

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

ANALYTICAL CHEMISTRY DIVISION
COMMISSION ON ANALYTICAL NOMENCLATURE*
COMMISSION ON GENERAL ASPECTS OF ANALYTICAL CHEMISTRY†

NOMENCLATURE OF INTERLABORATORY ANALYTICAL STUDIES

(IUPAC Recommendations 1994)

Prepared for publication by

WILLIAM HORWITZ

Center for Food Safety and Applied Nutrition, HFS-500, Food & Drug Administration, Washington,
DC 20204, USA

*Membership of the Commission during the period (1987–1989) when this report was being prepared is given hereunder (*Note: The Commission ceased to exist after the 35th IUPAC General Assembly, Lund, 1989*).

Chairman: R. E. Van Grieken (Belgium); *Secretary:* C. L. Graham (UK); *Titular Members:* C. A. M. G. Cramers (Netherlands), L. A. Currie (USA), W. Horwitz (USA), D. Klockow (FRG), M. Parkany (Switzerland); *Associate Members:* L. S. Etre (USA), D. M. Everaerts (Netherlands), Y. Gohshi (Japan), P. S. Goel (India), H. M. Kingston (USA), G. J. Patriarche (Belgium), D. L. Rabenstein (USA), J. W. Stahl (USA); *National Representatives:* Magda Ariel (Israel), K. Danzer (GDR), U. L. Haldna (USSR), D. R. Reeves (New Zealand), B. Schreiber (Switzerland), J. Stary (Czechoslovakia), C. Svehla (UK).

†Membership of the Commission during the preparation of this report (1989–93) was as follows:

Chairman: 1989–91 F. Ingman (Sweden); 1991–93 W. E. van der Linden (Netherlands); *Vice-Chairman:* 1989–91 R. E. van Grieken (Belgium); *Secretary:* 1989–91 W. E. van der Linden (Netherlands); *Co-Secretary:* 1989–91 C. L. Graham (UK); *Secretary:* 1991–93 C. L. Graham (UK); *Titular Members:* L. A. Currie (USA); 1989–93; St. Głab (Poland; 1989–93); W. Horwitz (USA; 1989–93); D. L. Massart (Belgium; 1989–93); M. Parkany (Switzerland; 1989–93); *Associate Members:* K. Danzer (FRG; 1991–93); P. S. Goel (India; 1989–91); Y. Gohshi (Japan; 1989–93); H. Müller (FRG; 1989–93); M. Otto (FRG; 1989–93); G. J. Patriarche (Belgium; 1989–91); S. V. Savvin (Russia; 1989–91); J. W. Stahl (USA; 1989–93); P. J. Worsfold (UK; 1989–93); *National Representatives:* Tania M. Tavares (Brazil; 1989–93); L. Sommer (Czechoslovakia; 1989–91); J. Garaj (Czechoslovakia; 1991–93); D. Klockow (FRG; 1989–91); K. Danzer (GDR; 1989–91); J. Inczédy (Hungary; 1989–93); D. Thorburn Burns (Ireland; 1989–93); R. D. Reeves (New Zealand; 1989–93); A. Hulanicki (Poland; 1989–91); J. F. van Staden (RSA; 1991–93); F. Ingman (Sweden; 1991–93); B. Schreiber (Switzerland; 1989–91); S. Ateş (Turkey; 1989–93); G. Svehla (UK; 1989–91).

Names of countries given after Members' names are in accordance with the *IUPAC Handbook 1991–93*; changes will be effected in the 1994–95 edition.

Rapid communication of comments on this document can be made to Dr. William Horwitz by telephone (+1-202-205-4046/4346) or by FAX (+1-202-401-7740)

Republication of this report is permitted without the need for formal IUPAC permission on condition that an acknowledgement, with full reference together with IUPAC copyright symbol (© 1994 IUPAC), is printed. Publication of a translation into another language is subject to the additional condition of prior approval from the relevant IUPAC National Adhering Organization.

