Appendix D: Guidelines for Collaborative Study Procedures To Validate Characteristics of a Method of Analysis

{*Note*: These guidelines incorporate symbols, terminology, and recommendations accepted by consensus by the participants at the IUPAC Workshop on Harmonization of Collaborative Analytical Studies, Geneva, Switzerland, May 4–5, 1987 [*Pure Appl. Chem*. **60**, 855–864(1988); published as "Guidelines for Collaborative Study of Procedure to Validate Characteristics of a Method of Analysis," *J. Assoc. Off. Anal. Chem*. **72**, 694–704(1989)]. The original guidelines were revised at Lisbon, Portugal, August 4, 1993, and at Delft, The Netherlands, May 9, 1994, *Pure Appl. Chem*. **67**, 331–343(1995). These revised, harmonized guidelines have been adopted by AOAC INTERNATIONAL as the guidelines for the AOAC Official Methods Program, *J. AOAC Int*. **78**(5), 143A–160A(1995). Although the directions were developed for chemical studies, some parts may be applicable to all types of collaborative studies.}

Summary Statement of AOAC Recommendation for Design of a Collaborative Study

Minimum Criteria for Quantitative Study

Minimum number of materials (see Note 1 on p. 4).—Five (only when a single level specification is involved for a single matrix may this minimum be reduced to 3).

Minimum number of laboratories.—Eight reporting valid data for each material (only in special cases involving very expensive equipment or specialized laboratories may the study be conducted with a minimum of 5 laboratories, with the resulting expansion in the confidence interval for the statistical estimates of the method characteristics).

Minimum number of replicates.—One, if within-laboratory repeatability parameters are not desired; 2, if these parameters are required. Replication should ordinarily be attained by blind replicates or split levels (Youden pairs).

Minimum Criteria for Qualitative Analyses

Ten laboratories reporting on 2 analyte levels per matrix, 6 test samples per level, and 6 negative controls per matrix. (*Note*: AOAC criteria for qualitative analyses are not part of the harmonized guidelines.)

1. Preliminary Work (Within One Laboratory)

1.1 Determine Purpose and Scope of the Study and Method

Determine purpose of the study (e.g., to determine attributes of a method, proficiency of analysts, reference values of a material, or to compare methods), the type of method (empirical, screening, practical, reference, definitive), and the probable use of the method (enforcement, surveillance, monitoring, acceptance testing, quality control, research). Also, on the basis of the relative importance of the various method attributes (bias, precision, specificity, limit of determination), select the design of the collaborative study. The directions in this document pertain primarily to determining the precision characteristics of a method, although many sections are also appropriate for other types of studies.

Alternatives for Method Selection

(*1*) Sometimes obvious (only method available).

(*2*) Critical literature review (reported within-laboratory attributes are often optimistic).

(*3*) Survey of laboratories to obtain candidate methods; comparison of within-laboratory attributes of candidate methods (sometimes choice may still not be objective).

(*4*) Selection by expert [AOAC-preferred procedure (selection by Study Director with concurrence of General Referee)].

(*5*) Selection by Committee (ISO-preferred procedure; often time-consuming).

(*6*) Development of new method or modification of existing method when an appropriate method is not available. (Proceed as a research project.) (This alternative is time-consuming and resource-intensive; use only as a last resort.)

1.2 Optimize Either New or Available Method

Practical Principles

(*1*) Do not conduct collaborative study with an unoptimized method. An unsuccessful study wastes a tremendous amount of collaborators' time and creates ill will. This applies especially to methods that are formulated by committees and have not been tried in practice.

(*2*) Conduct as much experimentation within a single laboratory as possible with respect to optimization, ruggedness, and interferences. Analysis of the same material on different days provides considerable information on variability that may be expected in practice.

Alternative Approaches to Optimization

(*1*) Conduct trials by changing one variable at a time.

(*2*) Conduct formal ruggedness testing for identification and control of critical variables. *See* Youden and Steiner (pp 33–36, 50–55). The actual procedure is even simpler than it appears. (This is an extremely efficient way for optimizing a method.)

(*3*) Use Deming simplex optimization to identify critical steps. *See* Dols and Armbrecht. The simplex concept can be used in the optimization of instrument performance and in application to analytical chemical method development.

1.3 Develop Within-Laboratory Attributes of Optimized Method

(Some items can be omitted; others can be combined depending on whether study is qualitative or quantitative.)

Determine calibration function (response vs concentration in pure or defined solvent) to determine useful measurement range of method. For some techniques, e.g., immunoassay, linearity is not a prerequisite. Indicate any mathematical transformations needed.

Determine analytical function (response vs concentration in matrix, including blank) to determine applicability to commodity(ies) of interest.

Test for interferences (specificity): (*1*) Test effects of impurities, ubiquitous contaminants, flavors, additives, and other components expected to be present and at usual concentrations. (*2*) Test nonspecific effects of matrices. (*3*) Test effects of transformation products, if method is to indicate stability, and metabolic products, if tissue residues are involved.

Conduct bias (systematic error) testing by measuring recoveries of analyte added to matrices of interest and to extracts, digests, or other treated solutions thereof. (Not necessary when method defines property or component.)

Develop performance specifications for instruments and suitability tests for systems (which utilize columns or adsorbents) to ensure satisfactory performance of critical steps (columns, instruments, etc.) in method.

Conduct precision testing at the concentration levels of interest, including variation in experimental conditions expected in routine analysis (ruggedness). In addition to estimating the "classical" repeatability standard deviation, s_r, the initiating laboratory may estimate the total within-laboratory standard deviation (s_e) whereby s_e is the variability at different days and with different calibration curves, by the same or different analysts within a single laboratory. This total within-laboratory estimate reflects both between-run (between-batch) and within-run (within-batch) variability.

Delineate the range of applicability to the matrices or commodities of interest.

Compare the results of the application of the method with existing, studied methods intended for the same purposes, if other methods are available.

If any of the preliminary estimates of the relevant performance of these characteristics are unacceptable, revise the method to improve them, and re-study as necessary.

Have method tried by analysts not involved in its development.

Revise method to handle questions raised and problems encountered.

1.4 Prepare Description of Method

Note: A collaborative study of a method involves practical testing of the written version of the method, in its specific style and format, by a number of laboratories on identical materials.

Prepare method description as closely as possible to format and style that will be used for eventual publication.

Clearly specify requirements for chromatographic materials, enzymes, antibodies, and other performance-related reagents.

Clearly describe and explain every step in the analytical method so as to discourage deviations. Use imperative directions; avoid subjunctive and conditional expressions as options as far as possible.

Clearly describe any safety precautions needed.

