Laboratory PT #4

September 2025

4th Round of Mercury Laboratory Proficiency Testing

Final Report

1. Objective

The main purpose of this proficiency testing (PT) programme was to evaluate the performance of mercury analyses conducted by the laboratories. It was expected to provide the individual proficiency levels of participating laboratories and the collective mercury monitoring capacity in the region.

2. Proficiency testing provider

This PT was organized by National Institute for Minamata Disease (NIMD) in collaboration with United Nations Environmental Programme (UNEP). Asian Institute of Technology Regional Resource Centre for Asia and the Pacific (AIT RRC.AP) distributed the test item and collected the analytical results. IDEA Consultants, Inc. prepared the test item for this PT.

3. Implementation period

Call for participation:	November 2024 –January 2025
Test item distribution:	February 2025
Duration of test (analysis):	Until 31 May 2025

4. Participation fee

Free.

5. Test item (sample) and parameter

One (1) human urine sample was used for analysing total mercury concentration. The test item was distributed in freeze-dried form, and participants dissolved the received items and conduct the analysis in liquid form.

5.1. Test item preparation

The test item was made from human urine collected from eight people in the general population. The urine was homogenized by repeatedly freeze-thawing and filtering it until no precipitates formed. 5 mL of the urine was then aliquoted into brown glass containers and freeze-dried. Two freeze-dried samples were sealed in aluminum pouches with nitrogen alongside a desiccant and were then vacuum-sealed. The obtained dried test items were then sterilized using gamma rays.

Two tubes of the test items were sealed in one aluminum-lined laminate pack for distribution to the participants. Thus, two bottles of the same testing items were distributed to each participant.

5.2. Homogeneity testing

The following homogeneity testing of the test item was conducted to ensure that there were no significant differences in the mercury concentrations in the test items between bottles that could affect the result of the PT.

After the preparation of the test item (packed in bottles), ten bottles were selected, and the total mercury analysis (acid digestion - aeration CVAAS measurement) was performed twice for each test item in a bottle.

The homogeneity of the test item was then analysed from the results of the total mercury concentrations. Since the analytical results include the uncertainty due to the (chemical) analytical procedure, homogeneity was judged by the following criterion:

Criterion: $S_s \leqq \sqrt{F_1 \times (0.3 \times \sigma_{ep})^2 + F_2 \times {S_w}^2}$

S_s: relative standard deviation of homogeneity testing

σ_{ep}: (expected) relative standard deviation of the reported results from participants

$w_i^2 = \Sigma (x_{gm}^2 - \bar{x}_g^2) / (m-1)$ x_{gm}: result of m times analysis of the bottle

$S_w^2 = \Sigma w_i^2 / g$ $\bar{x}_g$: average of the result of each bottle

F_1 and F_2 are values which are calculated from the probability distribution. In this homogeneity testing (20 bottles testing), F_1 and F_2 were applied following numbers:

F1 = 1.59 F2 = 0.57

(Even though these values are referred from the Annex B of ISO13528:2022, they are introduced from the random variables of χ^2 distribution and F distribution.)

Analysis results of this homogeneity testing are as follows:

S_s = 0.0138

S_w^2 = 0.000390

Also, relative standard deviation of the results from the participants (used for evaluation) was as follows:

σ_{ep} =0.227

This standard deviation should be used the value which was used for the performance evaluation for participants. As described in 8.2, the performance of participants was evaluated from the median and normalized interquartile range (NIQR) of the results, thus relative NIQR was used for the confirmation of the criterion.

Therefore, above criterion was judged as follows:

$$\sqrt{F_1 \times (0.3 \times \sigma_{ep})^2 + F_2 \times {S_w}^2}$$

$= \sqrt{1.59 \times (0.3 \times 0.227)^2 + 0.59 \times 0.00039}$

= 0.0872 > 0.0167 (S_s)

It was confirmed that the test item was sufficiently homogeneous to evaluate the performance of participants' results.

5.3. Stability testing

To ensure that the concentration of the target parameter (total mercury) was maintained without significant changes during the PT, a following stability testing was conducted after the duration of the analysis.