Nomenclature of interlaboratory analytical studies (IUPAC Recommendations 1994)

Abstract

There are three major types of interlaboratory studies in which a group of laboratories analyzes identical test portions from a homogeneous, stable test sample. Each type is characterized by a self-defining designation according to its purpose: (1) *Method-performance study*.--A study in which all laboratories follow the same written protocol and use the same test method to measure a property (usually concentration of an analyte) in order to assess the performance parameters of a method. (2) *Laboratory-performance study*.--A study in which laboratories use the method of their choice to measure a property in order to assess the performance of the laboratory or analyst, usually to evaluate or improve performance. (3) *Material-certification study*.--A study that assigns a reference value to a characteristic in the test material, usually with a stated uncertainty, using the "best" laboratories and the least-biased methods. Vague terms such as "round-robins," "intercalibrations," "ring tests," etc., should not be used.

INTRODUCTION

By definition, an interlaboratory study in analytical chemistry requires the active participation of more than one laboratory, but a minimum of five laboratories should be used to provide meaningful statistical conclusions. Current IUPAC guidelines call for eight laboratories. Interlaboratory studies differ greatly in design and interpretation, depending on the purposes of the studies. Failure to recognize that interlaboratory studies can have different purposes results in confusion in the designation of the type of study and in the interpretation of the acceptability of the results.

"Interlaboratory study" is the general term for any study requiring the active participation of more than one laboratory to obtain the desired information. Different types of studies are designed to evaluate or assess the important factors affecting the reliability of analytical results. Studies involving several laboratories with each preparing its own test materials and using its own methods, sometimes called *cooperative studies*, or those involving several laboratories merely because of the volume of work, are not included in this description, unless a deliberate attempt is made to coordinate their analytical output through quality control and quality assurance techniques. Interlaboratory studies of methods or materials requiring specialized expertise, techniques, or instrumentation may encompass all the existing laboratories that can conduct a specialized type of analysis. Statistical analysis of such specialized studies require the use of population rather than sampling statistics, where n , the number of laboratories, rather than $(n - 1)$, is used to calculate the standard deviation. The measures of reliability calculated from the initial, complete population studies may not necessarily apply to new laboratories entering the previously restricted analytical field.

Interlaboratory studies have been called "round-robins," "ring tests," "ring trials," "intercalibrations," "external schemes," or similar designations. Use of such terms, even with definitions, is discouraged. Synonyms have been used for "studies:" e.g., "exercises," "checks," "tests," "trials," and "evaluations." Although there is nothing fundamentally wrong with these terms, they all have a narrower scope in some respect than "study." "Exercise" includes the concept of regularity, which is not intended. "Check" and "evaluation" imply an already accepted method, as is often the case with some organizations, but not with others. "Trial," which is widely used in Europe, might be taken to have legal overtones. And "test" is being recommended as the word to replace "sample" in certain terms related to analytical work, as in "test solution." "Study" is general and neutral in connotation. "Evaluation" and "assessment" has also been used for "performance," but perhaps it overemphasizes the decision-requirement aspect. However, when a decision has been made as a result of an interlaboratory study, e.g., acceptance of a method, accreditation of a laboratory, or certification of a material, a specific term such as "validation," "evaluation," or "certification" may be substituted for "performance."

The evaluation types of interlaboratory studies can be characterized by parallel, self-defining terms, depending on which of the three primary factors -- methods, laboratories, materials -- is held constant. This set of internally consistent terms is recommended as a replacement for the conglomeration of unrelated designations currently in use. The proposed terms are:

- A *method-performance study*, which assesses the performance characteristics of a method of analysis;
- A *laboratory-performance study*, which assesses the performance of a laboratory(ies) (or analyst(s));
- A *material-certification study*, which assigns a most probable value of a quantity (e.g., concentration of a component) for a material with a stated uncertainty.

Each type of study requires its own statistical assumptions (model), statistical analysis, and outlier treatment. All such studies require that the laboratories be supplied with one or more homogeneous, stable materials to avoid confounding the results with sampling errors. It is usually undesirable to conduct a study to handle more than one of these objectives, but attaining more than one of these objectives can be an unexpected by-product. For example, if all the laboratories of a laboratory-performance study indicate that they used the same method, the results may also be interpreted as a method-performance study; if sufficient material remains from a method- or laboratory-performance study, the remaining test material with an assigned consensus value may be used for quality control.