Edit method for completeness, credibility (e.g., buffer pH consistent with specified chemicals, volumes not greater than capacity of container), continuity, and clarity.

Check for inclusion of performance specifications and system suitability tests, defined critical points, and convenient stopping points. Incorporate physical or chemical constants of working standards solutions, e.g., absorptivities, half-scale deflections, recoveries, etc., or properties of operating solutions and chromatographic materials, e.g., pH, volumes, resolution, etc., and any other indicators (e.g., sum equals 100%) that suggest analysis is proceeding properly.

If time and resources are available, conduct pilot study involving 3 laboratories.

1.5 Invite Participation

Selection of Collaborators/Candidate Laboratories

Laboratories invited to participate should have personnel experienced in the basic techniques employed; experience with the method itself is not a prerequisite for selection. Lists of possible participants can be developed through personal contacts, technical societies, trade associations, or literature search, and advertisements in the Referee section of AOAC's magazine, *Inside Laboratory Management*. Collaborators are chosen by the organizers of the collaborative study from a diversity of laboratories with interest in the method, including regulatory agencies, industry, and universities.

Letter of Invitation

Address a formal letter to the individual responsible for assignment of laboratory effort. State reason for selecting that laboratory (e.g., as a volunteer or has responsibility or familiarity with the problem or method), estimated number of person-hours required for performance, number of test samples to be sent, number of analyses to be required, expected date for test sample distribution, and target date for completion of the study. *Emphasize the importance of management support in assigning the necessary time for the project*. Enclose a copy of the method and a return form or card (with postage affixed, if appropriate), requiring only a check mark for acceptance or refusal of the invitation, a signature, space for address corrections, telephone and fax numbers, e-mail, and date.

Laboratory Coordinator

With large studies, involving several analysts per laboratory, several familiarization samples, receipt of items at different times, or similar recurrent situations, acceptance of the invitation should be followed by a letter suggesting that a Laboratory Coordinator be appointed. The Laboratory Coordinator should be responsible for receiving and storing the study materials, assigning the work, dispensing study materials and information related to the study, seeing that the method is followed as written, accumulating the data, assuring that the data are correctly reported, and submitting the collaborative study manuscript within the deadline.

1.6 Instructions and Report Forms

Carefully design and prepare instructions and forms, and scrutinize them before distribution. A pilot study is also useful for uncovering problems in these documents.

Send instructions and report forms immediately on receipt of acceptance, independent of study materials, if selection of laboratories is not to be based on performance in pilot or training studies. The instructions should include in bold face or capital letters a statement:

THIS IS A STUDY OF THE METHOD, NOT OF THE LABORATORY. THE METHOD MUST BE FOLLOWED AS CLOSELY AS PRACTICABLE, AND ANY DEVIATIONS FROM THE METHOD AS DESCRIBED, NO MATTER HOW TRIVIAL THEY MAY SEEM, MUST BE NOTED ON THE REPORT FORM.

Include instructions on storage and handling, markings, and identifications to be noted, any special preparation for analysis, and criteria for use of practice or familiarization samples, if included. Pre-code the form for each laboratory and provide sufficient space for as much sequential data as may be required for proper evaluation of the results, including a check of the calculations.

The initiating laboratory should indicate the number of significant figures to be reported, usually 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 means and relative standard deviations are rounded to accommodate the significant figures of the standard deviation. For example, if the reproducibility standard deviation $s_R = 0.012$, the mean is reported as 0.147, not as 0.1473 or 0.15, and RSD_R, relative reproducibility standard deviation, is reported as 8.2%. 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 2 times the number of figures in the data plus 1.

When recorder tracing reproductions are required to evaluate method performance, request their submission both in the instructions and as a check item on the form. Provide instructions with regard to labeling of recorder tracings, such as identification with respect to item analyzed, axes, date, submitter, experimental conditions, and instrument settings.

Include in the report form a signature line for the analyst and lines for a printed or typed version of the name and address for correct acknowledgement.

Provide for a review by the laboratory supervisor. An example of a completed form is helpful. A questionnaire may be included or sent after completion of the analyses in which the questions can be designed to reveal if modifications have been made at critical steps in the method.

Request a copy of the calibration curve or other relationship between response and concentration or amount of analyte so that if discrepancies become apparent after examining all of the data, it can be determined whether the problem is in the calibration or in the analysis.

1.7 Familiarization or Practice Samples

If deemed necessary, supply as far ahead as practicable, familiarization samples, with instructions, before actual materials are sent. When familiarization samples have been submitted, supply forms for reporting progress toward satisfactory performance.

2. Design of the Collaborative Study

2.1 General Principles

The purpose of a collaborative study is to determine estimates of the attributes of a method, particularly the "precision" of the method that may be expected when the method is used in actual practice. The AOACI uses 2 terms to define the precision of a method under 2 circumstances of replication: repeatability and reproducibility. Repeatability is a measure of the variation, s_r^2, between replicate determinations by the same analyst. It defines how well an analyst can check himself using the same method on blind replicates of the same material or split levels (Youden pairs), under the same conditions (e.g., same laboratory, same apparatus, and same time). Reproducibility is a composite measure of variation, s_R^2, which includes the between-laboratory and within-laboratory variations. It measures how well an analyst in a given laboratory can check the results of another analyst in another laboratory using the same method to analyze the same test material under different conditions (e.g., different apparatus and different time). The between-laboratory variation represents a systematic error that reflects variation arising from environmental conditions (e.g., condition of reagent and instruments, variation in calibration factors, and interpretations of the steps of the method) associated with the laboratories used in the study. Therefore, it is important to identify the causes of the differences among laboratories so that they may be controlled. Otherwise they will be summed into s_R^2.

Present test samples sent for analysis as unknowns (blind) and coded in a random pattern. If necessary to conserve analyst time, an indication of the potential range of concentration or amount of analyte may be provided. If spiking solutions are used, provide one coded solution for each material. All spiking solutions should be identical in appearance and volume. Do not provide a single solution from which aliquots are to be removed for spiking. Any information with regard to concentration (e.g., utilizing factorial aliquots or serial dilutions of the same spiking solutions) or known replication is likely to lead to an underestimate of the variability.

The study must be extensive enough to assure sufficient data surviving in the face of possible loss of materials during shipment, inability of collaborators to participate after acceptance, and a maximum outlier rate of 2/9 and still maintain valid data from a minimum of 8 laboratories.

