Evaluation of Novel Assays in Clinical Chemistry: Quantification of Plasma Total Homocysteine

Ebba Nexo,† Frode Engbaek,1 Per Magne Ueland,2 Christina Westby,3 Paudy O’Gorman,4 Carole Johnston,5 Bengt Frode Kase,6 Anne B. Guttormsen,2 Ingrid Alfheim,3 Joseph McPartlin,4 David Smith,5 Jan Møller,1 Karsten Rasmussen,1 Robert Clarke,5 John M. Scott,4 and Helga Refsum2

Background: There is a need for systematic evaluation of methods before their release to the market. We addressed this problem in novel homocysteine assays as part of an European Demonstration Project involving six centers in four countries.

Methods: Two immunological methods for measurement of plasma total homocysteine (P-tHcy), the fluorescence polarization immunoassay (FPIA) and the enzyme immunoassay (EIA), were compared with two comparison methods, HPLC and gas chromatography–mass spectrometry (GC-MS). All laboratories performed the following procedures: (a) familiarization; (b) determination of linearity and precision by analyzing five plasma samples with interrelated concentrations for 20 days; (c) correlation using patients’ samples; and (d) assessment of long-term performance.

Results: Both immunological methods were linear for P-tHcy between 5 and 45 µmol/L. The intralaboratory imprecision (CV) was <5% for FPIA and <9% for EIA used with a sample processor. The bias was −2% to 3% for FPIA and 2–4% for EIA used with a sample processor.

Conclusions: The immunological methods provide results with little bias compared with HPLC and GC-MS.

New assays should be thoroughly evaluated before introduction for routine use in clinical practice. However, this requires that diagnostic laboratories have the resources and the analytical capacity to compare the performance of the new assays with comparison methods. The outcome of such an evaluation will depend on the individual laboratories, thus leading to difficulties in distinguishing the performance of the method from the performance of the laboratory. This applies to both the new assay and the comparison method(s).

New methods for measurement of plasma total homocysteine (P-tHcy)7 currently are under evaluation. Until recently, measurement of P-tHcy required time-consuming and complex methods such as HPLC and gas chromatography–mass spectroscopy (GC-MS) (1, 2), but now an immunological method for the determination of P-tHcy has been developed and subsequently used in both a fluorescence polarization immunoassay (FPIA) format in conjunction with the IMX® analyzer (3) and in an enzyme immunoassay (EIA) format (4). These assays are simple and may be superior to the conventional chromatographic assays for routine use.

The introduction of these novel tests for P-tHcy illustrates a typical situation in clinical chemistry today. The assays were developed for commercial use, and the producer provided extensive documentation on specificity, analytical range, reproducibility, accuracy, and linearity. However, systematic data on the performance in individual routine laboratories of various sizes and experience

---

1 Department of Clinical Biochemistry AKH and Skejby Hospital, Aarhus University Hospital, DK-8000 Aarhus C, Denmark.
2 Department of Pharmacology, University of Bergen, 5021 Bergen, Norway.
3 Axis Biochemicals, 0510 Oslo, Norway.
4 Department of Biochemistry, Trinity College, University of Dublin, Dublin 2, Ireland.
5 Department of Pharmacology, University of Oxford, Oxford OX1 3QT, United Kingdom.
6 Department of Pediatric Research, University of Oslo, 0316 Oslo 3, Norway.

*Address correspondence to this author at: Department of Clinical Biochemistry, AKH, Aarhus University Hospital, Norrebrogade 44, DK-8000 Aarhus C, Denmark. Fax 45-8949-3068; e-mail ene@post9.tele.dk.

†Received December 27, 1999; accepted June 2, 2000.

7 Nonstandard abbreviations: P-tHcy, plasma total homocysteine; GC-MS, gas chromatography–mass spectroscopy; FPIA, fluorescence polarization immunoassay; and EIA, enzyme immunoassay.
are lacking, not only for the P-tHcy assays but for nearly all tests introduced into routine laboratory practice.

This report describes a strategy to evaluate novel P-tHcy assays. The study is part of a European Union-funded demonstration project involving six centers in four countries. The objective was to evaluate the performance of these assays in a range of settings reflecting their future use. The strategy involved the following procedures: (a) familiarization; (b) assessment of linearity and imprecision; (c) assessment of correlation with the comparison method; and (d) long-term performance using samples from patients.