OPERATIONAL DEFINITIONS

(Although absolute terms such as "homogeneous" and "identical" are used in the definitions, these terms should be understood to mean that the variability in the property they describe is negligible compared to the variability of the primary factor.)

1. Interlaboratory Study

A study in which several laboratories measure a quantity in one or more identical portions of homogeneous, stable materials under documented conditions, the results of which are compiled into a single report.

NOTE: The larger the number of participating laboratories, the greater the confidence that can be placed in the resulting statistical parameters. The IUPAC-1987 protocol (1) requires a minimum of eight laboratories for method-performance studies.

2. Method-performance Study

An interlaboratory study in which all laboratories follow the same written protocol and use the same test method to measure a quantity in sets of identical test samples. The reported results are used to estimate the performance characteristics of the method. Usually these characteristics are within-laboratory and among-laboratories precision, and when necessary and possible, other pertinent characteristics such as systematic error, recovery, internal quality control parameters, sensitivity, limit of determination, and applicability.

NOTES: (1) The materials used in such a study of analytical quantities are usually representative of materials to be analyzed in actual practice with respect to matrices, amount of test component (concentration), and interfering components and effects. Usually the analyst is not aware of the actual composition of the test samples but is aware of the matrix.

(2) The number of laboratories, number of test samples, number of determinations, and other details of the study are specified in the study protocol. Part of the study protocol is the procedure which provides the written directions for performing the analysis.

(3) The main distinguishing feature of this type of study is the necessity to follow the same written protocol and test method exactly.

(4) Several methods may be compared using the same test materials. If all laboratories use the same set of directions for each method and if the statistical analysis is conducted separately for each method, the study is a set of method-performance studies. Such a study may also be designated as a method-comparison study.

3. Laboratory-performance Study

An interlaboratory study that consists of one or more analyses or measurements by a group of laboratories on one or more homogeneous, stable test samples by the method selected or used by each laboratory. The reported results are compared with those from other laboratories or with the known or assigned reference value, usually with the objective of evaluating or improving laboratory performance.

NOTES: (1) Laboratory-performance studies may be used to accredit laboratories or to audit performance. If a study is conducted by an organization with some type of management control over the participating laboratories -- organizational, accreditation, regulatory, or contractual -- the method may be specified or the selection may be limited to a list of approved or equivalent methods. In such situations, a single test sample is insufficient to judge performance. It is expected that the results from 1 of 20 tests will be outside the limits of the specified performance mean \pm 2 standard deviations due just to chance fluctuations alone.

(2) Sometimes a laboratory-performance study may be used to select a method of analysis that will be used in a method-performance study. If all laboratories, or a sufficiently large subgroup of laboratories, use the same method, the study may also be interpreted as a method-performance study.

(3) Separate laboratories of a single organization with independent facilities, and with different local management, instruments, and calibration materials, are treated as different laboratories.

4. Material-certification Study

An interlaboratory study that assigns a reference value ("true value") to a quantity (concentration or property) in the test material, usually with a stated uncertainty.

NOTE: A material certification study often utilizes selected reference laboratories to analyze a candidate reference material by a method(s) judged most likely to provide the least-biased estimates of concentration (or of a characteristic property) and the smallest associated uncertainty.

5. Test Sample/Analytical Sample/Test Material

The homogeneous, stable material with a certain property, a specified composition, or containing one or more test components at given concentrations in a defined matrix that is subdivided into identical portions sent to the laboratories participating in an interlaboratory study.

NOTES (1) The homogeneous, stable parent substance prepared by the organizer of the interlaboratory study may be designated as the "material" and the subdivided portions may be designated as the "test sample," if it is necessary to distinguish between them. The portion removed from the test sample is the "test portion." If only analytical chemistry is involved, the term "analytical" may be substituted for "test."

(2) Although it is preferable that the composition of the test component be known by formulation, independent analysis, or by assignment so that recovery (or bias, trueness, systematic error) may also be measured, this is not always possible at trace levels, or with natural products, tissues, sediments, sludges, or environmental matrices such as waters and wastewaters, or with analytes that are defined by the method, e.g., moisture, boiling range, etc.

6. Matrix

The carrier of the test component (analyte); all of the constituents of the material except the analyte; or the material with as low a concentration of the analyte as is possible to obtain.

