

Recommendations for Future Research

In the course of this project, the working groups identified various areas where further research was needed, or a more comprehensive review of the documents developed for this project. The areas of further research include the following:

Validation Study Design and Statistical Analysis

1. Evaluate statistical approaches for qualitative and quantitative methods including: (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-Kärber method for estimating the LOD₅₀ for qualitative methods; and (3) evaluation of alternative approaches to the Spearman-Kärber method e.g. Logit, Probit and other statistical procedures currently under investigation by the ISO TC34/SC9/SWG. Active participation in the ISO committee discussions is encouraged. A comparison of the generalized Spearman-Kärber method to logit and probit analyses will be undertaken. It is important to determine what issues are important for an appropriate statistical method. The most appropriate method will depend on the study design and the assumptions of the statistical method. The consensus opinion of the task force is that more than two levels of contamination are needed for an LOD₅₀ analysis.
2. Use of existing AOAC data for assisting in design issues and choice of statistical methodology 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. Non-AOAC data (e.g., clinical data) should be considered as well since AOAC data has design limitations. Effort should be made to identify individuals outside of AOAC (e.g., through FDA's Center for Devices and Radiological Health) that may be doing innovative work in this area.
3. There is a concern about the statistical comparisons used that are usually weighted to the prevention of Type I error (stating a difference exists when one does not) over Type II error (stating no difference exists when one does). The statistical hypothesis in testing evaluates if there is a difference between two groups (two-tail test), one group is larger than another (upper-tail test), or one group is less than another (lower-tail test). There is no test for equivalence in significance-testing, yet that is often the major focus of AOAC testing. That is, that Lab A and Lab B results are not different. Perhaps a remedy as simple as increasing Type I error levels ($\alpha > 0.05$) and reducing Type II error levels ($\beta > 0.20$) would be useful.
4. The project timeline did not allow full discussion of the differences between Single Laboratory Validation (SLV), Multi-Laboratory Validation (MLV) and Collaborative Validation (HCV) in terms of the statistical confidence related to method performance and the effects of changes in number of samples, levels, analysts, labs, etc. The task force recommends that further work be done to elaborate these differences. In addition, the task force recommends investigation of the effectiveness of current AOAC Official Methods for Single Laboratory Validation (SLV) procedures, Multi-Laboratory Validation procedures (MLV) and harmonized Collaborative Validation studies (HCV), relative to the recommendations concerning the design of verification studies. Develop general guidelines for method validation protocols relative to

different applications (fit for purpose) and how these might be modified depending on the level of confidence required (how much uncertainty can be tolerated). Ideally, the guidelines would be flexible to allow for practical considerations, such as allowing an increased number of samples per lab to compensate for fewer labs in a study.

Confirmation of Results

5. As new innovative technologies are exploited for food pathogen detection, the gap between the LOD₅₀ of the alternative method and the LOD₅₀ of the reference cultural method (“gold standard”) is expected to widen. This can result in presumptive positive results for the alternative method that cannot be confirmed culturally. In addition, as new pathogens emerge, gold standard methods may not exist. Finally, we must consider validation of methods to detect organisms that are viable but not culturable or not easily culturable, such as mycobacteria and viruses. In these cases, it is necessary to develop new approaches. Several approaches to be evaluated include:

- a. Quantification of the confidence in the presumptive positive results in the method validation study. One proposal is to determine the incidence of positive results for a given uninoculated food matrix. Assuming the incidence is low, some statistical confidence is gained that presumptive positive results obtained in a validation study of the inoculated food matrix reflect the presence of the target analyte. The task force recommends that this concept, a modification of the clinical positive predictive values and negative predictive values, should be further discussed and developed.
- b. Confirmation using methods based on technology distinct from the alternate method being validated. For example, a PCR test may be used to validate an immunoassay result. RNA targets could be used to ensure detection of live cells.
- c. Confirmation based on detection of multiple analyte markers. In the absence of a suitable confirmatory test (high sensitivity and specificity), multiple tests could be used for confirmation and the level of agreement between these tests specified in order to achieve a true result.
- d. Comparison of fractionally positive results to the theoretical Poisson and/or other distributions.

Preparation of Inoculated Samples

6. Test the dilution to extinction method for preparing samples for validation studies. Dilution to extinction is essentially an MPN method based on probability with an assumed distribution. The method should be further developed and tested to determine the number of levels and number of samples per level that should be tested and what level of recovery and statistical confidence can be achieved. Further, experimentally determine if these techniques can reliably calculate the level of target organism at the limit of detection.

Method Verification

7. The task force recommends that a laboratory intending to adopt a method that has been validated, verify the performance of that method in their laboratory. Future work is required to

develop procedures for verification of all validated methods, so that the method description will include a minimal verification procedure.

Ruggedness Testing

8. Ideally, every method validation would be initiated with a single lab validation to assess a variety of method performance parameters. If a method is intended to be validated through a multi-lab or collaborative study, this would occur after successful completion of the single lab validation. The task force recommends that some ruggedness testing of the method be performed as part of the single lab validation study. Critical parameters to be tested in the SLV ruggedness studies depend on the type of method under consideration. Future work will include development of guidelines for choosing parameters and designing ruggedness studies.

Note: Ultimately, the goal of the BPMM is to produce new proposed AOAC guidelines for validation, verification, modification and extension of microbiological methods.