among laboratories can be improved by introducing high-quality reference materials and by instituting external quality assessment programs.

© 1999 American Association for Clinical Chemistry

In recent years, there has been much interest in improving the clinical specificity and sensitivity of the tests used to diagnose cobalamin deficiency (1–4). Although measurement of serum cobalamin provides adequate specificity, the diagnostic efficiency of the test is too low because of less-than-perfect sensitivity (5). Determination of serum methylmalonic acid (MMA) concentrations has received particular attention because some authorities have suggested that it may have better diagnostic accuracy than serum cobalamin for diagnosis of cobalamin deficiency (6). However, methods for measuring MMA are costly and cumbersome, and commercial assays currently are not available. In addition, information about interlaboratory variations and methodological differences is limited.

An earlier external quality assessment study addressed the question of interlaboratory variation and of recovery of added MMA to different matrices (aqueous, serum, and heparin/NaF plasma) but included only one sample for each matrix type (7).

To assess the comparability of results among laboratories and within laboratories (among-run), CDC invited national and international clinicians and laboratorians who routinely measure MMA in serum or plasma to participate in a round-robin interlaboratory comparison study. Each participant was asked to analyze aliquots of 8 serum and 11 EDTA-plasma specimens on 2 days.

Materials and Methods

Participating Laboratories

We invited laboratories to analyze two identical, blinded sets of 8 human serum (analysis required) and 11 human EDTA-plasma specimens (analysis optional) covering the normal and increased range of MMA (≤10,000 nmol/L) on 2 days. The participants included clinical research facilities, academic laboratories, one clinical reference laboratory, and one government laboratory. Laboratories 2 and 5 analyzed serum specimens only.

Received July 14, 1999; accepted September 10, 1999.

*Address correspondence to this author at: Centers for Disease Control and Prevention, 4770 Buford Hwy., NE, MS F-18, Atlanta, GA 30341-3724. Fax 770-488-4699; e-mail cfp8@cdc.gov.
MMA concentration within each group, but the CV corrected for the increasing variance and was relatively constant within each group. Thus, we calculated for each sample the among-laboratory CV of the participating laboratories. We expressed the imprecision as the mean CV (SD) for each sample group (native serum, native plasma, serum with added MMA, and plasma with added MMA). We calculated the among-laboratory, within-method imprecision for only two method groups because one group was represented by one laboratory only.

In the absence of target values for the samples analyzed and because all laboratories used a GC/MS method, we arbitrarily considered the consensus mean of all laboratories as a point of reference. Possible systematic biases were assessed by computing 95% confidence intervals (mean difference ± confidence limit) for the mean differences between the consensus results and each laboratory’s results (10, 11). We assessed limits of agreement by calculating the central 0.95 intervals (mean difference ± 2 SD). The mean differences and the mean between the consensus results and each laboratory’s results were correlated to test for a relationship between these two variables. To assess the mean proportional bias between the consensus results and each laboratory’s results, we calculated the relative ratios of the consensus and test-laboratory results.

We calculated recoveries individually for each sample containing added MMA: recovery (%) = (specimen with added MMA – specimen without added MMA)/added concentration of MMA. Recovery results were reported as the mean (SD) of the added MMA concentrations over the 2 days of analysis.

To test for methodological differences, we grouped laboratories by method of analysis (GC/MS using cyclohexanol/HCl, silylation, or ethylchloroformate as derivatization reagent) and performed a two-way ANOVA with laboratory and analytical method as variables using the SAS GLM procedure and the Bonferroni test (to correct for multiple comparisons). *P < 0.05* was considered statistically significant.

**Results**

**Participants and Methods**

Details to the methods used by the 13 participating laboratories are listed in Table 1. All laboratories use d3-MMA as the internal standard, and most laboratories use MMA as calibrator and calibrate daily (laboratories 2, 3, and 7–13); laboratory 4 calibrates on a monthly basis. None of the laboratories routinely performs duplicate analyses of the samples.

**Imprecision**

The mean among-laboratory CVs are similar for serum and plasma specimens without added MMA but are approximately one-half for serum and plasma specimens with added MMA (Table 2). For the GC/MS method using cyclohexanol/HCl for derivatization, the among-
laboratory, within-method CV was almost the same as the among-laboratory, among-method CV. However, for the GC/MS method using silylation for derivatization, the among-laboratory, within-method CV was higher than the among-laboratory, among-method CV. Although differences among laboratory means were statistically significant for all four sample groups, for serum and plasma without added MMA they were highly significant ($P < 0.0001$).

Approximately one-half of the laboratories showed among-run CVs higher than 10% for serum and plasma specimens without added MMA (Table 3). For serum and plasma specimens with added MMA, only $\sim$10% of the laboratories obtained among-run CVs higher than 10%.

**RECOVERY**

The mean among-laboratory differences between the three serum samples containing added MMA and the native serum sample were not significantly different from the expected results of 500, 2000, and 10 000 nmol/L, respectively ($P = 0.85, 0.19$, and $0.25$, respectively). The mean among-laboratory difference between the plasma sample containing added MMA and the native plasma sample was significantly different from the expected result of 5000 nmol/L ($P = 0.01$); it was 5.4% lower than expected. Recoveries were 85–115% for laboratories other than laboratory 7 (127% ± 8.4%; Table 3). The mean among-laboratory recovery was higher in serum (105%) than in plasma (95%), but in both matrices approached complete recovery.

