AOAC INTERNATIONAL
Presidential Task Force on
Best Practices for Microbiological Methodology
US FDA Contract #223-01-2464, Modification #12

Executive Summary
Statistics Working Group

Objective: The objective of the STWG has been to review current practices and alternative approaches for validation of microbiological test methods (including growth-related and chemical tests), and make recommendations based on the past experiences and ongoing activities of the group members. This was to include statistical methods for analyzing data from validation studies. At the outset of this project it was acknowledged by the BPMM Steering Committee and by the STWG that the project provided neither the time nor the resources to fully validate all the recommendations of the group. Some recommendations, such as the use of LOD_{50} for qualitative methods, may require further development, before being widely used.

Determining method performance: Performance standards should be based on criteria based on fitness for the intended use, including public health needs. In general, statistical methods should be used to assist in setting realistic performance standards. These methods should be based on control of Type I and Type II error, which implies the determination of levels of unsatisfactory performance that must be detected (with stated probability) and controlled. It also implies use of appropriately determined sample sizes to meet the stated goals relative to stated α and ß. This approach would be a change from current practices in which studies are accepted on the basis of standard designs for number of laboratories, materials, and replicates, and standard criteria for suitability of the summary statistics. The design specifications and resulting reliability estimates should form the basis of applicability statements for test and measurement methods. (Ref 1, 9, 10, 13, 14) (*Task 2: What are the scientific/statistical bases for developing performance standards against which the validation of methods should be based*?).

The committee supports the use of appropriate international consensus standards. For consensus standards that are currently under development, the STWG recommends active participation in the development and/or validation of the standards. In general, the STWG acknowledges the value of rigorous consensus processes and international harmonization of method validation procedures. Specific approved international consensus standards include the following:

a) ISO 16140 Microbiology of food and animal feeding stuffs – Protocol for the validation of alternative methods
b) ISO 5725 Series: Accuracy (trueness and precision) of measurement methods and results.
c) ISO 11843 Series: Capability of Detection
d) CLSI/NCCLS EP17-A: Limits of Detection and Limits and Quantitation for Quantitative Measurement Procedures.

Standards under development include ISO draft Technical Specification 19036: Microbiology of food and animal feeding stuffs – Guide on estimation of measurement uncertainty for quantitative determinations.

Statistical procedures recommended in ISO 16140 are appropriate for "alternative methods" where there is an accepted reference method, but many of the procedures can also be used where there is no reference method. This document recommends use of robust statistical procedures that do not necessarily assume a normal distribution and are not so severely affected by extremely large or small outlier results that can be misleading with more conventional procedures. It also recommends against the removal of outliers from collaborative studies, except for assignable causes. The STWG fully agrees with these recommendations.

The committee strongly urges caution in applying the concept of "false negative" and "false positive" results because of the difficulty of confirming all positives and negatives, and the likelihood of misinterpretation. Alternative confirmation procedures should be considered, such as nucleic acid testing. Any estimates of "sensitivity" for low level samples should be corrected using appropriate statistical methods, such as adjustments for expected true negatives predicted with the Poisson distribution. Protocols should continue to include the appropriate Chi Square test based on whether or not samples are paired. (Ref 1-5, 14) (*Task 10. What are the appropriate statistical tools to be used for interpretation of validation studies*?)

Predictor and response variables important to the study design as well as for validating methods must be discussed and accepted by all subcommittees and the Steering Committee after review of all reports. Initial considerations should include variables that have been identified in the reports from other task groups. (Ref 15) (*Task 11: What are the test variables (e.g., number of strains, foods, inoculum levels) that should be considered for each of the factors listed in Task 8?*)

Estimating uncertainty: Uncertainty in measurements using quantitative procedures is best estimated following an all-inclusive, or "top down" approach. This approach does not attempt to estimate all components of uncertainty separately and it does not require a detailed mathematical model of how those components are combined. This approach is in contrast to a "bottom up" approach, which provides an estimate of the uncertainty of the method rather than the measurement and requires estimation and combination of variances at all stages of an analysis. This cannot be done routinely, however, so standard, or assumed, variances are used which aligns the combined estimate to the basic method rather than the analytical result. The "bottom up" approach is likely to underestimate uncertainty due to sources of uncertainty that are not considered. By contrast, the "top down" approach makes no attempt to set generic estimates of uncertainty for specific test methods and rightly aligns the estimate of measurement uncertainty with a specific analysis (or set of analyses). The "top down" approach is consistent with the Guide to the Expression of Uncertainty in Measurement (GUM) principles that allow combination of sources of uncertainty that are difficult to estimate individually. Comprehensive estimates of uncertainty can be obtained from collaborative studies, from carefully designed validation studies, or in some cases from routine quality monitoring data.

