



Project Summary

Guidelines for Conducting Single Laboratory Evaluations of Biological Methods

William D. McKenzie and Theodore A. Olsson III

The single laboratory test is used to establish the data quality that can be achieved within a single laboratory. It provides a basis for deciding whether or not a given method merits collaborative testing and it more clearly defines a method's potential for inclusion as part of an operational monitoring network. This summary provides a brief description of the suggested procedures for single laboratory testing.

Phases of the single laboratory test include identification of procedural variables that must be carefully controlled (ruggedness testing), evaluation of method sensitivity, identification of the limits of reliable measurement, evaluation of systematic error (bias), and identification of method precision and accuracy.

The chemical composition of all sample material must be verified during the single laboratory test. Sample material should have a concentration range, in the same sample matrix, that would be encountered if the method was being routinely used for its intended purpose. Some phases of the test should make use of certified reference materials.

The resulting single laboratory test data and the revised (ruggedness tested) method protocol will ultimately be used as part of the basis for deciding whether or not to proceed with collaborative testing.

The Project Summary was developed by EPA's Environmental Monitoring Systems Laboratory, Las Vegas, NV, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction and Basic Test Procedures

This summary provides a brief description of the suggested procedures for single laboratory testing. These suggestions are presented primarily as guidance to EPA contract laboratories that are involved in the single laboratory testing of biological methods.

Single laboratory testing is used to establish the data quality that can be achieved within a single laboratory. It provides a basis for deciding whether or not a given method merits collaborative testing and it more clearly defines a method's potential for inclusion as part of an operational monitoring network. The single laboratory test includes identification of procedural variables that must be carefully controlled (ruggedness testing), evaluation of method sensitivity, identification of the limits of reliable measurement, evaluation of systematic error (bias), and identification of method precision and accuracy. A complete protocol, for the method being evaluated, must be received by the evaluating laboratory prior to test initiation. All method requirements and procedural instructions should be clearly presented. It is obviously important that these written instructions be technically correct, complete, and as unambiguous as possible. However, the laboratory conducting the single laboratory test is not usually responsible for the actual method itself. The laboratory must strictly follow the method procedures and method requirements (experimental conditions, reagents, laboratory equipment, storage of samples, maintenance of experimental organisms, blanks, standards, replicates, etc.) as they are written in the protocol.

For the purposes of this guidance document, biological methods will include procedures used to analyze biological tissues and fluids as well as the various biological tests for toxicity, mutagenicity, etc. Single laboratory test objectives will be somewhat different depending on the method that is being evaluated. To determine a method's capability for accuracy (and for systematic error), the testing laboratory must have a reference sample material and there must be a known response (true value) for the material. When a method calls for analysis of biological tissue or biological fluid, the testing laboratory will usually have a reference material for which there is a true response value, e.g., samples with certified compound concentrations or perhaps with certified enzymatic activity levels. However, when the technique being evaluated includes a toxicity test or perhaps an algal assay (population stimulation or inhibition) there is frequently no "true value" or "true response" and hence the method's single laboratory capability for accuracy cannot really be determined. Under these conditions, an average test response should be acquired by conducting successive analyses on the same concentration of the same reference material. Because of such differences, the specific requirements for each single laboratory test must always be confirmed with the sponsor before single laboratory testing begins.

Ten successive analyses (i.e., acquiring 10 valid responses by following the method protocol) have been suggested for several phases of the single laboratory test. The method precision determination, for example, requires that 10 successive independent analyses be conducted on the same sample material. Multistage calculations to determine the required number of analyses might be conducted during the single laboratory test as more information becomes available on the expected variance. However, 10 analyses will allow the test laboratory to estimate the standard deviation to within 45 percent of its true value (at a 95 percent confidence interval). The single laboratory test cost will rapidly increase as more of these successive analyses are required since each additional value must represent a valid test response and therefore will include whatever quality control analyses (blanks, replicates, etc.) are required in the original method protocol to insure a valid test response.

The single laboratory test cannot really evaluate a method's scope of applicability,

i.e., the array of sample types or environmental situations for which the method would be able to provide useful data. However, the single laboratory test should include various test samples having a concentration range, in the same sample matrix, that would be encountered if the method was being routinely used for its intended purpose. Chemical composition of all sample material must be verified during the single laboratory test.