**Materials and Methods**

**REAGENTS AND EQUIPMENT**

Both the FPIA and the EIA were supplied as test kits from AXIS, and the assays were carried out according to the instructions supplied by the manufacturer. The FPIA assay was run on an IMx analyzer (Abbott). The EIA assay was performed either manually with the use of plate washers and plate readers ( Labsystems or Wallac) or using a pipetting robot equipped with a plate reader and washer (Rosys Plato 7 or Tecan Genesis RMP 100). The FPIA and EIA used calibrators manufactured by AXIS as part of the test kits, whereas the GC-MS and HPLC methods used in-house calibrators. Selected information about the operating characteristics of the two immunological methods and the GC-MS and HPLC comparison methods are shown in Table 1.

**SAMPLES**

**Samples for testing linearity and imprecision.** Two pools of human plasma with no exogenous tHcy were prepared, with one containing 45 μmol/L P-tHcy [S1; close to the upper limit for the immunoassay (4)] and the other containing 5 μmol/L P-tHcy [S5; close to the lower reference limit of an adult population (5)]. S1 and S5 were then mixed so that samples S2, S3, and S4 contained 50%, 25%, and 12.5% S1 and 50%, 75%, and 87.5% S5, respectively. The tHcy concentrations of S2, S3, and S4 were calculated from the fractional composition of S1 and S5. Aliquots of each S sample were frozen at −20 °C. In each laboratory, samples S1–S5 were thawed and further aliquoted for the analysis of each sample on three occasions daily (FPIA) or twice daily as duplicates (EIA) for 20 consecutive working days.

**Patients’ samples used in the correlation analysis.** Plasma samples from 57 individuals covering a P-tHcy concentration range of 4–34 μmol/L were collected from individuals screened at the University of Bergen clinical chemistry laboratory and stored at −20 °C until analysis.

**Control samples.** Internal control samples at three concentrations (~8, 12, and 25 μmol/L) were supplied by the manufacturer (AXIS) and used to accept or reject the individual runs, according to prespecified rules (6). External quality assessment samples were supplied by the Danish Institute for External Quality Assurance for Hospital Laboratories and analyzed on six occasions to test the long-term performance at each center. All samples, with the exception of the external assessment samples, were shipped to each of the participating laboratories on dry ice.

**LABORATORIES**

A total of six laboratories in four countries participated in the demonstration project.

**Laboratory 1.** Laboratory 1 developed and carried out the premarket testing of the immunoassays. The laboratory participated in the protocol with the FPIA and the EIA using automatic sample processing. The inclusion of this laboratory permitted a comparison of the performance of the assay producer with the routine laboratories.

**Laboratory 2.** Laboratory 2 was a research laboratory with more than 15 years of experience with P-tHcy determinations and serves as a reference laboratory for P-tHcy determinations. This laboratory developed a method in 1985 that formed the basis for the novel immunoassays (7). The fully automated HPLC method, developed in 1989 and modified in 1993, has been widely used in several large clinical and epidemiological studies on P-tHcy (8–10). For this project, this laboratory measured P-tHcy with four methods, EIA, FPIA, a HPLC technique based on monobromobimane derivatization and fluorescence detection (11), and a GC-MS method involving

| Table 1. Selected characteristics for calibration, sample requirements, and performance of various P-tHcy assays. |
|---|---|---|---|---|---|---|
| **No. of calibrators** | **Sample volume, μL** | **No. of measurements per sample** | **No. of samples per run** | **Time per run, h** | **Ref.** |
| FPIA | 6 | 50 | 1 | 20 | 1 | (3) |
| EIA | 6 | 5 | 2 | 84 | 2.5 | (4) |
| HPLC | 1 | 30 | 1 | 68 | 24 | (8) |
| Lab 2 | 2 | 50 | 2 | 96 | 12 | (13, 14) |
| Lab 3 | 1 | 100 | 1 | 96 | 24 | (12) |
| Lab 6 | 3 | 200 | 1 | 90 | 24 | (15, 16) |
ethylchloroformate derivatization as described by Husek (12). The EIA method was run using an automated sample processor.