Improper preparation of reference standards and standard solutions can cause a significant portion of the analytical error. A decision must be made whether such error is to be considered separately or as part of the method, i.e., will the analysts procure their own standard solutions or will standards be provided by the Study Director. The decision depends primarily on the availability of the standard. If the standard is readily available, the analysts should prepare their own. If the standard is not readily available, the standard may be supplied, but physical constants, e.g., absorptivity of working standard solutions, should be incorporated into the description as a check on proper preparation of the solution.

Obtain the necessary administrative and operational approvals. Review by potential users of the method is also desirable.

2.2 Laboratories

Laboratories must realize the importance of the study. A large investment is being made in studying the method and this probably will be only collaborative study of the method that will performed. Therefore, it is important to have a fair and thorough evaluation of the method.

Type

The most appropriate laboratory is one with a responsibility related to the analytical problem. Laboratory types may be representative (selection of laboratories that will be using the method in practice), reference (assumed to be "best"), or the entire population of laboratories (usually certified or accredited) that will be using the method. Final selection of participants should be based on a review with the General Referee and others of each laboratory's capabilities and past performance in collaborative studies, followed up, if possible, by telephone conversations or by personal visits. Selection may

also be based on performance with familiarization samples. Sometimes only laboratories with dedicated or very specialized instruments must be used. If the study is intended for international consideration, laboratories from different countries should be invited to participate.

Number of Laboratories

Minimum of 8 laboratories submitting valid data (to avoid unduly large confidence bands about the estimated parameters). Only in special cases of very expensive equipment or specialized laboratories may the study be conducted with a minimum of 5 laboratories. Fewer laboratories widen the confidence limits of the mean and of the variance components (*see* design considerations). The optimum number of laboratories, balancing logistics and costs against information obtained, often is 8–10. However, larger studies are not discouraged.

For qualitative analyses, a minimum of 10 laboratories is needed; collaborative study must be designed to include 2 analyte levels per matrix, 6 test samples per level, and 6 negative controls per matrix. (*Note 1*: AOAC criteria for qualitative analyses are not part of the harmonized guidelines.)

Analysts

Most designs require only 1 analyst per laboratory. If analyst–within-laboratory variability is a desired variance component, multiple analysts should be requested from all participating laboratories. Ordinarily 2 analysts from the same laboratory cannot be substituted for different laboratories, unless standard solutions, reagents, chromatographic columns and/or materials, instrument calibrations, standard curves, etc., are prepared independently, and no consultation is permitted during the work. Different laboratories from the same organization may be used as separate laboratories if they operate independently with their own instruments, standards, reagents, and supervision.

2.3 Test Materials

Homogeneous Materials

Materials must be homogeneous; this is critical. Establish homogeneity by testing a representative number of laboratory samples taken at random before shipment. (A collaborator who reports an outlying value will frequently claim receipt of a defective laboratory sample.) The penalty for inhomogeneity is an increased variance in the analytical results that is not due to the intrinsic method variability.

Test Sample Coding

Code test samples at random so that there is no pre-selection from order of presentation.

Concentration Range

Choose analyte levels to cover concentration range of interest. If concentration range of interest is a tolerance limit or a specification level, bracket it and include it with materials of appropriate concentration. If design includes the determination of absence of analyte, include blank (not detectable) materials as part of range of interest.

Number of Materials

A minimum of 5 materials must be used in the collaborative study. Three materials are allowed but only when a single specification is involved for a single matrix.

Note 1: A material is an analyte (or test component)/matrix/concentration combination to which the method-performance parameters apply. This parameter determines the applicability of the method.

Note 2: The 2 test samples of blind or open duplicates are a single material (they are not independent).

The 2 test samples constituting a matched pair (called X and Y) are considered Youden matched pairs only if they are sufficiently close in composition. "Sufficiently close" would be considered as ≤5% difference in composition between X and Y. That is, given that the concentration of analyte in X (x_c) is higher than the concentration of the analyte in Y (y_c) then:

$$\frac{x_c \quad y_c}{x_c} \leq 0.05$$

or:

$$y_c \geq (x_c - 0.05x_c)$$

Note 3: The blank or negative control may or may not be a material, depending on the usual purpose of the analysis. For example, in trace analysis, where very low levels (near the limit of quantitation) are often sought, the blanks are considered as materials, and are necessary to determine certain statistical "limits of measurement;" however, if the blank is merely a procedural control, in macro-level analysis (e.g., fat in cheese), it would not be considered a material.

Nature of Materials

Materials should be representative of commodities usually analyzed, with customary and extreme values for the analyte.

Size of Test Samples

Furnish only enough test sample to provide the number of test portions specified in the instructions. If additional test portions are required, the collaborator must request them, with an explanation.

Interferences

If pertinent, some materials, but not all, should contain contaminants and interferences in concentrations likely to be encountered, unless they have been shown to be unimportant through within-laboratory study. The success of the method in handling interference on an intralaboratory basis will be demonstrated by passing systems suitability tests.

Familiarization Samples

With new, complex, or unfamiliar techniques, provide material(s) of stated composition for practice, on different days, if possible. The valuable collaborative materials should not be used until the analyst can reproduce the stated value of the familiarization samples within a given range. However, it should be pointed out that one of the assumptions of analysis of variance is that the underlying distribution

of results is independent of time (i.e., there is no drift). The Study Director must be satisfied that this assumption is met.

2.4 Replication

When within-laboratory variability is also of interest, as is usually the case, independent replication can be ensured by applying at least one of the following procedures (listed in suggested order of desirability; the nature of the design should not be announced beforehand):

(*1*) *Split levels (Youden pairs)*.—The 2 test materials, nearly identical but of slightly different composition (e.g., ≤5% difference in composition, *see 2.3 Number of Materials, Note 2*) are obtained either naturally or by diluting (or by fortifying) one portion of the material with a small amount of diluent (or of analyte). Both portions are supplied to the participating laboratories as test samples, each under a random code number, and each test sample should be analyzed only once; replication defeats the purpose of the design.

(*2*) *Split levels for some materials and blind duplicates for other materials in the same study*.—Obtain only single values from each test sample supplied.

(*3*) *Blind duplicate test samples, randomly coded*.—*Note*: Triplicate and higher replication are relatively inefficient when compared with duplicate test samples because replication provides additional information only on individual within-laboratory variability, which is usually the less important component of error. It is more effective to utilize resources for the analysis of more levels and/or materials rather than for increasing the number of replicates for the individual materials.

PRACTICAL PRINCIPLE: With respect to replication, the greatest net marginal gain is always obtained in going from 2 to 3 as compared to going from 3 to 4, 4 to 5, etc.

(*4*) *Independent materials*.—(*Note*: Unrelated independent materials may be used as a split level in the calculations of the precision parameters or for plotting. There should be ≤5% difference in composition for such materials (*see 2.3 Number of Materials, Note 2*). The more they differ in concentration, the less reliable the information they provide on within-laboratory variability.)

