HORIZON2020 Programme
Contract No. 733032 HBM4EU

Standard Operation Procedure

HBM4EU-SOP-QA-003

Evaluation of results from Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS)

Version / date of issue	2 / 28-03-2019
Author(s) (Short name of institute)	Hans Mol (RIKILT), Thomas Goen (IPASUM), Marta Esteban (ISCIII)
Approved by:	Thomas Goen (IPASUM), task leader WP9.4 Argelia Castaño (ISCIII), work package leader WP9, Pillar 2 leader
Validation status:	~~Basic / Single lab / Multi-lab~~ / Not applicable
Accredited:	~~Yes~~ / No

Table of contents

1 Aim and application area

This SOP is one out of four SOPs describing how Interlaboratory Comparison Investigations (ICIs) and External Quality Assurance Schemes (EQUAS) are done in HBM4EU:

HBM4EU-SOP-QA-001 "Organisation of Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS) of interlaboratory studies"

HBM4EU-SOP-QA-002 "Preparation of control materials for Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS)"

HBM4EU-SOP-QA-003 "Evaluation of results from Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS)"

HBM4EU-SOP-QA-004 "Reporting of results of Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS)"

This SOP describes the procedures for evaluation of the results submitted by laboratories participating in ICI / EQUAS studies. For drafting this SOP, requirements as outlined in ISO/IEC 17043:2010, "Conformity assessment – General requirements for proficiency testing " have been taken into account.

2 Definitions

All definitions have been compiled in HBM4EU-SOP-QA-001, and the reader is referred to that SOP.

3 Procedures

3.1 Evaluation of ICI results

The results submitted by the participants are transferred into an Excel spreadsheet 'as submitted' (so without any adjustment/corrections for obvious errors such as wrong unit etc.). Input of results from the participants by the organizer into an excel sheet or software tool is double checked. An Excel sheet or a dedicated software tool then calculates for each biomarker in each test material:
- the consensus value
- the uncertainty of the consensus value
- the robust standard deviation of the participants' results
- the Z-score (separately for each laboratory)

3.1.1 Calculation of the consensus value

The consensus value (C) is calculated from the results submitted by the participants. For this robust statistics is used. With this method, outliers are not discarded but have only a minor influence on the performance parameters. The robust mean is taken as consensus value. Robust statistics is performed in accordance with [Thompson 2006], the guidelines from [Analytical Methods Committee, 1989a&b], and ISO 13528. This provides the consensus value (C) and the ICI standard deviation ($\hat{\sigma}$ = standard deviation of the consensus value).

From this the uncertainty of the consensus value will be calculated as follows:

$$u = 1.25\frac{\hat{\sigma}}{\sqrt{n}} \qquad (1)$$

With: u = uncertainty of the consensus value;
$\hat{\sigma}$ = standard deviation of the participants' results (ICI standard deviation);
n = number of results used for calculation of the consensus value.

3.1.2 Target standard deviation

The standard deviation for proficiency (target standard deviation) determines the performance boundaries in an ICI/EQUAS. In the evaluation of ICI/EQUAS results (i.e. calculation of the Z-score), a pre-set fit-for-purpose target standard deviation is used rather than the standard deviation obtained from the participants' results. The fit-for-purpose target standard deviation reflects the maximum variability that is considered acceptable for a certain biomarker/(concentration)/matrix.

The value is set based on expert opinion, taking into account what is technically feasible and realistic in current routine practises. By default, the value is set at:

$$\sigma_T = 0.25C \qquad (2)$$

3.1.3 Calculation of the Z-score

In the calculation of the Z-score, the uncertainty of the consensus value and, if applicable, instability of the control material are taken into account.

First it is tested whether or not the uncertainty in the consensus value can be considered negligible. This is the case when:

$$u \leq 0.3\sigma_T \quad (3)$$

With: u = uncertainty of the consensus value (formula 1);
σ_T = the target standard deviation as set in 3.1.2).

Next, it is tested if the uncertainty of the consensus value is within acceptable limits with respect to use for statistical evaluation of the data and calculation of Z-scores. This is the case when:

$$u \leq 0.7\sigma_T \quad (4)$$

When the condition of formula (4) is not met, then no Z-scores are provided, the data set for the applicable biomarker/matrix/concentration is unfit for evaluating individual laboratory's performance.

Below, the calculation of Z-scores are described for four scenarios:

Uncertainty of consensus value negligible / biomarker in control material stable
If the uncertainty of the consensus value is negligible and there were no issues with the stability of the test material, then the Z-score (Z) is calculated as follows:

$$Z = \frac{x - C}{\sigma_T} \quad (5)$$

With: Z = Z-score for the submitted analysis result;
x = result submitted by the participant;
C = consensus value, determined according to 3.1.1;
σ_T = target standard deviation, set according to 3.1.2.

Uncertainty of consensus value NOT negligible / biomarker in control material stable
When the uncertainty of the consensus value cannot be considered negligible, i.e. the condition from formula (3) is not met, the uncertainty of the consensus value is taken into account and the Z-score (Z') is calculated as follows:

$$Z'_a = \frac{x - C}{\sqrt{\sigma_T^{\,2} + u^2}} \quad (6)$$

With: Z' = Z-score for the submitted analysis result;
x = result submitted by the participant;
C = consensus value, determined according to 3.1.1;
σ_T = target standard deviation, set according to 3.1.2;
u = uncertainty of the consensus value (formula 1).

Uncertainty of consensus value negligible / biomarker in control material NOT stable

When the biomarker turned out to be not stable in the test material (for stability testing see HBM4EU-QA-002), it is still possible to calculate a Z-score, but the instability has to be taken into account, and the Z-score (Z_i) is calculated as follows:

$$Z_i = \frac{x - C}{\sqrt{\sigma_T^2 + \Delta^2}} \qquad (7)$$

With: Z_i = Z-score for the submitted analysis result;
x = result submitted by the participant;
C = consensus value, determined according to 3.1.1;
σ_T = target standard deviation, set according to 3.1.2;
Δ = difference between mean concentrations of the biomarker at t=0 and t=end.