Laboratory 3. Laboratory 3 had performed P-tHcy determination with HPLC based on 7-fluorobenzofurazane-4-sulfonic acid derivatization and fluorescence detection (13,14) for ~5 years. This laboratory measured P-tHcy using both HPLC and FPIA.

Laboratories 4 and 5. Laboratories 4 and 5 were laboratories with no previous experience in the determination of P-tHcy, and both measured P-tHcy using the EIA with manual sample processing.

Laboratory 6. Laboratory 6 included two laboratory units. One was responsible for the GC-MS comparison method (15,16) and had almost 10 years of experience with P-tHcy determination in clinical practice as well as in research. The other unit was comparable to a routine clinical chemistry laboratory introducing P-tHcy determination (the FPIA variant) for the first time.

STUDY DESIGN
The following steps were performed.

Familiarization. After installation of the equipment, the technicians received practical training. This involved performing a run that included calibration and analysis of internal controls using a specified protocol. The acceptance of the internal controls was based on the limits supplied by the manufacturer.

Linearity and imprecision. Samples S1–S5 were run in random order on three occasions per day together with the three internal control samples for the FPIA or twice daily in duplicate for the EIA on 20 consecutive working days. Runs were accepted if the internal controls were within the limits stipulated by the manufacturer. The mean P-tHcy concentrations and SDs for the internal controls were calculated and used for acceptance or rejection in the subsequent runs.

Correlation analyses. The 57 samples from patients were analyzed once, and runs were accepted based on values obtained for the internal controls.

Long-term performance and external quality assessment. Over a 12-month period, six pairs of samples from an external quality assessment scheme were run within 2 weeks of receipt.

One laboratory continued the FPIA assay on a routine basis to test its practicability. The assay was run once a week by seven different technicians over a 5-month period. The number of rejected runs was recorded, and the imprecision was calculated based on the values obtained for the internal controls.

STATISTICAL METHODS
The within-day, between-day, intralaboratory, between-laboratory, and interlaboratory variances were calculated using nested analysis of variance. The interlaboratory variance is the variation between results obtained from samples run in various laboratories. The intralaboratory variance is the variation for results obtained from samples run over several days in the same laboratory. The within-day variance describes the variation for results obtained from samples assayed in the same run. The between-day variance was calculated by subtracting the within-day variance from the intralaboratory variance. Similarly, the between-laboratory variance was calculated by subtracting the intralaboratory variance from the interlaboratory variance. Bias was calculated from the results obtained for the 5 samples and for the patients’ samples. Bias was defined as the difference between the test method result and the comparison method result divided by the comparison method result. The results from the 57 patients’ samples obtained by the different methods were analyzed by linear regression and according to the procedure of Bland and Altman (17,18). Results obtained by GC-MS (laboratory 6) were used as the comparison method or “gold standard”, and this laboratory unit was not involved in running the novel immunological assays.

Results
Six laboratories participated in the study. The FPIA and the EIA methods were both evaluated at four sites, whereas the HPLC and GC-MS methods were both evaluated at two laboratories. The EIA format was performed either manually or using a sample processor (see Materials and Methods), and the results are reported separately.

FAMILIARIZATION
The six laboratories were familiarized with the immunological methods within a few weeks of installation of the equipment.

ASSESSMENT OF LINEARITY AND PRECISION
Based on the results of the analysis of five samples (S1–S5) with interrelated concentrations, the linearity and imprecision of the methods were assessed and compared with similar data for the comparison methods. In general, all methods showed linearity throughout the 5–45 μmol/L P-tHcy range (Fig. 1). A statistically significant deviation from linearity was observed for the FPIA run at laboratory 3 because of results obtained for sample S2.

The intralaboratory variation for the various assays is shown in Table 2. The intralaboratory imprecision (CV) was <5% for FPIA, <9% for the automated EIA, and <13% for the manual EIA. The interlaboratory imprecision for the FPIA was 3–5% for the range of values studied and was comparable to values obtained with the comparison methods. The interlaboratory imprecision for the EIA was somewhat higher: 6–9% when a sample processor...
was used, and 10–17% when the assay was performed by manual pipetting.