(*5*) *Known replicates*.—*Use of known replicates is a common practice*.—It is much preferable to use the same resources on blind replicates or split levels.

(*6*) *Quality control materials*.—Instead of obtaining repeatability parameters through the collaborative study, information can be obtained from use of quality control materials in each laboratory individually, for its own use, independent of the collaborative study, for a separate calculation of s_r, using 2 (or more) replicates from each quality control test, according to the pattern developed for each product.

2.5 Other Design Considerations

The design can be reduced in the direction of less work and less cost, but at the sacrifice of reduced confidence in the reliability of the developed information.

More work (values) is required if more confidence is needed, e.g., greater confidence is required to enforce a tolerance at 1.00 mg/kg than at 1.0 mg/kg. (The distinction is a precision requirement of the order of 1% rather than 10%.)

The estimate of the standard deviation or the corresponding relative standard deviation obtained from a collaborative study is a random variable that varies about its corresponding true value. For example, the standard deviation, s_r, which measures within laboratory or repeatability precision has associated with it a standard deviation (STD = s_r) describing its scatter about the true value σ_r. Therefore, s_r, whose STD (s_r) is a function of s_r^2, number of laboratories, and number of analyses per laboratory, will vary about σ_r from occasion-to-occasion even for the same test conditions and material. The STD s_R, which measures among laboratory or reproducibility precision, has a STD (s_R) that is a function of the random variables s_r^2 and s_L^2, number of laboratories, and number of analyses per laboratory. s_R will vary about its true value σ_R from occasion-to-occasion for the same test material.

The validity of extrapolating the use of a method beyond concentrations and components tested can be estimated only on the basis of the slope of the calibration curve (sensitivity) observed as a function of the nature and concentration of the matrix and contaminant components. If the signal is more or less independent of these variables, a reasonable amount of extrapolation may be utilized. The extrapolator assumes the burden of proof as to what is reasonable.

3. Preparation of Materials for Collaborative Studies

3.1 General Principles

Heterogeneity between test samples from a single test material must be negligible compared to analytical variability, as measured within the Study Director's laboratory.

The containers must not contribute extraneous analytes to the contents, and they must not adsorb or absorb analytes or other components from the matrix, e.g., water.

If necessary, the materials may be stabilized, preferably by physical means (freezing, dehydrating), or by chemical means (preservatives, antioxidants) which do not affect the performance of the method.

Composition changes must be avoided, where necessary, by the use of vapor-tight containers, refrigeration, flushing with an inert gas, or other protective packaging.

3.2 Materials Suitable for Collaborative Studies

Material and analyte stability: Ensure analyte and matrix stability over projected transport and projected length of study.

Single batch of homogenous, stable product such as milk powder, peanut butter, vegetable oil, starch, etc., is the best type of material.

Reference materials supplied by standards organizations such as National Institute of Standards and Technology (NIST, Gaithersburg, MD) and EC's Joint Research Center and Institute on Reference Materials and Methods (IRMM, Belgium) are excellent, unless they have easily recognizable characteristics (e.g., odor and color of NIST Orchard Leaves). However, they are of limited availability, composition, and analyte level. If available, they are expensive. Sometimes the certification organization may be interested in making reference materials available for the analyte under study, in which case it may assist in providing the material for the study.

Synthetic materials may be especially formulated with known amounts of analytes by actual preparation for the study. This procedure is best used for macro-constituents such as drugs or pesticide formulations.

Spiked materials consisting of normal or blank materials to which a known amount of analyte has been added may be used. The amount of analyte added should not be excessive in relation to the amount present (e.g., about 2×), and the analyte added should be in the same

chemical form as present in the commodities to be analyzed subsequently.

In drug and pesticide residue-type problems, it is often necessary to use spiked materials in order to assess recovery. However, because incurred residues are likely to present different problems from those of spiked residues, collaborative studies should include some test samples with incurred residues to ensure that the method is applicable under these conditions as well.

(*1*) *Preparation in bulk*.—This requires thorough and uniform incorporation of analyte, often by serial dilution of solids. The danger of segregation due to differences in densities always exists. Fluid materials susceptible to segregation should be prepared under constant agitation. Uniformity should be checked by direct analysis, with an internal standard, or by a marker compound (dye or radioactive label).

(*2*) *Test samples, individually prepared*.—A known amount of analyte is either weighed directly or added as an aliquot of a prepared solution to pre-measured portions of the matrix in individual containers. The collaborator is instructed to use each entire portion for the analysis, transferring the contents of the container quantitatively or a substantial weighed fraction of the portion. (This is the preferred alternative to spiked solid materials at trace [mg/kg] levels, at the expense of considerably more work.)

(*3*) *Concentrated unknown solutions for direct addition by collaborators to their own commodities*.—Should be used only as a last resort when instability of the analyte precludes distribution from a central point. To preclude direct analysis of the spiking solution, supply individual coded solutions to be added in their entirety to portions of the matrix for single analyses by each laboratory. All solutions should have the same volume and appearance. This type of material is analogous to that of test samples except for the source of matrix. This case should be used only for perishable commodities that are altered by all available preservation techniques.

Materials analyzed by another, presumably accurate, method, if available, in the Study Director's laboratory or by some or all the collaborators.

Only as an absolutely last resort (usually with unstable materials and preparation of material studies) should the collaborators be permitted to prepare their own materials for analysis. Since it is impossible to avoid the personal bias introduced by knowledge of the composition of the material, the materials should be prepared in each laboratory by an individual who will not be involved in the analyses.

3.3 Blanks

When the absence of a component is as important as its presence, when determinations must be corrected for the amount of the component or the presence of background in the matrix, or when recovery data are required, provision must be made for the inclusion of blank materials containing "none" (not detected) of the analyte. It is also important to know the variability of the blank and the tendency of the method to produce false positives. There are 2 types of blanks: matrix blanks and reagent blanks. Since laboratories often will utilize reagents from different sources, each laboratory should perform reagent blanks. Matrix blanks, when required, are an intrinsic part of the method, and the number of blanks needed depends on the combined variance of the material (s_M) and of the blank (s_B). Standard deviation reflecting the total variability of a blank corrected value will be $s = (s_M^2 + s_B^2)^{1/2}$.

3.4 Limit of Detection/Quantitation

If the limit of detection/quantitation is important, it is necessary to provide a design which gives special attention to the number of blanks, and to the necessity for interpreting false positives and false negatives. In all cases, the definition of limit of detection/quantitation used in the study must be given by the Study Director.