NOTES: Some interlaboratory studies will include the submission of a blank matrix. The instructions should then specify how to report low-level values. The preferred reporting procedure with instrumental methods applied to blank materials is to translate the signal into the corresponding concentration through the calibration graph, extrapolating if necessary. (It should be noted that in most other situations it is not good analytical practice to extrapolate beyond the points used to establish a calibration graph.) The resulting apparent

concentration should be reported as it appears -- positive, negative, or zero. Qualitative terms such as "less than," "below the detection limit of ___ units," assigned values of zero, or a multiple of the detection limit, e.g., 0.5, $\sqrt{2}$ etc. x the detection limit, should not be accepted. Such results cannot be handled by statistical techniques. The omission of such results from the responsible laboratory would diminish the reliability of the estimates of the performance parameters.

7. Laboratory

The place with physical facilities and an environment in which the analyst(s) operates to gather data for the interlaboratory study.

NOTES: (1) Although analytical results are ascribed to a laboratory, they usually reflect the output of an analyst or a team of analysts. When the proficiency of individual analysts within a laboratory is to be evaluated, provision for independent operation by each of the analysts must be specified in the study protocol.

(2) Separate laboratories of a single organization with independent facilities, and with different local management, instruments, and calibration materials, are treated as different laboratories.

8. Data Set

The group of estimated values of quantity or concentration, final results, or decisions (yes/no; accept/reject; present/absent) from the group of participating laboratories in an interlaboratory study of a specific material, at a specific level.

NOTE: The term "assay" has also been applied to this concept -- the data set from a given matrix/analyte level/method combination. The use of "assay" should be confined to the operation of analyzing the material.

9. Determination

The complete analytical (test) operation starting from the removal of the single test portion to reporting the final result.

NOTES: (1) The purpose of this definition is to provide a measure of the amount of work required to conduct the interlaboratory study.

(2) A determination (an operation) must be distinguished from a final result (a datum or estimate). Sometimes the average (mean) of the results from several replicate determinations or independent determinations is the final reported result.

10. Replicate

Each of the set of multiple determinations conducted on identical test portions from one test sample, by one laboratory by the same method and protocol.

NOTES: (1) To avoid ambiguity, the term "replicate" or "replication" used alone should be employed only in the context of measurement (analysis) and not in the sense of "preparation of multiple units" or "collect replicates" unless the usage is explicit, e.g., "Prepare replicate test samples from the laboratory sample."

(2) The analyst may or may not be aware that a test sample is a replicated material. If the determinations are conducted concurrently (regardless of knowledge of the identity of the test samples), the reported results can provide a measure of within-run (-batch, -group) precision, usually designated as repeatability precision. This type of measurement replication often provides an over-optimistic estimate of within-laboratory variability. Consequently, if the identity of the replicate test samples is disclosed, it is better to request the replicate analyses to be conducted at different times in order to obtain a more realistic estimate of within-laboratory variability. If the replication is conducted at different times, the reported values can provide an estimate of between-runs precision, which includes within-run precision. The between-runs precision parameter is intermediate between the ISO repeatability and reproducibility (among-laboratories) precision (2). Closely matched pairs (Youden pairs, split levels) are a permissible substitute for blind replication, but require a different statistical analysis technique.

(3) Replication from the very beginning of the removal of the test portion provides an estimate of repeatability. Repetition beginning at any later stage (e.g., aliquots from the same dissolved test portion) does not provide an estimate of repeatability since the variability introduced by the omitted steps is not included in the final measurement. Presenting a test solution repeatedly to an instrument provides an estimate of instrumental precision only. Instructions and reports should be very clear as to which readings are to be reported separately and which are to be combined.

(4) The final result from each of the series of measurements may be called a "replicate value," indicating exactly which item or step was replicated.

11. Protocol

The detailed set of instructions describing the design, conduct, and reporting of a study, test, or trial.

NOTES: (1) The protocol of an interlaboratory study should specify the minimum number of laboratories, the number and nature of the test samples, the details of the method(s) of analysis, and the number of replicate determinations to be performed, as applicable. It should also contain practical details of transport, receipt, and preservation of test samples, the performance of the statistical analysis (particularly the outlier removal techniques), and the reporting of final results. See Reference (1) for the IUPAC-1987 Harmonized Protocol for a method-performance study.