Some phases of the single laboratory test should make use of certified reference materials. These reference materials are samples that have a known chemical composition or enzymatic activity level. In some cases, they might be samples which are known to produce a certain response (true value) from a given test method. Any reference material used should be readily available to other testing laboratories. Both the National Bureau of Standards and the U.S. Environmental Protection Agency have certified various types of samples for use as reference materials.

Before the single laboratory test begins, the evaluating laboratory must review the

method protocol and make notations where ambiguous statements are made or where more detail is needed. These questions must be resolved with the sponsor, and, if necessary, a second protocol prepared, before the actual laboratory test begins. The single laboratory test should begin with ruggedness testing but it is important that the laboratory performing the single laboratory test plan all the assays required in advance in order to prevent duplication of effort. Some of the following evaluations may be performed simultaneously, depending on the nature of the test methods being investigated. Therefore, it is important to make advance preparations for the efficient use of time and available funds.

Figure 1 illustrates how the different analyses can provide data for more than one portion of the single laboratory test. This figure assumes that a single test material is being used, i.e., different concentrations of the same reference material. Many reference materials are available and the type of sample material(s) used should always be discussed

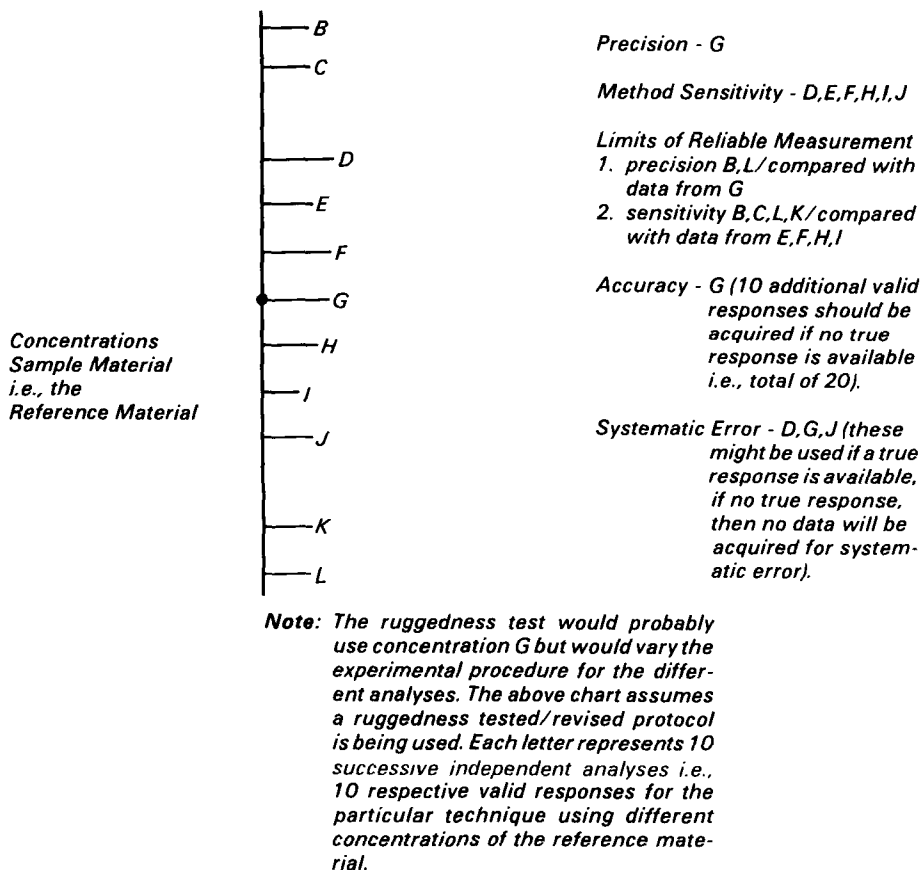


Figure 1. Example of how the evaluation analyses can provide data for multiple portions of the single laboratory test.

with the sponsor prior to beginning a single laboratory test.