For qualitative methods, measurement uncertainty for the result cannot be expressed directly – instead, the measurement uncertainty relates to the probability of reporting an incorrect result. This can be estimated with false negative and false positive rates, for those methods with confirmation procedures (Ref 7). For some measurement procedures, uncertainty can be expressed as the standard error of a limit value estimation e.g. the LOD_{50}, as estimated by the Spearman-Karber or some alternative method (Ref 11,15,16). This procedure estimates uncertainty where it is most important, which is at the border of the determination of "present" or "absent" (that is, in the area of the detection limit). The work of ISO Technical Committee 34, Subcommittee 9 is not yet completed, so the STWG recommends active participation in the efforts of this subcommittee. (Ref 5) (*Task 6. What are the effective means for articulating the uncertainty associated with microbiological methods?*)

Limit of Detection: The detection limit for qualitative tests is best described as the "LOD_{50}", or number of organisms per gram of sample at which 50% of the tests are positive. This is determined with a nonparametric (distribution free) version of probit analysis, and an experimental study using at least 4 dilutions in which at least two of the dilutions have "fractional positives" in order to better estimate the LOD_{50} and perhaps allow for estimates of other percentiles, such as the LOD_{90} (number of organisms per gram of sample where 90% of results are positive).This procedure also assumes that one dilution level has 0% positive results and one dilution level has nearly 100% positive results (allowing for measurement error in the test laboratories). (Ref 12,13, 17, 18).

For quantitative methods, the committee recommends use of the ISO 16140 procedure, which presents limits of detection and quantification as functions of the variability of blank (or very low) samples. The committee recognizes, however, that alternative procedures exist that should be investigated, such as the ISO 11843 Series on capability of detection, or the nonparametric analog of that procedure, as described in the CLSI document EP17-A on Limits of Detection and Quantitation. These procedures recognize the importance of Type I and Type II errors, and that variances of signals from truly negative and truly positive samples can be different (Ref 1, 3, 4). There are related strategies for designing experiments to use the ISO/CLSI approach (*Task 4: What are scientific/statistical bases for determining the lower limit of detection for microbiological methods? How is the lower limit of detection validated during the validation of a method? How is the relative performance of a method determined as the lower limit of detection is approached and what is the best way of characterizing this performance?*)

Topics for further research

In the course of this review, the STWG identified several areas where further research was needed, or a more comprehensive review of the documents developed for this study. The areas of further review include the following (Ref 19):

1. Further development of procedures for describing the Limit of Detection for quantitative methods.
2. Further development of recommendations for use of the generalized Spearman-Karber method for estimating the LOD_{50} for qualitative methods.

3. Evaluation of alternative approaches to the Spearman-Karber method e.g. Logit, Probit and other statistical procedures currently under investigation by the ISO TC34/SC9/SWG.
4. Investigation of the effectiveness of current AOAC Official Methods for Single Laboratory Validation (SLV) procedures, Multiple laboratory Validation procedures (MLV) and harmonized Collaborative Validation studies (HCV), relative to the recommendations concerning the design of verification studies.
5. Use of existing AOAC study data to evaluate the alternative statistical methods proposed.
6. Use of existing AOAC data for assisting in design issues for future validation studies. This could include proper consideration of Type II error in addition to Type I error, and should develop a structured approach for making decisions based on the data.

References

1. ISO 16140: Microbiology of food and animal feeding stuffs – Protocol for the validation of alternative methods

2. ISO 5725 Series (Parts 1-6): Accuracy (trueness and precision) of measurement methods and results.

3. ISO 11843 Series (Parts 1-4): Capability of Detection

4. NCCLS EP17-A: Protocols for the Determination of Limits of Detection and Limits of Quantitation. Clinical and Laboratory Standards Institute, Wayne, PA., 2004.

5. ISO TS19036: Microbiology of food and animal feeding stuffs – Guide on estimation of measurement uncertainty for quantitative determinations (draft)

6. Feldsine, Abeyta, Andrews: Journal of AOACI Vol 85, No.5, 2002: AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Qualitative and Quantitative Food Microbiological Official Methods of Analysis

7. Wilrich, P-T (2005b) *The determination of precision of measurement methods with qualitative results by interlaboratory experiments.* Discussion paper for the ISO TC34 SC9 Statistics Group Meeting, Paris, April 2005

8. Linnet K, Kondratovich M. Partly Nonparametric Approach for Determining the Limit of Detection. *Clinical Chemistry* 50(4): 732-740; 2004

9. McClure F D, Lee J K. Sample Sizes Needed for Specified Margins of Relative Error in the Estimates of the Repeatability and Reproducibility Standard Deviations. *JAOACI*, in press, 2005.

10. McClure F D. Design and Analysis of Qualitative Collaborative Studies: Minimum Collaborative Program. *JAOACI*, 73(6): 953-960. 1990

11. Miller J, Ulrich R. On the analysis of psychometric functions: The Spearman-Karber method. Perception & Psychophysics 63(8): 1399-1420. 200

12. Paulson, D. S. (2003). *Applied Statistical Designs for the Research.* New York: Marcel-Dekker.

13. Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for Experiments: An Introduction to Design, Data Analysis, and Model Building*. New York: John Wiley & Sons.

14. STWG report: Task2&10.doc

15. STWG report: Study Variables.doc 022505

16. STWG report: Combined report BPMM SWG_Uncertainty consolidated_rev 1.pdf

17. STWG report: LOD50% 20050721.doc

18. STWG Excel worksheet: LOD50% Spearman-Karber.xls

19 STWG report: Recommendations for Future Research.doc