Pure & Appl. Chem., Vol. 60, No. 6, pp. 855–864, 1988.
Printed in Great Britain.

INTERNATIONAL UNION OF PURE
AND APPLIED CHEMISTRY

ANALYTICAL, APPLIED AND CLINICAL CHEMISTRY DIVISIONS

INTERDIVISIONAL WORKING PARTY FOR HARMONIZATION OF
QUALITY ASSURANCE SCHEMES FOR ANALYTICAL
LABORATORIES*†

PROTOCOL FOR THE DESIGN, CONDUCT AND INTERPRETATION OF COLLABORATIVE STUDIES

Resulting from the IUPAC Workshop on the Harmonization of Collaborative Analytical Studies, Geneva, Switzerland, 4–5 May 1987

Prepared for publication by

WILLIAM HORWITZ

Center for Food Safety and Applied Nutrition, HFF–7,
Food and Drug Administration, Washington, DC 20204, USA

* Membership of the Working Party during 1985–89:

Chairman: 1985–87 F. Pellerin (France); 1987–89 G. Svehla (UK); *Secretary*: 1985–87 G. Svehla (UK); 1987–89 M. Parkany (Switzerland); *Members*: 1987–89 S. S. Brown (UK); 1985–89 L. E. Coles (UK); 1985–89 G. den Boef (Netherlands); 1985–89 W. Horwitz (USA); 1987–89 S. H. H. Olrichs (Netherlands); 1985–87 M. Parkany (Switzerland); 1987–89 W. D. Pocklington (UK).

†Title 1985–87: Interdivisional Working Party for the Harmonization of Collaborative Analytical Studies

Protocol for the design, conduct and interpretation of collaborative studies

Abstract - The minimum requirements accepted by consensus of 27 participants at an IUPAC Workshop for the design, conduct, and interpretation of collaborative interlaboratory studies are given.

This document deals with a collaborative study (as distinct from a proficiency study or a certification study). The document emphasizes the need for a preliminary optimization of the candidate method within at least a single laboratory to avoid wasting the power and expense of a collaborative study on a relatively untested method.

The recommended protocol developed for a collaborative study of the optimized method indicates the required numbers of materials, laboratories, and replicate analyses arranged according to specific designs. The protocol also specifies the method for removal of outliers, designates the statistical parameters to be calculated from the valid data, and suggests a format for reporting the results of the statistical analysis.

INTRODUCTION

This document summarizes the minimum requirements for a collaborative study based upon the recommendations accepted by consensus of the 27 participants at the IUPAC Workshop on the Harmonization of Collaborative Analytical Studies, held in Geneva, Switzerland, 4- 5 May 1987. If a collaborative study is to be indicated as complying with the "IUPAC-1987 Protocol," it must be in conformity with the minimum rules given below. Additional requirements may be imposed by other organizations for their specific needs.

These harmonized requirements are the result of efforts begun by the late Dr. Harold Egan, Laboratory of the Government Chemist, United Kingdom, who organized a meeting of interested international organizations in London, England, in March 1978. This was followed by Symposia in Helsinki, Finland, in 1981, and in Washington, DC, USA, in 1984, and the Workshop in Geneva.

The following organizations have participated in one or more of these meetings (* indicates attendance at the Workshop):

*American Oil Chemists' Society
American Society for Testing and Materials
*Association of Official Analytical Chemists
*Association Francaise de Normalisation
*British Pharmacopoeia Commission
*British Standards Institution
CEC Community Bureau of Reference
*Collaborative International Pesticides Analytical Council
European Brewery Convention
European Council of Federations of Industrial Chemistry
European Monitoring and Evaluation Programme
*Food and Agricultural Organization/Codex Alimentarius Commission
International Agency for Research on Cancer
International Atomic Energy Agency
*International Association for Cereal Science and Technology
*International Commission for Uniform Methods of Sugar Analysis
International Committee for Standardization in Haematology
*International Dairy Federation
International Federation of Clinical Chemistry
*International Office of Cocoa and Chocolate
International Olive Oil Council
*International Organization for Standardization (Central Secretariat and Technical Committees 34 and 102 and their Subcommittees)

International Sugar Confectionery Manufacturers' Association
International Office of Wine
*International Union of Pure and Applied Chemistry (Analytical Chemistry, Clinical Chemistry, and Applied Chemistry Divisions)
*Nordic Committee on Food Analysis
*Swiss Commission of Food Analysis
(United Kingdom) Assay Offices
(United Kingdom) Laboratory of the Government Chemist
*(United Kingdom) Ministry of Agriculture, Fisheries and Food
(United Kingdom) Royal Society of Chemistry (Analytical Methods Committee)
*(United States) Food and Drug Administration
(United States) National Bureau of Standards
United States Pharmacopeia
*Water Research Centre (United Kingdom)
World Health Organization

PROTOCOL

1.0 Preliminary work

Collaborative studies require considerable effort and should be conducted only on methods that have received adequate prior testing. Such within-laboratory testing should include, as applicable, information on the following:

1.0.1 Preliminary estimates of precision

Estimates of the total within-laboratory standard deviation of the analytical results over the concentration range of interest; as a minimum at the upper and lower limits of the concentration range, with particular emphasis on any standard or specification value.