Test Phases

Ruggedness Testing

The single laboratory test should first identify any procedural variables that must be carefully controlled. If the given method is "rugged" it will not be susceptible to the inevitable, modest departures in routine and the results obtained will not be altered by these minor variations. If the results are altered by small procedural variations, it is important to emphasize in the protocol that a specific step must be strictly followed or, in some cases, to indicate the limits of allowable variability. For example, the ruggedness test results might indicate that a certain (protocol directed) temperature requirement of 20°C was a critical procedure and that slight variations in this temperature (at a given phase of the test) would produce an altered test result. The method protocol will then need to be revised to emphasize that the stated temperature requirement must be strictly followed and, perhaps, to provide more detail on any quality control steps associated with temperature determinations. Depending on the sponsor's requirements, additional tests might be conducted to determine the specific narrow range of temperatures that would be allowable, i.e., a protocol revision noting that the temperature cannot vary except by a stated amount. The minor departures in routine selected for ruggedness testing do not always need to be variations in the protocol. Unjustified latitude is sometimes given in a protocol simply because of insufficient detail.

A single concentration of one test material can be used for the ruggedness test. The suggested approach does not seek to study each separate variable in an individual sequential fashion, but rather it provides for the introduction of several changes (protocol variations) simultaneously. At least seven variables should be selected which will require that the test laboratory conduct eight separate analyses. However, one of these variables could be a meaningless variable thus providing a modified control for the ruggedness test itself. Basically, the differences between the "protocol directed" result and the "protocol altered" result for each variable are compared. If one or two variables are having an effect on the test result, their respective differences (directed vs. altered) will be substantially larger than the group of differ-

ences associated with the other variables. Most of the modest procedural alterations should have little effect on the test result since the variations should only be of a magnitude that could be made by a qualified laboratory following the written method protocol.

Table 1 summarizes a ruggedness test using seven variables which will therefore require eight separate analyses. The varied condition is to be either slightly above or slightly below the "protocol directed" condition. The "protocol directed" conditions are designated as A through G, and the varied conditions are designated as a through g. The evaluation is concerned with identifying respective variations in the final test result due to the specific procedural differences, i.e., A-a, B-b, C-c, D-d, E-e, F-f, and G-g. Each of the eight trials consists of a single analysis conducted using eight respective aliquots of a single test material. The final test results are indicated as s, t, u, v, w, x, y, and z.

Table 1. Experimental Design for a Seven Variable Ruggedness Test*

Experiment Number	Factor Level	Analysis Result
1	A BCDEFG	s
2	ABcDefg	t
3	AbCDEfg	u
4	AbcdeFG	v
5	aBCdeFg	w
6	aBcdEFG	x
7	abCDEfG	y
8	abcDEFg	z

*Based on W. J. Youden, 1969, *The Collaborative Test*, p. 151-158, in *Precision Measurement and Calibration*, H. H. Ku, Editor, U.S. Department of Commerce, National Bureau of Standards, 436 pp.

The average of $A = (s + t + u + v)/4$, compared with the average of $a = (w + x + y + z)/4$, can serve as a rapid means of assessing the effect of changing variable A to a. Since each of the two groups of four determinations contain the other six variables, twice at the upper case level and twice at the lower case level, the effect of these variables (if present), tend to cancel out, leaving only the effect of changing variable A to a. The relative effect of the other variables can also be estimated by examining the following averages:

$$B = \frac{(s + t + w + x)}{4} \quad b = \frac{(u + v + y + z)}{4}$$

$$C = \frac{(s + u + w + y)}{4} \quad c = \frac{(t + v + x + z)}{4}$$

$$D = \frac{(s + t + y + z)}{4} \quad d = \frac{(u + v + w + x)}{4}$$

$$E = \frac{(s + u + x + z)}{4} \quad e = \frac{(t + v + w + y)}{4}$$

$$F = \frac{(s + v + w + z)}{4} \quad f = \frac{(t + u + x + y)}{4}$$

$$G = \frac{(s + v + x + y)}{4} \quad g = \frac{(t + u + w + z)}{4}$$

After tabulating the above averages, the differences between each respective variable would be computed, e.g.,

$$A - a = \frac{(s + t + u + v)}{4} - \frac{(w + x + y + z)}{4}$$

Examination of these respective differences enables the evaluating laboratory to assess which variables are probably effecting the test result. As stated above, most of the modest procedural alterations should have little or no effect on the result. Considerable information can be gained by merely comparing these differences (A-a, B-b, C-c, D-d, E-e, F-f, and G-g). The evaluating laboratory may wish to conduct multiple tests (repeat analyses) for each variable combination (e.g., 10 successive independent analyses for each of the 8 combinations) depending on the sponsor's requirements. Under these conditions, if any of the respective differences between averages are greater than two times the standard deviation for each variable (experiment), the testing laboratory would have another indication that the particular variable is effecting the test result. After discussing the results with the sponsor, additional studies might be planned to define the limits of acceptable variation for a particular critical test procedure.