**DIFFERENCES AMONG LABORATORIES AND METHODS**

Fig. 1 shows the performance of each laboratory for native serum and plasma samples (Fig. 1A) and for serum and plasma samples with added MMA (Fig. 1B). We obtained very similar graphs when we examined serum specimens only or plasma specimens only (data not shown).

Among-laboratory differences were assessed with native samples only to avoid unbalanced results by a few samples with high concentrations. Because of the relatively small number of samples in this study, we reported the median, minimum, and maximum differences between the consensus results and each laboratory’s results in addition to the mean differences (Table 4). The 95% confidence intervals of the mean differences showed an apparent positive bias for laboratories 1, 4, 6, and 13, and an apparent negative bias for laboratories 3, 5, 7, and 9–12. Laboratories 2 and 8 showed no apparent bias with respect to the consensus mean.

The central 0.95 interval (mean difference ± 2 SD) gives an indication of the agreement between the consensus results and each laboratory’s results in addition to the mean differences (Table 4). The 95% confidence intervals of the mean differences showed an apparent positive bias for laboratories 1, 4, 6, and 13, and an apparent negative bias for laboratories 3, 5, 7, and 9–12. Laboratories 2 and 8 showed no apparent bias with respect to the consensus mean.

**Table 1. Participating laboratories using GC/MS methodology.**

<table>
<thead>
<tr>
<th>Lab</th>
<th>Extraction/derivatization</th>
<th>Calibrator</th>
<th>Calibration range, nmol/L</th>
<th>Calibration in</th>
<th>Sample volume, mL</th>
<th>Reference interval, nmol/L</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bond-Elut SAX/cyclohexanol-HCl</td>
<td>d$_3$-MMA</td>
<td>NA</td>
<td>Buffer</td>
<td>0.55</td>
<td>79–376</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Bond-Elut SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>0–1000</td>
<td>Buffer</td>
<td>0.20</td>
<td>53–345</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Bond-Elut SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>0–20000</td>
<td>Water</td>
<td>0.28</td>
<td>60–280</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Accu-Bond SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>0–20000</td>
<td>Water</td>
<td>0.60</td>
<td>&lt;400</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>AG MP-1/silylation</td>
<td>d$_3$-MMA</td>
<td>NA</td>
<td>Serum</td>
<td>0.40</td>
<td>73–271</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>AG MP-1/silylation</td>
<td>d$_3$-MMA</td>
<td>NA</td>
<td>Water</td>
<td>0.50</td>
<td>73–376</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>100–1000</td>
<td>Methanol</td>
<td>0.50</td>
<td>&lt;400</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>Isolute SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>50–2500</td>
<td>Water</td>
<td>0.50</td>
<td>80–280</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>Isolute SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>100–2000</td>
<td>Bovine albumin</td>
<td>1.0</td>
<td>100–2000</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>Isolute SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>0–1000</td>
<td>Water</td>
<td>0.60</td>
<td>280</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>SAX/silylation</td>
<td>MMA</td>
<td>200–1000</td>
<td>Water</td>
<td>0.25</td>
<td>&lt;400</td>
<td>8, 18</td>
</tr>
<tr>
<td>12</td>
<td>Isolute SAX/cyclohexanol-HCl</td>
<td>MMA</td>
<td>0–750</td>
<td>PBS</td>
<td>1.0</td>
<td>80–280</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>EtOH, CHCl$_3$/ethylchloroformate</td>
<td>Defined serum</td>
<td>550</td>
<td>Serum</td>
<td>0.10</td>
<td>50–260</td>
<td>19</td>
</tr>
</tbody>
</table>

* SAX, strong anion exchange; PBS, phosphate-buffered saline; EtOH, ethanol.

**Table 2. Imprecision: Among-laboratory, among-method CV and among-laboratory, within-method CV.**

<table>
<thead>
<tr>
<th></th>
<th>Serum</th>
<th>Plasma</th>
<th>Serum + MMA</th>
<th>Plasma + MMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among-laboratory, among-method CV, mean (SD), %</td>
<td>13</td>
<td>19 (6.9)</td>
<td>21 (6.0)</td>
<td>9.3 (2.1)</td>
</tr>
<tr>
<td>Among-laboratory, within-method CV, mean (SD), %</td>
<td>9</td>
<td>18 (7.9)</td>
<td>21 (5.6)</td>
<td>7.6 (1.6)</td>
</tr>
<tr>
<td>GC/MS (cyclohexanol/HCl)</td>
<td>3</td>
<td>28 (7.3)</td>
<td>25 (11.2)</td>
<td>13 (4.4)</td>
</tr>
</tbody>
</table>

* n, number of laboratories.