NOTE 1: The total within-laboratory standard deviation is a more inclusive measure of imprecision than the ISO repeatability standard deviation, §3.3 below. This parameter is the maximum within-laboratory standard deviation to be expected from the performance of a method, at least on different days and preferably with different calibration curves. It includes between-run (between-batch) as well as within-run (within-batch) variations. In this respect it can be considered as a measure of within-laboratory reproducibility. Unless this value is well within acceptable limits, it cannot be expected that the between-laboratory standard deviation (reproducibility standard deviation) will be any better. This precision term is not estimated from the minimum collaborative study described in this protocol.

NOTE 2: The total within-laboratory standard deviation may also be estimated from ruggedness trials that indicate how tightly controlled the experimental factors must be and what their permissible ranges are. These experimentally determined limits should be incorporated into the description of the method.

1.0.2 Systematic error (bias)

Estimates of systematic error of the analytical results over the concentration range and in the commodities of interest; as a minimum at the upper and lower limits of the concentration range, with particular emphasis on any standard or specification value. The results obtained by applying the method to relevant reference materials should be noted.

1.0.3 Recoveries

The recovery of "spikes" added to real materials and to extracts, digests, or other treated solutions thereof.

1.0.4 Applicability

The ability of the method to identify and measure the physical and chemical forms of the analyte likely to be present in the materials.

1.0.5 Interference

The effect of other substances that are likely to be present at appreciable concentrations in matrices of interest and which may interfere in the determination.

1.0.6 Method comparison
The results of comparison of the application of the method with existing tested methods intended for similar purposes.

1.0.7 Calibration procedures
The procedures specified for calibration and for blank correction must not introduce important bias into the results.

1.0.8 Method description
The method must be clearly and unambiguously written.

1.1 Significant figures
The initiating laboratory should indicate the number of significant figures to be reported, based on the output of the measuring instrument.

NOTE: In making statistical calculations from the reported data, the full power of the calculator or computer is to be used with no rounding or truncating until the final reported mean and standard deviations are achieved. At this point the standard deviations are rounded to 2 significant figures and the mean and relative standard deviations are rounded to accommodate the significant figures of the standard deviation. For example, if $s_R = 0.012$, $\bar{x}$ is reported as 0.147, not as 0.1473 or 0.15, and RSD_R is reported as 8.2%. (Symbols are defined in Appendix 1.) If standard deviation calculations must be conducted manually in steps, with the transfer of intermediate results, the number of significant figures to be retained for squared numbers should be at least 1 plus 2 times the number of figures in the data.

2.0 Design of the collaborative study

2.1 Number of materials
(Material = analyte/concentration level/matrix combination)

At least 5 materials must be used; only when a single-level specification is involved for a single matrix may this minimum required number of materials be as low as 3. For this design parameter, the 2 portions of a split level (4 of a double split level) and the k individual portions of blind replicates per laboratory are considered as a single material.

2.2 Number of laboratories
At least 8 laboratories must report results for each material; only when it is impossible to obtain this number (e.g., very expensive instrumentation or specialized laboratories required) may the study be conducted with less, but with an absolute minimum of 5 laboratories. If the study is intended for international use, laboratories from different countries should participate.

2.3 Number of replicates
The repeatability precision parameters must be estimated by using one of the following set of designs (listed in approximate order of desirability):

2.3.1 Split level (single or double)
For each level which is split and which constitutes only a single material for purposes of design and statistical analysis, use 2 nearly identical materials that differ only slightly in analyte concentration. For the single split level, each laboratory is to make only 1 determination on each (split) level (total = 2 per material); for the double split level, 2 known (nonblind) determinations are made on each (split) level (total = 4 per material). Alternatively, for the double split level, the 2 replicates for each (split) level may be submitted as blind replicates (1 determination on each portion submitted (total = 4 per material)).

2.3.2 Combination blind replicates and split level
Use split levels for some materials and blind replicates for other materials in the same study (single values from each submitted portion).