Ten test items were selected from the stored (not distributed to participants), and total mercury analysis (acid digestion - aeration CVAAS measurement) was performed twice for each test item in a bottle.

The stability of the test item was then analysed by comparing the results before and after the distribution of the test item. The stability of the test item was judged by the following criterion:

Criterion: $| \bar{x} - \bar{y} | \leqq 0.3 \times \sigma_{pt} + 2 \times \sqrt{u_{(x)}^2 + u_{(y)}^2}$

$\bar{x}$: average of the item before distribution

$\bar{y}$: average of the item after proficiency testing

$u_{(x)}$: uncertainty of the item before distribution

$u_{(y)}$: uncertainty of the item after proficiency testing

σ_{pt}: standard deviation for the proficiency evaluation. In this program, NIQR was applied to evaluation of performance of the participant.

Analysis results of test items before and after the PT are as follows:

$\bar{x}$ = 0.612 $u_{(x)}$ = 0.00228

$\bar{y}$ = 0.616 $u_{(y)}$ = 0.00228

Standard deviation of the result of all participants was as follows:

σ_{pt} = 0.128

This standard deviation should be used the value which was used for the performance evaluation for participants. As described in 8.2, the performance of participants was evaluated from the median and normalized interquartile range (NIQR) of the results, thus NIQR was used for the confirmation of the criterion.

Therefore, above criterion was judged as follows:

$0.3 \times \sigma_{pt} + 2 \times \sqrt{u_{(x)}{}^2 + u_{(y)}{}^2}$

$= 0.3 \times 0.128 + 2 \times \sqrt{0.00228^2 + 0.00228^2}$

$= 0.0448 > 0.004$ ($|\bar{x} - \bar{y}|$)

It was confirmed that the concentration of total mercury in test item was not changed during the PT.

6. Target parameter

The target parameter for the PT was total mercury. Two test items with the same content were distributed to each participant. Participants conducted two total mercury analyses for each test item and reported all four results.

The test items were distributed in freeze-dried form, and participants dissolved them into a liquid before analysis. The procedure for dissolving the test items is as follows:

1. Add 5 mL of pure water precisely to each test item.
2. Mix the item by shaking (reciprocal or rotation) for 30 minutes.
3. If any suspended solids are found in the item, continue to shaking for an additional 30 minutes.

7. Participating institutions

This PT was intended for public or university laboratories performing mercury analyses. The analysis was requested with a lower detection limit of less than 0.1 µg/L for a 2 mL test item.

50 laboratories registered for the PT, and 10 registered laboratories were unable to receive the test items due to several issues such as customs clearance. Among 40 laboratories received the samples, 35 laboratories reported their analysis results. The number of laboratories and their statuses are shown inTable 1.

Table 1 Number of participating laboratories

Category	Number of laboratories
Registered	50
Sample received	40
Result delivered	35

8. PT Analysis Result

In total, 35 laboratories submitted results. One laboratory performed multiple analysis by different groups individually and reported multiple results. The following reported results were excluded from the statistical analysis and calculation of Z score due to analytical difficulties: 1) reports with insufficient number of replicates, 2) reports with non-detects (below the LOD), 3) results identified as outliers by Grubbs's test (p = 0.05). The statistical data were calculated from the average of each participant's results (2 x 2 = 4). Although one laboratory reported more than four results; however, all data were included in the statistical analysis.

8.1. Basic statistical data of the PT result

The basic statistics of the result of PT are shown inTable 2.

Table 2 Summary of the results of the PT

Statistical data of the results (unit: μg/L)	Reject excluded
n	30
Mean	0.612
Median	0.565
Standard deviation:	0.285
Minimum	0.149
Maximum	1.390
25th percentile	0.488
75th percentile	0.660
Interquartile range (IQR)	0.173
Normalized IQR (NIQR)	0.128
Parameter related to distribution	
Skewness of distribution	1.257
Kurtosis of distribution	2.144

After exclusion of rejected results, both skewness (1.257) and kurtosis (2.144) were substantially reduced, indicating a distribution closer to normal.

The distribution of the average values of the all results excluding one whose reported values were all ND is illustrated in Fig. 1.