The number of variables to select should be discussed with the sponsor prior to the test. However, reference tables for designing ruggedness tests are available when certain numbers of procedural variables are selected. Having completed the ruggedness evaluation, the subsequent phases of the single laboratory test (precision, method sensitivity, etc.) can then be conducted using a revised method procedure.

Method Precision

The only requirement for the precision test is to conduct 10 separate determinations on the same sample (preferably using a reference material). Each separate

determination must represent a valid test response as required by the particular method protocol. It is also recommended that the separate precision determinations be conducted on alternate days, i.e., an interval of at least one day between the completion of one analysis and the start of the next. The resulting data can be expressed either as a standard deviation, a standard error, or as a coefficient of variation.

Method Accuracy

To determine a method's single laboratory capability for accuracy (and for systematic error), the testing laboratory must have both a standard reference material and a known method response (true response) to this reference material. When a method calls for analysis of biological tissue or biological fluid, there will usually be a standard reference material available to the testing laboratory, e.g., samples with certified compound concentrations or perhaps with certified enzymatic activity levels. In these instances, the method's single laboratory capability for accuracy can be assessed by determining the differences between the observed single laboratory result, using the reference sample, and the known true value. Ten separate determinations (10 valid responses) should be conducted using a single concentration of the reference sample material. Each of the 10 determinations must represent a valid test response as directed in the particular method protocol. The method protocol presents whatever requirements are necessary for replicates, blank samples, etc., in order to provide a valid response. The Student t-Test would be used to determine the significance of the difference between the observed single laboratory test result and the known true value.

If the method being single laboratory tested is a toxicity test or perhaps an algal assay test (population stimulation or inhibition), there will usually be no "true response" available for a reference material and hence the method's single laboratory capability for accuracy (or for systematic error) cannot really be determined. Under these conditions, the testing laboratory should first select a reference material and then determine an average test response for a single concentration of the reference sample. When determining the average test response, it is recommended that 20 independent determinations (valid test response as indicated by method protocol)

be conducted. While the literature data base may provide valuable background information for a method's average response to various sample types, the single laboratory testing group should still conduct these determinations (to acquire an average test response to a reference material) using the ruggedness tested/revised protocol.

Systematic Error

A method's capability for (minimizing) systematic error can be considered as a part of the method's capability for accuracy. If a true response (known value) is not available for a reference material, the single laboratory test will not be able to acquire data on the method's systematic error. The testing laboratory should probably remind the sponsor that, under these conditions, no data will be provided for this phase of the single laboratory test. Comparison of test data with the results of a reference method can be used in certain situations, but for the purposes of this program, the use of reference methods is not considered as part of single laboratory testing. It should also be remembered that it is the bias of the method, not the bias of the laboratory that is being addressed. Single laboratory testing does not really address laboratory bias even though it will obviously affect test results.

In the case of a bioanalytical method for which a standard reference material is available, the testing laboratory should estimate a method's systematic error by using various dilutions of the reference sample. A single concentration of the standard reference material would be aliquoted into at least three equal samples. Two of these aliquots would then be diluted to different total volumes, thus creating three different sample sizes. Ten respective independent analyses would then be conducted on each of the three sample groups. For some methods it might be preferable to make the three groups all have the same sample size, i.e., different total amount of analyte. Results from each group (results from two groups would first be corrected by the respective dilution factor) would then be compared (Student t-Test) with the true value for the reference material. The consistently present systematic error should be noted in each of these three test groups. Additional concentrations/dilutions as well as additional independent determinations would probably be beneficial for the single laboratory assessment of systematic error

(depending on the sponsor's requirements).

Method Sensitivity

For purposes of a single laboratory test, a method's sensitivity is defined as the method's capability to detect (or distinguish between) small changes in sample concentration, i.e., concentration of analyte. A chemically characterized reference material should be used as sample material during sensitivity testing.