2.3.3 Blind replicates
For each material, use blind identical replicates; when data censoring is impossible (e.g., automatic input, calculation, and printout), nonblind identical replicates may be used.

2.3.4 Known replicates
For each material, use known replicates (2 or more analyses of portions from the same test sample), but only when it is not practical to use one of the preceding designs.

2.3.5 Independent replicate analyses
Use only a single portion from each material (i.e., do not replicate) in the collaborative study, but rectify the inability to calculate repeatability parameters by quality control parameters and other within-laboratory data obtained independently of the collaborative study.

3.0 Statistical analysis (See attached flowchart)
In the statistical analysis of the collaborative study data, the required statistical procedures listed below must be performed and the results reported. Supplemental, additional procedures are not precluded.

3.1 Valid data
Only valid data should be reported and subjected to statistical treatment. Valid data are those data that would be reported as resulting from the normal performance of laboratory analyses; they are not marred by method deviations, instrument malfunctions, unexpected occurrences during performance, or by clerical or typographical errors.

3.2 One-way analysis of variance
One-way analysis of variance and outlier treatments must be applied separately to each material to estimate the components of variance and repeatability and reproducibility parameters.

3.3 Initial estimation
Calculate the mean, $\bar{x}$ (= average of laboratory averages), repeatability relative standard deviation, RSD_r, and reproducibility relative standard deviation, RSD_R, with no outliers removed, but using only valid data.

3.4 Outlier treatment
The estimated precision parameters that must also be reported are based on the initial valid data purged of all outliers flagged by the harmonized 1987 outlier procedure. This procedure essentially consists of sequential application of the Cochran and Grubbs tests (at 1% probability (P) level, 1-tail for Cochran, 2-tail for single Grubbs, overall for paired Grubbs) until no further outliers are flagged or until a drop of more than 22.2% (= 2/9) in the original number of laboratories would occur.

NOTE: The Grubbs tests are to be applied one material at a time to the set of replicate means from all laboratories, and not to individual values from replicated designs, because their differences from the overall mean for that material are not independent.

3.4.1 Cochran test
First apply the Cochran outlier test (1-tail test at P = 1%) and remove any laboratory whose critical value exceeds the tabular value given in the table, Appendix A.3.1.

3.4.2 Grubbs tests
Apply the single-value Grubbs test (2-tail) and remove any outlying laboratory; if no laboratory is flagged, then apply the pair-value test (two values at the same end and one value at each end, P = 1% overall). Remove any laboratory(ies) flagged by these tests, using the table, Appendix A.3.1, but stop removal if more than 22.2% (2 of 9 laboratories) would be removed.

3.4.3 Final estimation
Recalculate the parameters as in §3.3 after the laboratories flagged by the preceding procedure have been removed. If no outliers were removed in the Cochran-Grubbs sequence, terminate testing. Otherwise, reapply the Cochran-Grubbs sequence to the data purged of the flagged outliers until no further outliers are flagged or until more than a total of 22.2% (2 of 9 laboratories) would be removed. See flowchart, A.3.4.

4.0 Final report
The final report should be published and should include all valid data. Other information and parameters should be reported in a format similar (with respect to the reported items) to the following (as applicable):

[x] collaborative tests carried out at the international level in [year(s)] by [organization] in which [y and z] laboratories participated, each performing [k] replicates, gave the following statistical results:

TABLE OF COLLABORATIVE STUDY PARAMETERS
Analyte; Results expressed in [units]
Material [Description and listed across the top in increasing order of magnitude of means]

Number of laboratories retained after eliminating outliers
Number of outliers
Mean
True or accepted value, if known
Repeatability standard deviation (s_r)
Repeatability relative standard deviation (RSD_r)
Repeatability value, r ($2.8 \times s_r$)

Reproducibility standard deviation (s_R)
Reproducibility relative standard deviation (RSD_R)
Reproducibility value, R ($2.8 \times s_R$)

4.1 Symbols
A set of symbols for use in collaborative study reports and publications is attached as Appendix 1 (A.1).

4.2 Definitions
A set of definitions for use with collaborative study reports and publications is attached as Appendix 2 (A.2).

4.3 Miscellaneous

4.3.1 Recovery
Recovery of added analyte as a control on method or laboratory bias should be calculated as follows:

[Marginal] Recovery, % =
(Total analyte found - analyte originally present) x 100/(analyte added)

Although the analyte may be expressed as either concentration or amount, the units must be the same throughout. When the amount of analyte is determined by analysis, it must be determined in the same way throughout.