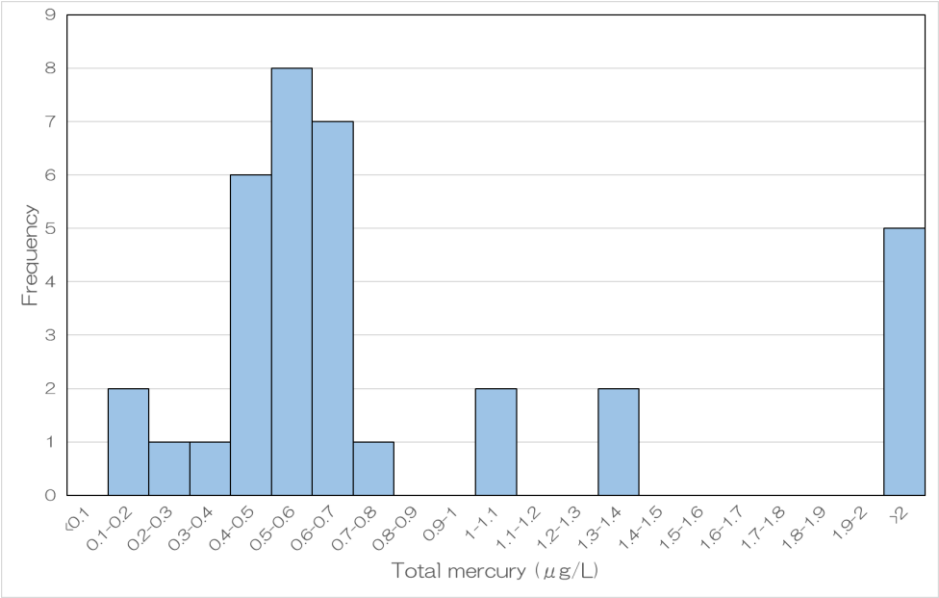


Fig. 1 Histogram of the report data

8.2. Performance evaluation for participants

Median data of laboratories excluding rejected laboratories was applied as agreement value. Performance of the results was evaluated by the robust z score, which was calculated from the median and normalized interquartile range (NIQR).

z score of each participant was calculated from the following equation.

z = [(average of reported result) – (median of laboratories excluding rejected laboratories)] / NIQR