(2) To avoid misunderstandings, the protocol must be completely specific about which readings are to be reported separately and which are to be combined (averaged), and whether reported replicate results apply to the complete determination or only to a portion of it. The statistical analysis must be consistent with the instructions to the participating laboratories. See also Item 10, NOTE 3.

(3) Participants should not provide more or fewer data than requested, and should report the same number of significant figures. These requirements simplify the subsequent statistical analysis.

COMMENTARY

A method-performance study focuses on the method, not the analyst. It requires the use of the same set of directions for the conduct of the analysis by all participants. The participants may be an "open set," permitting the participation of any laboratory with the necessary skill and facilities to use the method in practice. The study may also utilize a "closed set" of participants who are recognized experts in the specialized type of analysis, who may operate under professional or legal accreditation or under voluntary or imposed quality assurance programs. The report from a method-performance study usually does not provide information on the skill and experience of the participants. The usual statistical analysis makes the assumption that all laboratories perform with equal variability. Although this assumption is necessary for logical and statistical reasons, it is the confidence interval around the reported average that provides the range within which to expect future results from typical similar laboratories to lie, with a specified degree of confidence (usually 95 or 99%).

Method-performance studies have been called "collaborative studies" (AOAC), "method performance studies" (EPA), "precision studies" (ISO), and "round-robins" (ASTM). "Calibration" and "intercalibration" have occasionally been used to replace performance in "method-performance study," particularly when a new method is compared with an established or reference method. The use of these alternative terms should be discontinued in favor of internationally acceptable terminology.

Method-performance studies may include examination of the same materials by several methods, particularly when an established or reference method is compared with a new method or when, for example, a hazardous solvent is being replaced by a safer reagent. Such studies may be designated "method-comparison studies." The preferred design then requires all laboratories to perform all methods exactly as stated. Sometimes this is not possible. For example, in the comparison of an automated method with a manual method, laboratories with automated instruments may lack the experience required for performance of the manual method or laboratories performing the manual method may not have the automated instrument. Studies involving concurrent use of two or more methods on identical materials have been called method-comparison studies or method-equivalency studies; they may be carried out by a group of laboratories or more often within a single laboratory. The ruggedness of a method, i.e., the ability of a method to resist changes in the final result arising from minor changes in local or environmental variables, should not generally be the subject of an interlaboratory study. Such a study is best performed in a single laboratory or in a few to avoid wasting the resources of an interlaboratory study.

A laboratory-performance study focuses on the laboratories. It may permit the free choice of the method of analysis by any laboratory wanting to know how its own results compare with those of other participating laboratories. This type of study is also used for accrediting laboratories or analysts, in which case the choice of method may be restricted. Although the term "laboratories" is used throughout, this term really refers to the analyst together with the

physical facilities and environment of the laboratory. In addition, the term "analyst" includes those situations where a team of analysts is used, with different analysts performing separate steps of the entire method.

The terms "proficiency" and "evaluation" have been used instead of "performance" when the object of the study is appraising laboratory or analyst performance. This type of study should not be confused with auditing procedures conducted for oversight review or for accrediting or certifying laboratories. Such procedures may or may not require the demonstration of performance on test materials. Often their objective is merely documentation that such studies have been performed.

A material-certification study focuses on the material, not the methods. Such a study is usually conducted by an organization with legal or professional responsibility to prepare, characterize, and certify reference materials. The certified value is usually calculated from data obtained by "selected" or "referee" laboratories, using a definitive method of analysis or methods with small systematic errors. Occasionally, e.g., as with Kjeldahl nitrogen, a method-dependent certification may be provided. The certified value should be accompanied by a statement of uncertainty. Values with a lesser degree of reliability may be provided for information or as nominal values. Sometimes a "true value" is not required, but rather a consensus value obtained by representative laboratories, or even by the same laboratory over a period of time. Such a material may be used subsequently as a basis for comparison in interlaboratory studies or for quality control by individual laboratories.

REFERENCES

1. Pure Appl. Chem. 60, 855 (1988).
2. ISO 5725-1986 (currently being revised).