Analytical results should be reported uncorrected for recovery. Report recoveries separately.

4.3.2 When s_L is negative

By definition, s_R is greater than or equal to s_r in collaborative studies; occasionally the estimate of s_r is greater than the estimate of s_R (the range of replicates is greater than the range of laboratory averages and the calculated s_L^2 is then negative). When this occurs, set $s_L = 0$ and $s_R = s_r$.

APPENDIX 1

A.1

Use the following set of symbols and terms for designating parameters developed by a collaborative study.

Mean (of laboratory averages)	$\bar{x}$	
Standard deviations:	s	(estimates)
Repeatability	s_r	
"Pure" between-laboratory	s_L	
Reproducibility	s_R	
Variances:	s^2	(with subscripts r, L, and R)
Relative standard deviations:	RSD	(with subscripts r, L. and R)
Maximum tolerable differences (as defined by ISO 5725:1986; see A.2.4 and A.2.5)		
Repeatability value	$r = (2.8 \times s_r)$	
Reproducibility value	$R = (2.8 \times s_R)$	
Number of replicates per laboratory	k	(general)
Average number of replicates per laboratory	$\bar{k}$	(for a balanced design)
Number of replicates for laboratory i	k_i	
Number of laboratories	L	
Number of materials	m	
Total number of values in a given assay	n	(= kL for a balanced design)
Total number of values in a given study	N	(= kLm for an overall balanced design)

If other symbols are used, their relationship to the recommended symbols should be explained fully.

A.2 **APPENDIX 2**

Use the following definitions:

A.2.1 Collaborative study
A collaborative study is an interlaboratory study in which each laboratory uses the defined method of analysis to analyze identical portions of homogeneous materials to assess the performance characteristics obtained for that method of analysis.

A.2.2 Proficiency study
A proficiency study is an interlaboratory study consisting of one or more assays conducted by a group of laboratories on one or more identical materials, by whatever method is in use in each laboratory, for the purpose of comparing the results of each laboratory with those of other laboratories, with the objective of evaluating or improving performance.

A.2.3 Certification study
A certification study is an interlaboratory study in which a group of selected laboratories analyze a candidate reference material by methods judged most likely to provide the least biased estimates of concentration (or of a property) and the smallest associated uncertainty, for the purpose of providing a reference value of the analyte concentration (or property) in the material.

A.2.4 Repeatability value (r)
When the mean of the values obtained from 2 single determinations, performed [simultaneously or] in rapid succession by the same operator, using the same apparatus under the same conditions for the analysis of the same test sample, lies within the range of the mean values cited in the Table, 4.0, the difference between the 2 values obtained should not be greater than the repeatability value (r) that can generally be inferred by linear interpolation from the Table.

A.2.5 Reproducibility value (R)
When the values for the final result, obtained by operators in different laboratories using different apparatus under different conditions for the analysis of the same laboratory sample, lie within the range of the mean values cited in the Table, the difference between the values for the final result obtained by those operators should not be greater than the reproducibility value (R) that can generally be inferred by linear interpolation from the Table.

NOTE 1: When the results of the interlaboratory test make it possible, the value of (r) or (R) can be indicated as a relative value (e.g., as a percentage of the determined mean value), or as an absolute value.

NOTE 2: When the final reported result is an average derived from more than a single value, i.e., k is greater than 1, the value for R must be adjusted according to the following formula:

$$R' = [R^2 + r^2(1 - [1/k])]^{\frac{1}{2}}$$

Similar adjustments must be made for replicates constituting the final values for s_R and RSD_R, if these will be the reported parameters used for quality control purposes.

NOTE 3: The repeatability value (r) may be interpreted as the amount by which 2 determinations should agree with each other within a laboratory 95% of the time. The reproducibility value (R) may be interpreted as the amount by which 2 separate determinations conducted in different laboratories should agree with each other 95% of the time.

A.2.6 One-way analysis of variance
One-way analysis of variance is the statistical procedure for obtaining the estimates of within-laboratory and between-laboratory variability on a material-by-material basis. Examples of the calculations for the single-level and single-split-level designs can be found in ISO 5725-1986. The calculations for the double split level can be found in Netherlands Standardization Organization Standard NEN 6303.