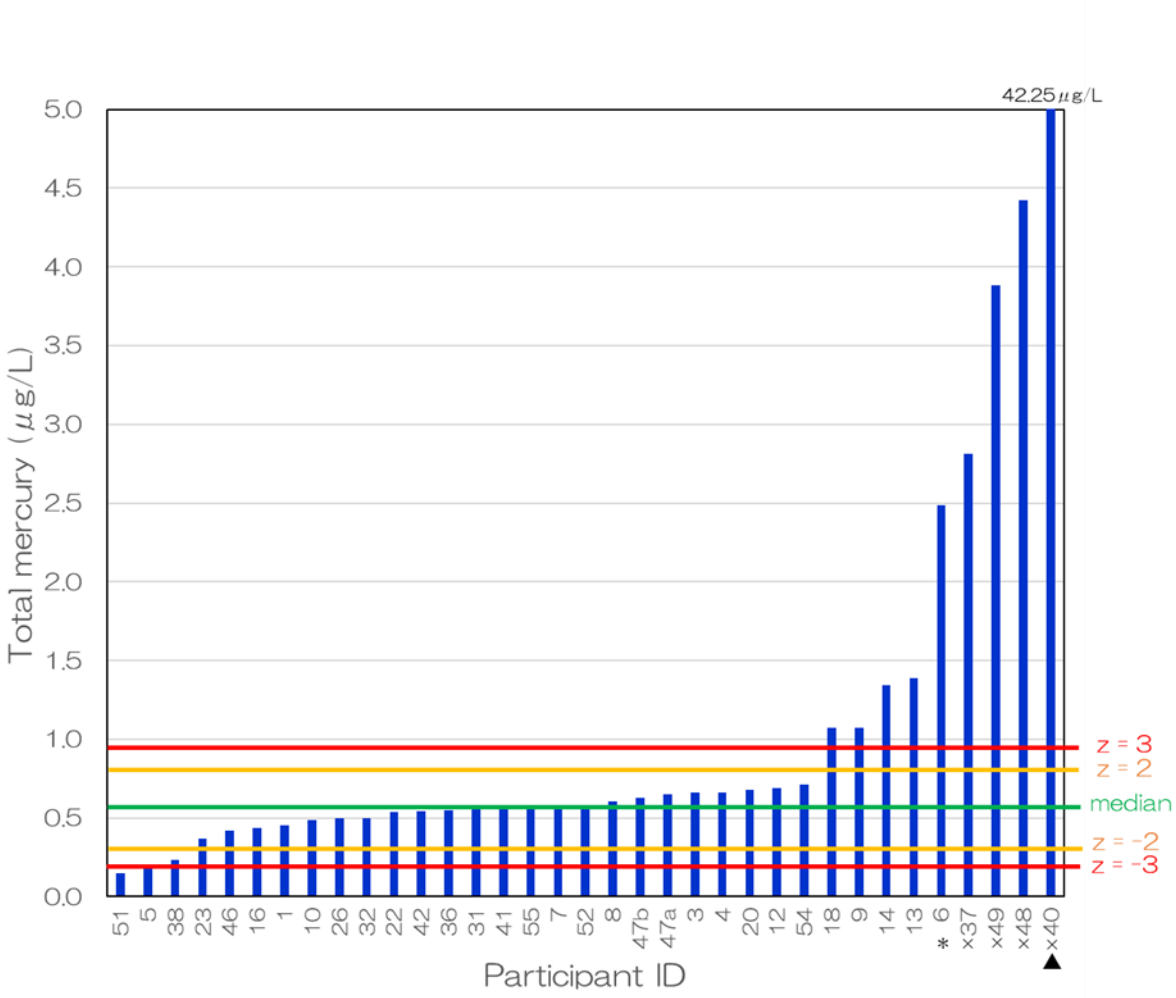
Performance of the result is classified by z score as follows:

|z| ≦2: Performance is satisfactory (satisfactory)

2 < |z| < 3: Performance is questionable (caution)

|z| ≧ 3: Performance is unsatisfactory (action)

The results and performances of laboratories are shown in Fig. 2. (The result that all reported data were non-detected is not shown.)



(Symbols below the Participant ID denote results excluded from the Z-score calculation:
× = Grubbs outlier, ▲ = insufficient replicates, * = ND report.

Fig. 2 Result of each participant

The numbers of laboratories disaggregated for each z score range are shown inTable 3. (The result that all reported data were non-detected is not included.)

Table 3 Number of the laboratories in the range of z score

z score	z ≦ -3	-3 <z< -2	-2 ≦z≦ 2	2 < z < 3	z ≧3
n	2	1	23	0	9

Although the kurtosis was large, some of the reported results were distant from the median. The outliers, 31% of the reports were "action" results (|z| ≥ 3).

8.3. Performance by regions and of type of participating institutions

The summary of the reported PT results by region is shown inTable 4. All data are included for the participant whose results were all non-detects. Due to the small number of participants in some regions, regional differences were not statically evaluated.

Table 4 Basic statistical data by Region

Statistical data of the results (unit: µg/L)	Africa	Asia and the Pacific	Eastern Europe	Latin America and the Caribbean	Western Europe and Other Group
Median	3.882	0.654	0.562	0.593	0.436
IQR	-	0.237	-	0.255	0.079
Range (min - max)	0.149-42.25	0.37-4.423	0.498-2.812	0.177-2.487	0.233-0.577

(The IQR is not shown in Region with insufficient number of results.)

The summary of the reported PT results by type of participating institutions is shown in Table 5. (All data are included for the participant whose results were all non-detect). Due to the small number of participants in some laboratory type, differences between the type of institution were not statistically evaluated.

Table 5 Basic statistical data by participating institutions

Statistical data of the results (unit: µg/L)	Academic	Government	Non-Government
Median	0.606	0.562	21.454
IQR	1.262	0.210	-
Range (min - max)	0.418-4.423	0.149-1.39	0.659-42.25

(The IQR is not shown in the institution type with insufficient number of results.)