Assume concentration A (Figure 2) had been selected as the concentration of sample material used previously in the method precision test. For the sensitivity test, the laboratory would select one concentration greater than (C, Figure 2), and one concentration less than (B, Figure 2) the concentration used during the precision test. These concentrations should probably be equally distant from the precision test concentration (A, Figure 2). The laboratory should conduct 10 independent analyses for each new concentration (i.e., 10 separate valid responses acquired by following the method protocol). Assuming that the procedure can distinguish between A and C, and between A and B, then the test laboratory will reduce the concentration interval by one-half (to C' and B', Figure 2). If the method is still capable of distinguishing between A and C', and between A and B', then the test laboratory will again reduce the sample concentration interval by one-half (to C'' and B''). The 10

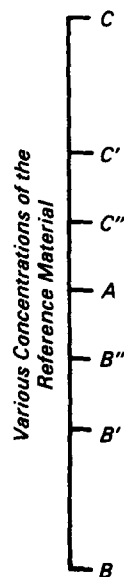


Figure 2. Example of different reference material concentrations used in sensitivity testing.

independent analyses on these last respective concentrations (C" and B", Figure 2) will complete the sensitivity test even if the method can distinguish between A and C", and between A and B". The sponsor can indicate if additional information on the method's sensitivity is required and, if so, direct the test laboratory to continue this process or to repeat the process using a different reference material.

A relatively poor method capability for sensitivity does not necessarily imply limited method usefulness or that the method would be an unlikely candidate for collaborative testing. The intended purpose of the respective method must always be considered when reviewing single laboratory test data.

Limits of Reliable Measurement

In determining a method's limits of reliable measurement, the single laboratory test data may simply verify that the method capabilities for sensitivity, precision, and accuracy (if applicable) do not deteriorate at the upper and lower extremes of the detection range. The same reference material should be used for these tests as was originally used during the method precision and method sensitivity evaluations.

The single laboratory test is not required to establish an upper and lower detection limit. It is assumed that a sufficient literature data base for the method exists to grossly estimate an upper and lower limit of detection using the particular sample material. The evaluating laboratory should initially select two concentrations of the sample material. One of these concentrations will be near the upper extreme of the method's detection range and the other concentration will be near the lower extreme of the detection range. Ten analyses would be conducted on each concentration to provide precision data (expressed as a coefficient of variation). The coefficient of variation will frequently show a dramatic increase at the extreme limits of detection and, therefore, precision data provide a distinct indication of the limits of reliable measurement. Two additional concentrations, one at each extreme of the estimated response range should then be selected in order to conduct the sensitivity determinations. Thus, two sample concentrations at each extreme of the estimated response range can be compared (in terms of precision and sensitivity) with the previously acquired data. If the method is a bioanalytical technique with an available true value

for the reference material, accuracy data would also be acquired.

When using this evaluation plan, a true limit of reliable measurement may not actually be established. However, even under these conditions, data would still be available to indicate that the technique was, or was not, capable of providing reliable measurements at the extreme concentrations selected. Additional sample concentrations, or additional test substances, can also be selected based on the sponsor's needs.

Conclusions

The single laboratory test data and the revised (ruggedness tested) method protocol will ultimately be used as part of the basis for deciding whether or not to proceed with collaborative testing. If the technique is selected, different laboratories will analyze aliquots of the same sample material (strictly following the method protocol) in order to validate the method's performance. Data from the collaborative test will be used to determine the reproducibility (overall between-laboratory variability) that can be expected when the procedure is used by different qualified laboratories.

William D. McKenzie and Theodore A. Olsson III, are with Bioassay Systems Corporation, Woburn, MA 01801.

William W. Sutton is the EPA Project Officer (see below).

The complete report, entitled "Guidelines for Conducting Single Laboratory Evaluations of Biological Methods," (Order No. PB 84-124 841; Cost: \$10.00, subject to change) will be available only from:

*National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650*

The EPA Project Officer can be contacted at:

*Environmental Monitoring Systems Laboratory
U.S. Environmental Protection Agency
P.O. Box 15027
Las Vegas, NV 89114*

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Official Business
Penalty for Private Use \$300

PS 0000329
U S ENVIR PROTECTION AGENCY
REGION 5 LIBRARY
230 S DEARBORN STREET
CHICAGO IL 60604