Uncertainty of consensus value NOT negligible / biomarker in control material NOT stable

In case the biomarker turned out to be not stable in the test material (for stability testing see HBM4EU-QA-002) and the uncertainty of the consensus value cannot be neglected (i.e. the condition from formula (3) is not met), then the Z-score (Z'_i) is calculated as follows:

$$Z'_i = \frac{x - C}{\sqrt{\sigma_T^2 + u^2 + \Delta^2}} \qquad (10)$$

With: Z'_i = Z-score for the submitted analysis result;
x = result submitted by the participant;
C = consensus value, determined according to 3.1.1;
σ_T = target standard deviation, set according to 3.1.2;
u = uncertainty of the consensus value (formula 1);
Δ = difference between mean concentrations of the biomarker at t=0 and t=end.

3.1.4 Classification of analysis results submitted by the participants

In accordance with ISO 13528 and ISO 17043 and the deliverable D 9.4 "*The Quality Assurance/Quality Control Scheme in the HBM4EU project*, Z-scores are classified as presented in Table 1.

Table 1: Classification of Z-scores

$\lvert Z\rvert \le 2$	Satisfactory
$2 < \lvert Z\rvert < 3$	Questionable
$\lvert Z\rvert \ge 3$	Unsatisfactory

3.2 Evaluation of EQUAS results

For an EQUAS study, Z-scores are calculated using equation 5, with C = expert value determined as described in HBM4EU-SOP-QA-001, section 3.3.2.

3.3 Dealing with false positives and <LOQ

In case a laboratory reports a concentration for a biomarker that was not present in the control material (i.e. as assessed during preparation of the test material and based on data from the other participants) this will be indicated in the report drafted by the organizer as a false positive. When the reported concentration by the participant is lower than the LOQ of the other participants, additional investigations need to be done by the organizer to judge whether or not the reported biomarker was a false positive.

When a biomarker is reported as "<LOQ-value", AND an assigned value could be established for the biomarker in the control material, a further assessment is done to verify whether this result might be a false negative and to judge whether the LOQ is considered adequate (low enough) for analysis in the frame of HBM4EU. A result is a false negative when the LOQ of a biomarker is well below the assigned value, but the laboratory did not report a quantitative value. The LOQ is considered not adequate (too high) in relation to the assigned value and in comparison to the other laboratories when:

1) the LOQ is substantially above the assigned value
2) the assigned value represents a realistic concentration of real samples in the frame of HBM4EU
3) quantitatively determination is feasible by the majority of the laboratories

In order to judge "<LOQ" results in a quantitative way, 'proxy-Z-scores' are calculated using equation (5) and the LOQ-value as result. When no LOQ is specified, zero will be used. Proxy-Z-scores are indicated in the report as between brackets and are for information. They are not included in the graphical representations of z-scores of the participants. The interpretation is as follows:

proxy-Z ≤ -3	based on the LOQ provided, the laboratory should have been able to detect and quantify the biomarker. The result is classified as a false negative (FN) and is interpreted as 'unsatisfactory' performance.
-3 ≤ proxy-Z < -2	based on the LOQ provided, it is highly likely that the laboratory should have been able to detect and quantify the biomarker. The result is classified as a false negative (FN) and should be interpreted as 'questionable'.
-2 ≤ proxy-Z ≤ 2	-2 to 0: based on the assigned value and the LOQ provided, the result cannot be classified as false negative. 0 to 2: benchmark: the LOQ is in the range of what is analytically feasible*.
2 < proxy-Z < 3	benchmark: the LOQ is high compared to what is analytically feasible* and might be high in relation to HBM4EU analysis. The laboratory should consider to lower their LOQ.
proxy-Z ≥3	benchmark: the LOQ is too high compared to what is analytically feasible* and might be too high in relation to HBM4EU analysis. The laboratory should consider to lower their LOQ.

* the analytical feasibility is derived from the ICI/EQUAS results. When an assigned value can be determined, this means that reliable quantitative determination at a certain low level is feasible.

4 Registration

The calculated parameters and Z-scores are registered in the Excel sheet used (or an export from a dedicated software tool) which will be saved and archived. All other relevant records regarding the evaluation of the data are also archived either electronically (with at least one backup) or as paper files for a period of at least 5 years.

5 References

HBM4EU-SOP-QA-001 "Organisation of Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS) of interlaboratory studies”

HBM4EU-SOP-QA-002 “Preparation of control materials for Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS)”

HBM4EU-SOP-QA-004 “Reporting of results of Interlaboratory Comparison Investigations (ICI) and External Quality Assurance Schemes (EQUAS)”

ISO/IEC 17043:2010, Conformity assessment – General requirements for proficiency testing

ISO 13528, 2015, Statistical methods for use in proficiency testing by interlaboratory comparison.

Thompson M., Ellison R. and Wood, R., 2006, The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories, Pure Appl. Chem, 78(1), 145-196.

Analytical Methods Committee, 1989a, Robust statistics - How not to reject outliers Part 1. Basic concepts, Analyst, 114, 1693-1697.

Analytical Methods Committee, 1989b, Robust statistics - How not to reject outliers Part 2. Interlaboratory trials, Analyst, 114, 1699-1702.

Official Methods of Analysis Program Manual, 2002, Appendix D: Guidelines for Collaborative Study Procedures To Validate Characteristics of a Method of Analysis. Association Of Analytical Communities International. http://www.aoac.org/vmeth/Manual_Part_6.pdf.