The proportions of the variance derived from the within-day, between-day, and between-laboratory variances were calculated using nested analyses of the variance (17); the results are shown in Fig. 2. For the FPIA, the within-day variance accounted for 27–63% of the variance and the between-day variation accounted for most of the

---

**Table 2. Assessment of intralaboratory imprecision (CV) for immunological assays and comparison methods for P-tHcy determination.**

<table>
<thead>
<tr>
<th>Method</th>
<th>S1 (44.7 μmol/L)</th>
<th>S2 (24.7 μmol/L)</th>
<th>S3 (14.7 μmol/L)</th>
<th>S4 (9.8 μmol/L)</th>
<th>S5 (4.7 μmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPIA</td>
<td>3.2</td>
<td>4.5</td>
<td>3.5</td>
<td>3.7</td>
<td>4.9</td>
</tr>
<tr>
<td>aEIAab</td>
<td>6.2</td>
<td>6.4</td>
<td>6.9</td>
<td>7.6</td>
<td>8.8</td>
</tr>
<tr>
<td>mEIAc</td>
<td>10</td>
<td>9.0</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>HPLC</td>
<td>4.1</td>
<td>3.5</td>
<td>3.4</td>
<td>4.8</td>
<td>5.8</td>
</tr>
<tr>
<td>GC-MS</td>
<td>2.8</td>
<td>2.1</td>
<td>2.1</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>3.5</td>
<td>3.5</td>
<td>3.0</td>
<td>2.9</td>
</tr>
</tbody>
</table>

* The five samples (S1–S5) with interrelated concentrations (see Materials and Methods) were analyzed two or three times daily for 20 days.
* aEIA, EIA run with sample processor.
* mEIA, EIA run with manual sample handling.
remaining variance. The between-laboratory contribution was marginal. A similar pattern was obtained with the EIA performed with a sample processor, whereas the between-laboratory component was considerably larger for EIA carried out by manual pipetting.

The bias relative to the comparison method is shown in Table 3. Based on the samples, the FPIA showed a negative bias of ~2%, whereas one of the laboratories showed a small positive bias for the EIA. The most significant bias was observed between the two GC-MS methods, and this was most likely attributable to differences in the calibrators used in each laboratory.

**Table 3. Assessment of bias for the immunological assays relative to the GC-MS comparison method for P-tHcy determination using both the S samples and the patients’ samples.**

<table>
<thead>
<tr>
<th></th>
<th>Bias, mean (SE), %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S samples (n = 5)</td>
</tr>
<tr>
<td>FPIA</td>
<td></td>
</tr>
<tr>
<td>Lab 1</td>
<td>-1 (1.0)</td>
</tr>
<tr>
<td>Lab 2</td>
<td>-2 (0.6)</td>
</tr>
<tr>
<td>Lab 3</td>
<td>-1 (1.1)</td>
</tr>
<tr>
<td>Lab 6</td>
<td>-2 (0.6)</td>
</tr>
<tr>
<td>aEIA</td>
<td></td>
</tr>
<tr>
<td>Lab 1</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Lab 2</td>
<td>3 (1.2)</td>
</tr>
<tr>
<td>mEIA</td>
<td></td>
</tr>
<tr>
<td>Lab 1</td>
<td>10 (4)</td>
</tr>
<tr>
<td>Lab 2</td>
<td>-2 (2.4)</td>
</tr>
<tr>
<td>HPLC</td>
<td></td>
</tr>
<tr>
<td>Lab 2</td>
<td>-3 (0.9)</td>
</tr>
<tr>
<td>Lab 3</td>
<td>ND</td>
</tr>
<tr>
<td>GC-MS</td>
<td></td>
</tr>
<tr>
<td>Lab 2</td>
<td>-5 (0.6)</td>
</tr>
</tbody>
</table>

*Significance: *P* < 0.05; *P* < 0.01.