Notably, the results from Africa and from non-governmental laboratories, the presence of extreme values influencing the overall statistics.

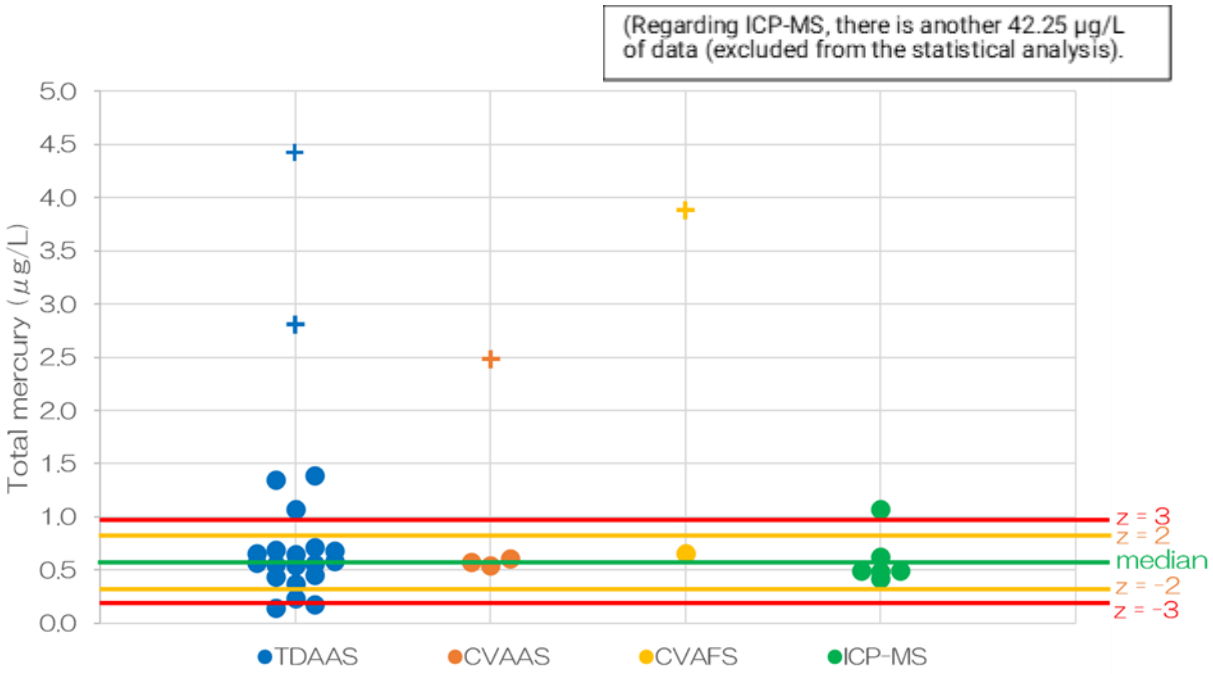
8.4. Analysis methods and results

The method of analysis was not specified for the PT and the participants performed analysis by the method that they usually used in their routine analysis, or they were planning to use in the future.

The participants performed analysis of total mercury by the following methods:

- Thermal Decomposition Cold Vapour Atomic Absorption Spectrometry (TDAAS)
- Acid digestion, aeration Cold Vapour Atomic Absorption Spectrometry (CVAAS)
- Acid digestion, Cold Vapour Atomic Fluorescence Spectrometry (CVAFS)
- Acid digestion, Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

The distribution of the results from participants by analysis method is shown in Fig. 3.



(Data excluded from the analysis are indicated by "+" marker.)

Fig. 3 Distribution of the report data by analysis method

No significant difference was found among the analysis methods performed. (Kruskal-Wallis one-way ANOVA on ranks, p=0.76. It does not contain excluded data.)

9. Deviation between sample analyses

To evaluate deviations in the chemical analyses and the effects of test item preparation (reconstitution with water), the relationship between the difference in the results of the analysis of each item (bottle) and the difference in the analysis results of the same item (bottle) is shown in Fig. 4. Results reported as less than two replicates per item, including not-detected results, are not shown in Fig. 4.