A.3 **APPENDIX 3**

A.3.1 Critical values for the Cochran maximum variance test, 1-tail, at the P = 1% level, expressed as a critical variance ratio; and critical values for the Grubbs tests, at the P = 1% level, expressed as the percent reduction in standard deviation caused by removal of the suspect value(s)

L = Number of laboratories for the given material

	Cochran test: Critical variance ratio					Grubbs tests	
	Number of replicates from each laboratory (k)					Single	Pair
L	2	3	4	5	6	value	value
2	---	0.995	0.979	0.959	0.937	---	---
3	0.993	0.942	0.883	0.834	0.793	99.3	---
4	0.968	0.864	0.781	0.721	0.676	91.3	99.7
5	0.928	0.788	0.696	0.633	0.588	80.7	95.4
6	0.883	0.722	0.626	0.564	0.520	71.3	88.3
7	0.838	0.664	0.568	0.508	0.466	63.6	81.4
8	0.794	0.615	0.521	0.463	0.423	57.4	75.0
9	0.754	0.573	0.481	0.425	0.387	52.3	69.4
10	0.718	0.536	0.447	0.393	0.357	48.1	64.6
11	0.684	0.504	0.418	0.366	0.332	44.5	60.5
12	0.653	0.475	0.392	0.343	0.310	41.5	56.8
13	0.624	0.450	0.369	0.322	0.291	38.9	53.6
14	0.599	0.427	0.349	0.304	0.274	36.6	50.8
15	0.575	0.407	0.332	0.288	0.259	34.6	48.3
16	0.553	0.388	0.316	0.274	0.246	32.8	46.0
17	0.532	0.372	0.301	0.261	0.234	31.2	44.0
18	0.514	0.356	0.288	0.249	0.223	29.8	42.1
19	0.496	0.343	0.276	0.238	0.214	28.5	40.4
20	0.480	0.330	0.265	0.229	0.205	27.3	38.9
21	0.465	0.318	0.255	0.220	0.197	26.2	37.4
22	0.450	0.307	0.246	0.212	0.189	25.2	36.1
23	0.437	0.297	0.238	0.204	0.182	24.3	34.9
24	0.425	0.287	0.230	0.197	0.176	23.4	33.7
25	0.413	0.278	0.222	0.190	0.170	22.7	32.7
26	0.402	0.270	0.215	0.184	0.164	21.9	31.7
27	0.391	0.262	0.209	0.179	0.159	21.2	30.8
28	0.382	0.255	0.202	0.173	0.154	20.6	29.9
29	0.372	0.248	0.196	0.168	0.150	20.0	29.1
30	0.363	0.241	0.191	0.164	0.145	19.5	28.3
35	0.325	0.213	0.168	0.144	0.127	17.1	25.0
40	0.294	0.192	0.151	0.128	0.114	15.3	22.5

Although the table is strictly applicable only to a balanced design (same number of replicates from all laboratories), it can be applied to an unbalanced design without too much error, if there are only a few deviations.

Source: Cochran table abbreviated from ISO 5725. Grubbs table: Dr. Patrick Kelly, Canada Packers, Toronto, Canada. Single critical values were calculated from available formulas; pair critical values were obtained by simulation and fitting and should be accurate to 0.2% absolute. (To be submitted for publication to Technometrics.)

A.3.2 Calculation of Cochran outlier test value

Compute the within-laboratory variance for each laboratory and divide the largest of these by the sum of all of the variances. The resulting quotient is the Cochran statistic which indicates the presence of a removable outlier if this quotient exceeds the critical value listed above in the Cochran table for the number of replicates and laboratories specified.

A.3.3 Calculation of the Grubbs test values

To calculate the single Grubbs test statistic, compute the average for each laboratory and then calculate the standard deviation (SD) of these L averages (designate as the original s). Calculate the SD of the set of averages with the highest average removed (s_H); calculate the SD of the set of averages with the lowest average removed (s_L). Then calculate the percentage decrease in SD as follows:

$$100 \times [1 - (s_L/s)] \quad \text{and} \quad 100 \times [1 - (s_H/s)]$$

The higher of these 2 percentage decreases is the single Grubbs statistic, which signals the presence of an outlier to be omitted at the P = 0.01 level, 2-tail, if it exceeds the critical value listed in the single-value column, for L laboratories, Appendix A.3.1.

To calculate the Grubbs pair value test statistic, proceed in an analogous fashion, except calculate the standard deviations s_{2L}, s_{2H}, and s_{HL}, following removal of the 2 lowest, the 2 highest, and the highest and the lowest averages, respectively, from the original set of averages. Take the smallest of these 3 SD values and calculate the corresponding percentage decrease in SD from the original s. A Grubbs outlier pair is present if the selected value for the percentage decrease from the original s exceeds the critical value listed in the Grubbs pair value column, Appendix A.3.1.

A.4 APPENDIX 4

A.4.1 Flowchart for outlier removal