3.5 Controls

When separation from interferences is critical to the analysis, appropriate materials incorporating these interferences must be included.

PRACTICAL ADVICE: Always allow for contingencies and prepare more sets (e.g., 25% more) of laboratory samples than there are collaborators. Some packages may never arrive, some materials may spoil, and some may be lost or the container broken. New laboratories may have to be substituted for those which are unable to complete the promised work. Some sets may have to be analyzed at a later time for different purposes, such as to verify stability on storage.

4. Submission of Test Samples

4.1 Sending Collaborative Study Material

Notify collaborators of shipping arrangements, including waybill numbers, arrival time, and required storage conditions.

Label test samples legibly and without ambiguity.

Pack shipping cartons well and label properly to avoid transportation delays. If the containers are breakable, pack well to minimize possibility of breakage. If material is perishable, ship frozen with solid CO_2, sufficient to last several days longer than anticipated travel time. Use special transportation services, if necessary. For international delivery, mark as "Laboratory samples—no commercial value" or other designation as required by customs regulations of the country to which the package is being sent. Hazardous materials must be packed and labeled as required by transportation regulations. Animal and plant products sent across international borders may require special certification from health authorities.

Include a return slip, to confirm safe receipt, with each package. If not sent previously, include copy of method, instructions, and report forms.

Provide instructions for proper storage of test samples between unpacking and analysis. Note that analysts should not use thawed or decomposed test samples without consulting the Study Director.

When it is important to have instruments calibrated with the same reference material, supply reference material to collaborators. Provision for supplying reference standards is particularly important when commercial sources of standards have not yet been developed. The inclusion of a working standard solution as an unknown is useful to establish a consensus value for standardization of quality control parameters, such as absorptivity, retention time, and sensitivity (change in signal intensity divided by the change in concentration).

4.2 Obligations of Collaborators

Analyze test samples at times indicated, according to submitted protocol. With unstable materials (e.g., with microbial or decomposition problems), analyses must be started at specified times.

FOLLOW METHOD EXACTLY (*this is critical*). If method is unclear, contact Study Director. Any deviation, such as the necessity to substitute reagents, columns, apparatus, or instruments, must be

recorded at the time and reported. If the collaborator has no intention of following the submitted method, he or she should not participate in the study. If the collaborator wishes to check another method on the same materials, additional test samples should be requested for that purpose, to be analyzed separately.

Conduct exactly the number of determinations stated in the instructions. Any other number complicates the statistical analysis. Too few determinations may require discarding the results from that laboratory for that material or inserting "missing values"; too many values may require discarding the contribution of that laboratory or at least some of the values. If a laboratory cannot follow instructions as to number of analyses to perform, it raises a question as to its ability to follow the method.

Report individual values, including blanks. Do not average or do other data manipulations unless required by the instructions. Undisclosed averaging distorts statistical measures. If blank is larger than determination, report the negative value; do not equate negative values to zero. Follow or request instructions with regard to reporting "traces" or "less than." Descriptive (i.e., nonquantitative) terms are not amenable to statistical analysis and should be avoided. When results are below the limit of determination, report actual calculated result, regardless of its value.

Supply raw data, graphs, recorder tracings, photographs, or other documentation as requested in the instructions.

Since collaborators may have no basis for judging whether a value is an outlier, the results should be communicated to the Study Director as soon as the protocol is complete and before time and equipment are reassigned, so that repeat assays may be performed at once, if necessary and if permitted by the protocol.

Note: The sooner an apparent outlier is investigated, the greater the likelihood of finding a reason for its occurrence.

The most frequent causes of correctable outliers are:

- Incorrect calculations and arithmetic errors.
- Errors in reporting, such as transposition of numbers, misplacement of the decimal point, or use of the wrong units.
- Incorrect standards due to weighing or volumetric errors (check physical constants or compare against freshly prepared standard solutions).
- Contamination of reagents, equipment, or test samples.

5. Statistical Analysis

5.1 Initial Review of Data (Data Audit)

The Study Director may first plot the collaborative study results, material by material (or one value against the other for a split level [Youden pair]), value vs laboratory, preferably in ascending or descending order of reported average concentration. Usually major discrepancies will be apparent: displaced means, unduly spread replicates, outlying values, differences between methods, consistently high or low laboratory rankings, etc.

Only valid data should be included in the statistical analysis. Valid data are values that the Study Director has no reason to suspect as being wrong. Invalid data may result when: (*1*) the method is not followed; (*2*) a nonlinear calibration curve is found although a linear curve is expected; (*3*) system suitability specifications were not met; (*4*) resolution is inadequate; (*5*) distorted absorption curves arise; (*6*) unexpected reactions occur; or (*7*) other atypical phenomena materialize. Other potential causes of invalid data are noted previously.

5.2 Outliers

Collaborative studies seem to have an inherent level of outliers, the number depending on the definition of outliers and the basis for calculation (analytes, materials, laboratories, or determinations). Rejection of more than 2/9 of the data from each material in a study, without an explanation (e.g., failure to follow the method), is ordinarily considered excessive. Study must maintain valid data from a minimum of 8 labs. For larger studies, a smaller acceptable percentage of rejections may be more appropriate. Determine the probability that the apparent aberrant value(s) is part of the main group of values considered as a normal population by applying the following tests in order:

(*1*) *Cochran test* for removal of laboratories (or indirectly for removal of extreme individual values from a set of laboratory values) showing significantly greater variability among replicate (within-laboratory) analyses than the other laboratories for a given material. Apply as a 1-tail test at a probability value of 2.5%.

To calculate the Cochran test statistic: Compute the within-laboratory variance for each laboratory and divide the largest of these by the sum of all of these variances. The resulting quotient is the Cochran statistic which indicates the presence of a removable outlier if this quotient exceeds the critical value listed in the Cochran table for P = 2.5% (1-tail) and L (number of laboratories), **Appendix 1**.

(*2*) Grubbs tests for removal of laboratories with extreme averages. Apply in the following order: single value test (2-tail; P = 2.5%); then if no outlier is found, apply pair value test (2 values at the highest end, 2 values at the lowest end, and 2 values, one at each end, at an overall P = 2.5%).

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 averages with the lowest average removed (s_L). Then calculate the percentage decrease in SD as follows:

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

The higher of these 2 percentage decreases is the single Grubbs statistic, which signals the presence of an outlier to be omitted if it *exceeds* the critical value listed in the single Grubbs tables at the P = 2.5% level, 2-tail, for L laboratories, **Appendix 2**.

To calculate the Grubbs pair 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 table at the P = 2.5% level, for L laboratories, **Appendix 2**.