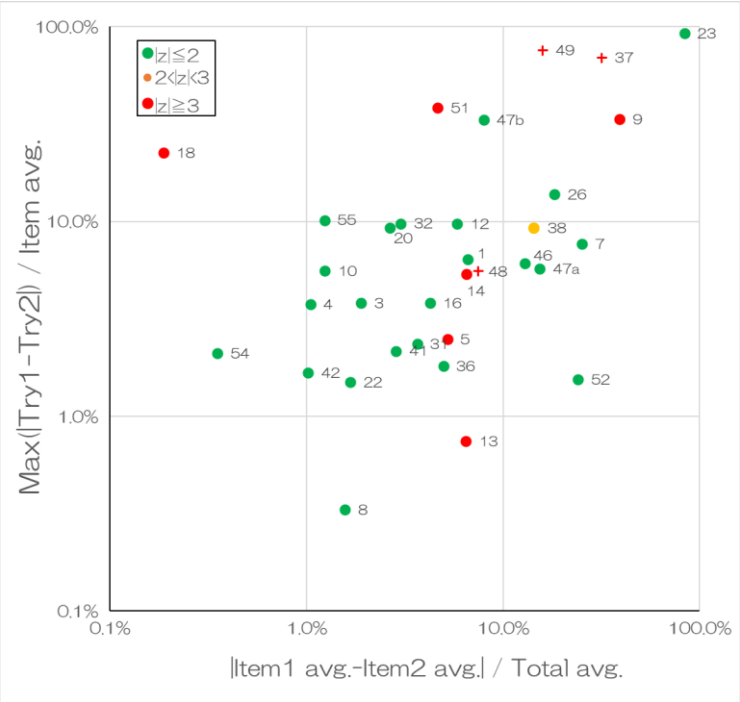


Fig. 4 Inter- and intra-item deviations expressed relative to mean values

In Fig. 4, the X-axis represents the relative difference between the mean analysis results of the distributed test items (bottles). Since the distributed items were identical, the X-axis values mainly reflects variability introduced during test items preparation process (dissolving the freeze-dried items in water). The Y-axis shows the difference between the analytical results of the same item (the larger of the two items). The Y-axis represents the relative difference between replicate analysis results of the same item. Thus, the Y-axis values indicate the contribution of analytical variability. Results excluded from statistical analysis are shown as "+" markers.

10. Conclusion

In this proficiency testing, the results reported by the participants were relatively dispersed over a wide range. Some results were much larger than the median and were excluded from the statistical analysis. The distribution of results was skewed toward the large side of the range, and more results were "action" ($|z| > 3$) on the large side of the distribution than on the small side. There were no significant differences in mercury concentrations between the analysis methods.

.

Appendix: List of participating laboratories (Non exhaustive)

Agricultural University Plovdiv	Ministry of Natural Resources and Environment of Thailand
Badan Riset dan Inovasi Nasional, Indonesia	National Central University of Taiwan
Balai Besar Biomedis dan Genomika Kesehatan, Indonesia	National Institute for Environmental Studies, Japan
Central Africa Region Mercury Monitoring and Evaluation Network, Gabon	National Institute for Minamata Disease
Esslingen University of Applied Science	National Institute of Industrial Technology, Argentina
GEOMAR Hemholtz Centre for Ocean Research Kiel	National Public Health Surveillance Laboratory, Lithuania
Georgetown University	Peruvian Amazon Research Institute
Hokkaido University	Prefectural University of Kumamoto
Institute of Forensic Medicine, Faculty of Medicine, University of Belgrade	Sokoine University of Agriculture
Instituto Adolfo Lutz	Universidad de Panamà
Instituto de Toxicología de la Defensa de España	Universidad Nacional del Litoral
Instituto Polo Tecnológico de Pando	Università degli Studi dell'Insubria
Laboratorio Tecnológico del Uruguay	Universitas Trisakti
Laboratorio UPL - La Comisión de Investigaciones Científicas (CIC), Buenos Aires, Argentina	