**CORRELATION ANALYSES**
Analysis of patients’ samples throughout the range of P-tHcy values between 4 and 34 μmol/L allowed a comparison of results obtained with the novel immunological methods with the GC-MS method used as gold standard (Fig. 3). Exclusion of one outlier (34 μmol/L) did not alter the overall results of the regression analysis (data not shown). Regression analysis based on all 57 samples showed no significant deviation from a slope of 1 and an intercept of 0 for three of the four laboratories performing FPIA and three of the four laboratories performing EIA. The deviations were significant for the remaining two laboratories (laboratories 3 and 5), showing slopes varying 8–11% from a slope of 1 and intercepts deviating −0.5 and 2 μmol/L for the FPIA and EIA assays, respectively. By comparison, similar data for the two HPLC methods (laboratories 2 and 3) showed 9% and 11% deviation from a slope of 1, and −0.8 and 0.9 μmol/L deviation from an intercept of 0. The other GC-MS method (laboratory 2) showed a deviation of 4% from a slope of 1 and no deviation for the intercept. Laboratory 3 obtained a similar deviation for the FPIA and HPLC. This deviation is unlikely to reflect calibration because all immunological assays used the same calibrators, whereas the HPLC used local calibrators.

Bias was calculated using the patients’ samples, and the results are shown in Table 3. One of the manual EIA assays showed a positive bias of 9%, but none of the other assays had a bias exceeding 4%.

**LONG-TERM PERFORMANCE**
All laboratories participated in an external quality assessment program (19). Over a 12-month period, each laboratory returned results for 12 samples. The results obtained with the FPIA method and three of the comparison methods were all within the 90th percentile. Four of the 48 results obtained with the EIA method carried out in three laboratories were outside the 90th percentile. In addition, four of the results obtained with the comparison method (HPLC; laboratory 3) were outside the 90th percentile.
The performance of the FPIA run by seven technicians over a 5-month period was excellent. Only 1 of the 60 runs required repeat measurement because of internal controls being outside the accepted limits, and the CV (obtained for the internal controls covering the range 7–25 μmol/L) was <3%.

Discussion

An increasing number of new analyses developed by commercial firms are being introduced into routine clinical laboratories. Frequently, the introduction involves a comparison with existing methods and an evaluation of the reproducibility of the new method (20, 21). We used a simple model that can be used as a template for multilaboratory testing of new methods before their introduction into routine clinical practice. The model was used to evaluate the performance of novel immunological assays for quantification of P-tHcy.

Our strategy for the initial testing of a new method involved familiarization; estimation of linearity, imprecision, and bias; assessment of correlation with comparison methods; and evaluation of long-term performance. Our protocol included two additional strategies compared with recommended guidelines for the implementation of a new methodology into the routine laboratory (20, 21). We compared the performance in different laboratories and used human samples with interrelated concentrations. Assessment of performance using samples with interrelated concentrations allows a systematic evaluation of the new methods that is independent of the comparison method (22). This is of particular importance when the new assay is superior to the currently used methodology or when no comparison method is available.

Our evaluation of the immunological assays for P-tHcy confirms and expands the results obtained from previous studies (23–27). The practicability and low imprecision of
the FPIA method throughout the range of values tested suggest that it is suitable for routine use in laboratory practice. The accuracy of the FPIA method makes it feasible to directly compare values obtained in different laboratories. This is an important issue both in clinical studies and in routine clinical chemistry. The EIA has the advantage of a high throughput and low sample volume requirement, but the imprecision is higher than that of the FPIA, especially when combined with manual sample handling. These features make the EIA format suitable mainly for screening purposes, where detection of a substantial change in P-tHcy is required. The results obtained by the immunological methods compared well with results obtained with the comparison method, and no systematic bias of significant magnitude was observed.

In conclusion, when implementing novel assays in routine clinical practice, it is important to evaluate these assays in the relevant laboratory settings and take account of laboratory performance as well as assay performance. This project illustrates an approach that could be used when introducing other novel assays for routine clinical practice. Manufacturers of novel assays should have the responsibility to evaluate their performance in this way before their introduction for routine clinical practice.

This work was supported by European Union Commission Demonstration Project Contract No. BMH4-98-3549. The Department of Pharmacology, University of Bergen, received support from the Top Research Program and the Program for Advanced Technical Equipment of the Norwegian Research Council. We gratefully acknowledge the technical assistance of Ove Aarseth, Nasser Bastani, Elfried Blomdal, Lene Christensen, Jennifer Gjerde, Marte Hveem, Gry Kvalheim, Marit Krokeide, Catrine Kvam, Erik Larsen, Lillian Manniche, Birgit Ortmann, Ove Netland, and Bodil Thomsen.

References