(*3*) If the single value Grubbs test signals the need for outlier removal, remove the single Grubbs outlier and recycle back to the Cochran test as shown in the flow chart, **Appendix 3**.

If the single value Grubbs test is negative, check for masking by performing the pair value Grubbs test. If this second test is positive,

remove the 2 values responsible for activating the test and recycle back to the Cochran test as shown in the flow chart, **Appendix 3**, and repeat the sequence of Cochran, single value Grubbs, and pair value Grubbs. Note, however, that outlier removal should stop before more than 2/9 laboratories are removed.

(*4*) If no outliers are removed for a given cycle (Cochran, single Grubbs, pair Grubbs), outlier removal is complete. Also, stop outlier removal whenever more than 2/9 of the laboratories are flagged for removal. With a higher removal rate, either the precision parameters must be taken without removal of all outliers or the method must be considered as suspect.

Note: The decision as to whether a value(s) should be removed as an outlier ultimately is not statistical in nature. The decision must be made by the Study Director on the basis of the indicated probability given by the outlier test and any other information that is pertinent. (However, for consistency with other organizations adhering to the harmonized outlier removal procedure, the estimate resulting from rigid adherence to the prescribed procedure should be reported.)

5.3 Bias (Systematic Deviation) of Individual Results

Bias is defined as follows:

(Estimated) bias =
mean amount found – amount added (or known or assigned value)

Single-value error and recovery are defined as follows:

Error of a single value =
the single value – amount added (true value)

There are 2 methods for defining percent recovery: marginal and total. The formulas used to estimate these percent recoveries are provided in the following:

$$\text{Marginal \%Rec} = 100R_M = 100((C_f - C_u)/C_A)$$

$$\text{Total \%Rec} = 100R_T = 100(C_f)/(C_u + C_A)$$

where C_f is the amount found for the fortified concentration, C_u is the amount present originally for the unfortified concentration, and C_A is the amount added for the added concentration. The amount added is known or fixed and should be a substantial fraction of, or more than, the amount present in the unfortified material; all other quantities are measured and are usually reported as means, all of which have variations or uncertainties. The variation associated with the marginal percent recovery is $\text{var}(100R_M) = (100^2/C_A^2)[\text{var}(C_f) + \text{var}(C_u)]$ is larger than the variation associated with the total percent recovery. The variation associated with total percent recovery is $\text{var}(100R_T) = [100^2/(C_u + C_A)^2][\text{var}(C_f) + (R_T^2)\text{var}(C_u)]$. In each formula var means variance and refers to the concentration variation for the defined concentrations.

A true or assigned value is known only in cases of spiked or fortified materials, certified reference materials, or by analysis by another (presumably unbiased) method. Concentration in the unfortified material is obtained by direct analysis by the method of additions. In other cases, there is no direct measure of bias, and consensus values derived from the collaborative study itself often must be used for the reference point.

Notes: (*1*) Youden equates "true" or "pure" between-laboratory variability (not including the within-laboratory variability) to the variability in bias (or variability in systematic error) of the individual laboratories. Technically, this definition refers to the average squared difference between individual laboratory biases and the mean bias of the assay.

(*2*) The presence of random error limits the ability to estimate the systematic error. To detect the systematic error of a single laboratory when the magnitude of such error is comparable to that laboratory's random error, at least 15 values are needed, under reasonable confidence limit assumptions.

5.4 Precision

The precision of analytical methods is usually characterized for 2 circumstances of replication: within laboratory or repeatability and among laboratories or reproducibility. Repeatability is a measure of how well an analyst in a given laboratory can check himself using the same analytical method to analyze the same test sample at the same time. Reproducibility is a measure of how well an analyst in one laboratory can check the results of another analyst in another laboratory using the same analytical method to analyze the same test sample at the same or different time. Given that test samples meet the criteria for a single material, the repeatability standard deviation (s_r) is:

$$s_r = (\Sigma d_i^2/2L)^{1/2}$$

where d_i is the difference between the individual values for the pair in laboratory i and L is the number of laboratories or number of pairs.

The reproducibility standard deviation (s_R) is computed as:

$$s_R = (1/2(s_d^2 + s_r^2))^{1/2}$$

where $s_d^2 = \Sigma(T_i - \overline{T})^2/(2(L-1))$, T_i is the sum of the individual values for the pair in laboratory i, $\overline{T}$ is the mean of the T_i across all laboratories or pairs, L is the number of laboratories or pairs, and s_r^2 is the square of $s_r = (\Sigma d_i^2/2L)^{1/2}$.

When the pairs of test samples meet the criteria for Youden matched pairs, i.e., when:

$$[(x_c - y_c)/x_c] \le 0.05$$

or

$$y_c \ge (x_c - 0.05x_c),$$

s_t, a practical approximation for repeatability standard deviation, is calculated as:

$$s_r = [\Sigma(d_i - \overline{d})^2/(2(L-1))]^{1/2}$$

where d_i is the difference between the individual values for the pair in laboratory i, $\overline{d}$ is the mean of the d_i across all laboratories or pairs, and L is the number of laboratories or pairs. The reproducibility standard deviation, s_R, which reflects the square root of the average of the reproducibility variances for the individual materials (i.e., $s_R = [\frac{1}{2}(s_{Rx}^2 + s_{Ry}^2)]^{1/2}$), previously called X and Y, should be determined only if the individual variances are not significantly different from each other. To compare s_{Rx}^2 and s_{Ry}^2, the following formula may be used.

$$t = \frac{(s_{Rx}^2 \quad s_{Ry}^2)(L \quad 2)^{1/2}}{2[(s_{Rx}^2)(s_{Ry}^2) \quad (cov_{xy})^2]^{1/2}}$$

where $s_{Rx}^2 = [1/(L-1)][\Sigma x_i^2 - (\Sigma x_i)^2/L]$, $s_{Ry}^2 = [1/(L-1)][\Sigma y_i^2 - (\Sigma y_i)^2/L]$, and $cov_{xy} = [1/(L-1)][\Sigma x_i y_i - (\Sigma x_i \Sigma y_i)/L]$. If t is greater than or equal to the tabular t-value for L – 2 degrees of freedom for a significance level of $\alpha = 0.05$, this may be taken to indicate that s_{Rx}^2 and s_{Ry}^2 are not equivalent and should not be pooled for a single estimate of s_R^2. That is, s_{Rx}^2 and s_{Ry}^2 should be taken as the reproducibility variance estimates for the individual test materials X and Y, respectively. This means that there is no rigorous basis for calculating s_r^2 because the within laboratory variability cannot be estimated directly.