* GC/MS (ethylchloroformate) was represented by only one laboratory.
sponds to a proportional bias of 29.7% and 27.3%, respectively. Ninety-five percent of the results from all other laboratories were up to 15% lower or higher than the consensus mean. When we correlated the mean differences and the mean between the consensus results and each laboratory’s results, we found a relationship between these two variables for laboratories 7 (\( r^2 = 0.801 \)), 12 (\( r^2 = 0.628 \)), and 13 (\( r^2 = 0.632 \)). For all other laboratories, \( r^2 \) was 0.4.

When we grouped laboratories by method, we found no significant difference between the results of the method groups.

### Discussion

This interlaboratory comparison study for serum and plasma MMA brought together 14 highly esteemed national and international laboratories performing the most frequently used methods to assess MMA. The study design allowed us to make among-run comparisons in addition to among-laboratory comparisons. Furthermore, we were able to assess recoveries of added synthetic MMA to serum and EDTA plasma. However, laboratory- and method-specific bias comments and conclusions in this study are, to a certain extent, speculative because of the relatively small number of samples analyzed.

The mean among-laboratory CVs of 19% and 21% for serum and plasma samples without added MMA in this study are in good agreement with the 18% and 17% among-laboratory CVs found by Møller et al. (7) for one serum and one heparin/NaF plasma specimen, respectively. For serum and plasma samples with added MMA, we found smaller among-laboratory CVs (9% and 8%) than did Møller et al. (17% and 11%) (7).

The present study, in which each laboratory and method used its own calibrators, has demonstrated again that among-laboratory, within-method imprecision can exceed among-laboratory, among-method imprecision. The same observation was made in our recently performed homocysteine interlaboratory comparison study (12). Of the nine laboratories performing GC/MS with cyclohexanol/HCl derivatization, eight showed no or an apparent small negative or positive bias relative to the consensus mean, whereas one laboratory reported results on average 30% higher than the consensus mean. Of the three laboratories performing GC/MS with silylation for derivatization, two showed an apparent small negative proportional bias relative to the consensus mean, whereas one laboratory reported results on average 27% higher than the consensus mean. This suggests that differences among laboratories are most likely related to calibration issues. Although calibration curves for MMA are linear over a wide concentration range, the intercept of the calibration curve increases with an increasing calibration range. Thus, the calibration range must be optimized to obtain the most accurate results for the clinically critical MMA concentrations. If the calibration range is too wide, low MMA concentrations might be underestimated. If the calibration range is too narrow, high MMA concentrations might be overestimated. It seems reasonable to strive for highest accuracy at low and slightly increased MMA concentrations (up to 1000 nmol/L) because diagnosis of B12 deficiency is more difficult to determine at these concentrations.

Although we found no correlation between the performance of the different laboratories and the calibration range or the matrix in which calibration was performed, results indicated that it might be inaccurate to use the internal standard \( d_3 \)-MMA as a calibrator. Two of the three laboratories that did not calibrate with MMA reported results that were on average almost 30% higher than the consensus mean. Thus, if the results of those two laboratories were excluded from the consensus mean,
Fig. 1. Plots showing the performance of each laboratory for native serum and plasma samples (A) and for serum and plasma samples with added MMA (B).

(A), the individual results (mean of day 1 and day 2 measurements) of each laboratory are represented by a triangle on a vertical bar representing the range of results. The horizontal line represents the consensus mean of all laboratories. The means of all results of each laboratory are connected with a line. Laboratories 2 and 5 analyzed serum samples only. (B), the results (mean of day 1 and day 2 measurements) of each laboratory for a particular sample are connected by a line. Each sample is represented by a different symbol. Laboratories 2 and 5 analyzed serum samples only.
This international round robin for serum and plasma MMA showed no method differences, but it did show significant among-laboratory imprecision among some of the most experienced laboratories using different GC/MS methods. The analysis for analytical quality specifications has shown that there is an urgent need to improve analytical imprecision. In addition, the analysis for among-laboratory imprecision suggests that most of this variation can be attributed to calibration issues. Differences among laboratories can be improved by introducing high-quality reference materials and instituting more external quality assessment programs.

This work was supported by the National Center for Environmental Health, Centers for Disease Control and Prevention. We thank the following laboratorians for their participation and support: Margaretha Fornelius, Nova Medical AB, Eskilstuna, Sweden; Vito Gatautis, Cleveland Clinic Foundation, Cleveland, OH; Anders Johnsen and Kirsten Amstrup, The National University Hospital, Copenhagen, Denmark; Anders Liden, University Hospital, Lund, Sweden; Stellan Lindberg, Hospital of Engelholm, Engelholm, Sweden; Yvonne Marde, Karolinska Hospital, Stockholm, Sweden; Joshua Miller and Ralph Green, University of California Davis, Sacramento, CA; Jan Møller and Karsten Rasmussen, Aarhus University, Aarhus, Denmark; Christine Pfeiffer, CDC NHANES Laboratory, Atlanta, GA; Piero Rinaldo, Mayo Clinic, Rochester, MN; Joern Schneede, University of Bergen, Bergen, Norway; Sally Stabler and Robert Allen, University of Colorado, Denver, CO; and Peter Wentz, LabCorp, Burlington, NC. We also thank the staff members of the NHANES Laboratory, CDC, for their contributions in preparing and shipping these materials to participating laboratories.
References