Though s_r and s_R are the most important types of precision, it is the relative standard deviations (RSD_r % = $100s_r$/mean and RSD_R % = $100s_R$/mean) that are the most useful measures of precision in chemical analytical work because the RSD values are usually independent of concentration. Therefore, the use of the RSD values facilitates comparison of variabilities at different concentrations. When the RSD increases rapidly with decreasing concentration or amount, the rise delineates the limit of usefulness of the method (limit of reliable measurement).

5.5 HORRAT

HORRAT value is the ratio of the reproducibility relative standard deviation, expressed as a percent (RSD_R, %) to the predicted reproducibility relative standard deviation, expressed as a percent ($PRSD_R$, %), i.e.,

$$HORRAT = \frac{RSD_R,\%}{PRSD_R,\%}$$

where $PRSD_R$, % = $2C^{-0.1505}$ and C = the estimated mean concentration. HORRAT values between 0.5 to 1.5 may be taken to indicate that the performance value for the method corresponds to historical performance. The limits for performance acceptability are 0.5–2.

The precision of a method must be presented in the collaborative study manuscript. The HORRAT will be used as a guide to determine the acceptability of the precision of a method.

The HORRAT is applicable to most chemical methods. HORRAT is not applicable to physical properties (viscosity, RI, density, pH, absorbance, etc.) and empirical methods [e.g., fiber, enzymes, moisture, methods with indefinite analytes (e.g., polymers) and "quality" measurements, e.g., drained weight]. Deviations may also occur at both extremes of the concentration scale (near 100% and $\leq 10^{-8}$). In areas where there is a question if the HORRAT is applicable, the General Referee will be the determining judge.

The following guidelines should be used to evaluate the assay precision:

- HORRAT ≤ 0.5—Method reproducibility may be in question due to lack of study independence, unreported averaging, or consultations.
- 0.5 < HORRAT ≤ 1.5—Method reproducibility as normally would be expected.
- HORRAT > 1.5—Method reproducibility higher than normally expected: the Study Director should critically look into possible reasons for a "high" HORRAT (e.g., were test samples sufficiently homogeneous, indefinite analyte or property?), and discuss this in the collaborative study report.
- HORRAT > 2.0—Method reproducibility is problematic. A high HORRAT may result in rejection of a method because it may indicate unacceptable weaknesses in the method or the study. Some organizations may use information about the HORRAT as a criterion not to accept the method for official purposes (e.g., this is currently the case in the EU for aflatoxin methods for food analysis, where only methods officially allowed are those with HORRATs ≤ 2).

5.6 Incorrect, Improper, or Illusory Values (False Positive and False Negative Values)

These results are not necessarily outliers (no *a priori* basis for decision), since there is a basis for determining their incorrectness (a positive value on a blank material, or a zero (not found) or negative value on a spiked material). There is a statistical basis for the presence of false negative values: In a series of materials with decreasing analyte concentration, as the RSD increases, the percent false negatives increases from an expected 2% at an RSD = 50% to 17% at an RSD = 100%, merely from normal distribution statistics alone.

When false positives and/or false negatives exceed about 10% of all values, analyses become uninterpretable from lack of confidence in the presence or absence of the analyte, unless all positive laboratory samples are re-analyzed by a more reliable (confirmatory) method with a lower limit of determination than the method under study. When the proportion of zeros (not necessarily false negatives) becomes greater than approximately 30%, the distribution can become bimodal and even more uninterpretable (is the analyte present or absent?).

5.7 Final Collaborative Study Manuscript

The final manuscript should contain a description of the materials used, their preparation, any unusual features in their distribution, and a table of all *valid* data, including outliers. When replication is performed, the individual values, not just averages, must be given, unless the method requires averages (e.g., microbiological methods). Values not used for specified reasons, such as decomposition, failure to follow method, or contamination, should not be included in the table since they may be included erroneously in subsequent recalculations. AOAC INTERNATIONAL requires the calculation and reporting of mean, percent recovery (% Rec), HORRAT, repeatability (within-laboratory, s_r) and reproducibility (interlaboratory, s_R) standard deviations, and repeatability and reproducibility relative standard deviations (RSD_r and RSD_R, respectively). The accuracy (bias, trueness) of a method measuring a specific, identifiable analyte should be presented in the collaborative study manuscript as a recovery of added (spiked) analyte, as the results of analysis of a reference material, or by comparison with results by a reference method. Methods that are unable to report accuracy because of the unavailability of an accepted "true" value, or because of the nature of the method (empirical, microbiological, quality factors) should mention the reason in the manuscript. Proofread tables very carefully since many errors are of typographical origin. Give the names of the participants and their organizations, including complete contact information (name, preliminary address, telephone and fax numbers, and e-mail address).

The final manuscript should be published in a generally accessible publication, or availability of the report from the organization sponsoring the method should be indicated in the published method. Without public documentation, the significance of the study is very limited.

The manuscript should be sent to all participants, preferably at the preliminary stage, so that clerical and typographical errors may be corrected before publication. If changes in values from the original submission are offered, they must be accompanied by an explanation.

Example of Table of Interlaboratory Study Results: *See* **Table 1**.

The summary table as it will appear in the *Official Methods of Analysis of AOAC INTERNATIONAL* is given in **Table 2**.

6. References

(1) W.J. Youden & E.H. Steiner (1975) *Statistical Manual of the AOAC*, AOAC INTERNATIONAL, 481 N. Frederick Ave, Suite 500, Gaithersburg, MD 20877-7077, USA. The fifth printing (1987) contains several explanatory footnotes.

(2) G.T. Wernimont (1985) *Use of Statistics to Develop and Evaluate Analytical Methods*, W. Spendley (Ed.) AOAC INTERNATIONAL, 481 N. Frederick Ave, Suite 500, Gaithersburg, MD 20877-7077, USA.

(3) T. Dols & B. Armbrecht (1976) *J. Assoc. Off. Anal. Chem.* **59**, 1204–1207.

(4) International Organization for Standardization Guide 18, ISO, Case Postale 56, CH-1211 Geneva, Switzerland, and other national standards organizations.

(5) International Organization for Standardization ISO 5725, ISO, Case Postale 56, CH-1211 Geneva, Switzerland, and other national standards organizations.

Appendix 1 Critical values for the Cochran maximum variance ratio at the 2.5% (1-tail) rejection level, expressed as the percentage the highest variance is of the total variance

L = number of laboratories at a given level (concentration)
r = number of replicates per laboratory

L	r = 2	r = 3	r = 4	r = 5	r = 6
4	94.3	81.0	72.5	65.4	62.5
5	88.6	72.6	64.6	58.1	53.9
6	83.2	65.8	58.3	52.2	47.3
7	78.2	60.2	52.2	47.3	42.3
8	73.6	55.6	47.4	43.0	38.5
9	69.3	51.8	43.3	39.3	35.3
10	65.5	48.6	39.9	36.2	32.6
11	62.2	45.8	37.2	33.6	30.3
12	59.2	43.1	35.0	31.3	28.3
13	56.4	40.5	33.2	29.2	26.5
14	53.8	38.3	31.5	27.3	25.0
15	51.5	36.4	29.9	25.7	23.7
16	49.5	34.7	28.4	24.4	22.0
17	47.8	33.2	27.1	23.3	21.2
18	46.0	31.8	25.9	22.4	20.4
19	44.3	30.5	24.8	21.5	19.5
20	42.8	29.3	23.8	20.7	18.7
21	41.5	28.2	22.9	19.9	18.0
22	40.3	27.2	22.0	19.2	17.3
23	39.1	26.3	21.2	18.5	16.6
24	37.9	25.5	20.5	17.8	16.0
25	36.7	24.8	19.9	17.2	15.5
26	35.5	24.1	19.3	16.6	15.0
27	34.5	23.4	18.7	16.1	14.5
28	33.7	22.7	18.1	15.7	14.1
29	33.1	22.1	17.5	15.3	13.7
30	32.5	21.6	16.9	14.9	13.3
35	29.3	19.5	15.3	12.9	11.6
40	26.0	17.0	13.5	11.6	10.2
50	21.6	14.3	11.4	9.7	8.6

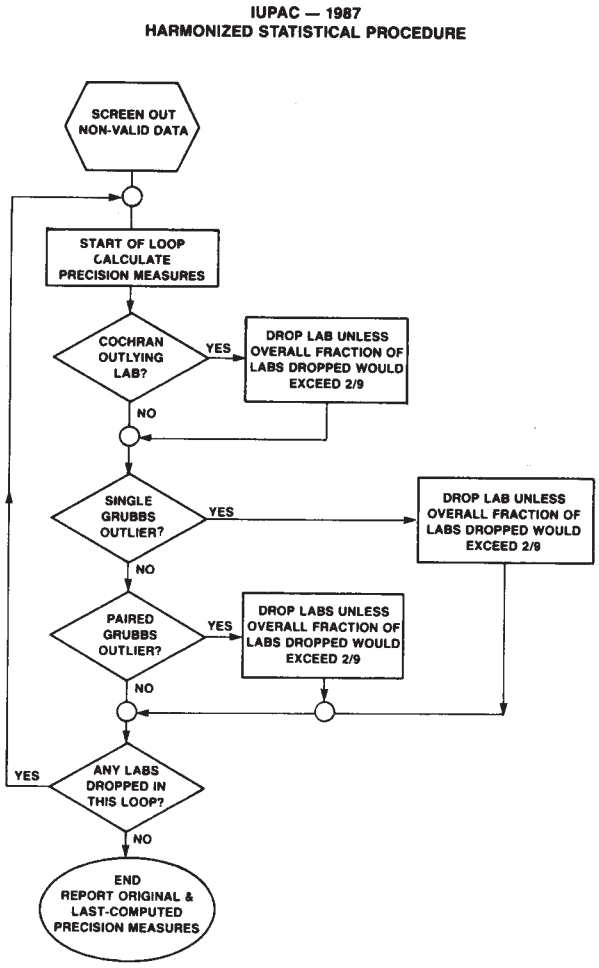
Cochran statistic = (largest individual within-laboratory variance)/(sum of all the within-laboratory variances).

Appendix 2 Critical values for the Grubbs extreme deviation outlier tests at the 2.5% (2-tail), 1.25% (1-tail) rejection level, expressed as the percent reduction in the standard deviations caused by removal of the suspect value(s) (*see* text for calculating the Grubbs statistics)

L = number of laboratories at a given level (concentration)

L	One highest or lowest	Two highest or two lowest	One highest and one lowest
4	86.1	98.9	99.1
5	73.5	90.3	92.7
6	64.0	81.3	84.0
7	57.0	73.1	76.2
8	51.4	66.5	69.6
9	46.8	61.0	64.1
10	42.8	56.4	59.5
11	39.3	52.5	55.5
12	36.1	48.5	51.6
13	33.8	46.1	49.1
14	31.7	43.5	46.5
15	29.9	41.2	44.1
16	28.3	39.2	42.0
17	26.9	37.4	40.1
18	25.7	35.9	38.4
19	24.6	34.5	36.9
20	23.6	33.2	35.4
21	22.7	31.9	34.0
22	21.9	30.7	32.8
23	21.2	29.7	31.8
24	20.5	28.8	30.8
25	19.8	28.0	29.8
26	19.1	27.1	28.9
27	18.4	26.2	28.1
28	17.8	25.4	27.3
29	17.4	24.7	26.6
30	17.1	24.1	26.0
40	13.3	19.1	20.5
50	11.1	16.2	17.3

Source: Both tables were calculated by R. Albert (October 1993) by computer simulation involving several runs of approximately 7000 cycles each for each value, and then smoothed. Although the table of **Appendix 1** 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.



Appendix 3—Flowchart.

Table 1 [x] Collaborative tests carried out at the international level in [year(s)] by [organization(s)] in which [y and z] laboratories participated, each performing [k] replicates, gave the following statistical results [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 outlying laboratories removed

Mean ($\bar{}$)
True or accepted value, if known

Repeatability standard deviation (s_r)
Repeatability relative standard deviation (RSD_r)
Repeatability value, r ($2.8 \times s_r$)
Total within laboratory standard deviation (s_e)—optional if s_r is not valid.

Reproducibility standard deviation (s_R)
Reproducibility relative standard deviation (RSD_R)
HORRAT
Reproducibility value, R ($2.8 \times s_R$)
Percent recovery (% Rec), if applicable

The repeatability and reproducibility values may also be expressed as a relative value (as a percentage of the determined mean value), when the results so suggest.
If the recovery and precision values are more or less constant for all materials or for group of materials, an overall average value may be presented. Although such averaging may not have statistical validity, it does have practical value.

Table 2 Model table for presentation of chemistry results from AOAC Official Methods

Table 200X.XX Interlaboratory results for [analyte] by [technique]

Material						Reproducibility	
Matrix	Level (units)	No. of labs[a(b)]	Mean (units)	Recovery, %	Repeatabiltiy RSD_r, %	RSD_R, %	HORRAT

[a(b)] a = Number of laboratories remaining after removal of the number of outliers indicated by (*b*).