AOAC INTERNATIONAL
Presidential Task Force on
Best Practices for Microbiological Methodology

US FDA Contract #223-01-2464, Modification #12

Executive Summary
Detection Limits Working Group (DLWG)

EXECUTIVE SUMMARY

For the Limits of Detection (LOD) Working Group, five novel recommendations are offered for consideration:

1. In addition to more precisely defining how preparation and stabilization of inoculated samples should occur, the LOD Working Group has proposed an alternative challenge procedure, Dilution to Extinction (DTE). This methodology has an advantage in that it does not necessarily require the simultaneous analysis of the matrix by a cultural reference method. In certain instances, for example, where the alternative method may be more sensitive than the reference method DTE may have advantages in that the consideration of false positives and false negatives is eliminated. The calculation of the inocula levels of the target analyte organism on the day of initiation of analyses is done only in the organizing laboratory and not in the collaborating laboratories. The performance calculations may be applied to both method comparison and collaborative studies.

2. Alternatives are presented for novel approaches to be taken when the proposed method is suspected of being more sensitive than a reference method.

3. In addition to the considerations in (2) (above), specific consideration is given to molecular based methods validation. While the LOD Working Group is not specifically endorsing a position that molecular based methods are superior or preferred to assays based on other detection technologies, there was a consensus in the Group that confirmation of molecular based assays using traditional cultural procedures may be problematic. Specifically, when molecular assays have improved sensitivity and/or a better limit of detection compared to cultural methods, this may result in the incorrect perception of higher levels of false positive results. On the other hand, molecular assays may be more prone to matrix associated inhibition, leading to reduced assay specificity and a higher incidence of false negative results. Certainly, for routine analysis of samples, alternative methods based on detection technology other than traditional culture or molecular are suitable for many applications. The impact of these assay designs on assay sensitivity and specificity must be considered when establishing the limits of detection for these alternative molecular assays.
4. With regard to the statistical validity of low level contamination, a supplemental statistical treatment is presented. This technique, the LOD_{50}, is not suggested as a replacement for existing tests for significant differences, e.g. Chi Square, but

rather offered as a data treatment that could provide some measurement of the potential variability associated with low level contamination. This may be particularly relevant given that the LOD Working Group believes that low level contamination of matrices, including levels providing high and low fractional recovery and levels at or near the endpoint of recovery, is the preferred method for defining assay performance.

5. With regard to quantitative methods validation, the LOD Working Group supports methodologies presented in ISO 16140. It should be noted, however, that when the only available reference method is based on 3 tube MPN analysis, validation may best be performed using the DTE approach. This alternative may be preferred because of the lack of precision associated with MPN measurements.

Objective 4:

What are the 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?

Summary of Recommendations

It is the opinion of this Working Group that achieving an endpoint of microorganism recovery for the alternative method is the most reliable means for defining method performance and equivalence to a reference procedure. The reference procedure chosen may be a traditional culture procedure or a well-defined rapid method. Endpoint analysis may be applied to detection of bacteria, fungi, viruses and toxigenic compounds, assuming that a detection procedure and a reference procedure are available. In the case of viruses and toxins, enrichment procedures do not apply as these materials do not replicate in culture media.

Even though the homogeneity of the sample cannot be assumed, protocols can be developed to minimize this impact. Furthermore, specific protocols can be designed for different categories of food matrices (high moisture food and low moisture food). The Limits of Detection Work Group has divided the consideration of this topic into:

i. Inoculum preparation and uniform contamination of the matrix of interest
ii. Confirmation of results, particularly when the alternative method may be more sensitive than the reference method
iii. Analyzing and presenting summarized data

Discussion for sections i) and ii) pertain to contract question C1-4 and follow. Discussion for section iii) pertains to contract questions C1-7 and is presented in the next section.

When validating quantitative methods, it is preferable to inoculate the food matrix at three contamination levels. These levels should occur at approximately one logarithm increments within the range expected in the food matrix.

Preparation of Inoculum and Artificial Contamination

To determine the detection limit of a qualitative method, the method should be tested on appropriate food samples naturally contaminated or inoculated with microorganisms above and below the anticipated detection limit. For quantitative methods, a minimum of three inoculation levels should be prepared within the expected range of application for the method and the food matrix.

Inoculated food samples for validation of methods should be prepared according to the standard protocols used for AOAC precollaborative and collaborative studies (see article by Andrews, W. A.: J Assoc. Off. Anal. Chem. 1987 Nov-Dec: 70(6):931-6).

Recently, reference materials and certified reference materials have become available that may be more precise than inoculation doses prepared by traditional dilution methods, and may also be utilized when the appropriate levels of target organisms can be obtained (see below).

For viruses and toxins, a high level concentration titered by an accepted reference procedure is prepared in the food matrix, stabilized and then diluted in the food matrix as described by Andrews (1987) for bacteria and fungi.

Precise Reference Materials and Certified Reference Materials

Recent developments have made it possible to produce samples that contain precise numbers of microorganisms for use as quantified standards in microbiological analyses. Flow cytometry has been adapted as the platform to analyze and sort cells, and dispense precise numbers of the cells in liquid or freeze dried forms. These precise samples can be used as quantitative Reference Materials.

The International Organization for Standardization has an accreditation system for reference materials known as ISO 34 that enables the production of Certified Reference Materials (CRMs). These CRMs are supplied with a certificate that specifies the amount of bacteria and the variability.

General Protocols for Limit of Detection Studies

Determine level of viable target organism(s) in the “seed.” Normally this has been accomplished by MPN procedure with the reference method. However, if the target level of organism present in the “seed” is higher than the background flora, this may be accomplished by non-selective plating or MPN procedures, followed by confirmation of colonies or growth that is typical of the target organism. If a certified reference material

is used, the data provided on the QC certificate should be used for determination of inoculum level.

Once the level of target organisms present in the seed is determined, the seed can be used to prepare method validation samples by either dry or wet dilution methods. This can be done in the organizing laboratory only, or can be done in multiple laboratories by splitting the seed matrix and sending portions to additional laboratories.

1. Wet Dilution Method

    a. Prepare an enrichment of the seed matrix.
    b. Prepare 90 ml enrichments of uninoculated product matrix (same product used to prepare seed) one part matrix to 9 parts enrichment broth.
    c. Set up dilutions of the seed by adding 10 ml of seed homogenate into 90 ml of uninoculated enrichment, thus producing 10-fold serial dilutions of the seed with the same ratio of matrix to broth.
    d. Continue to serial dilute into uninoculated enrichments, until the expected lower limit of detection is exceeded.
    e. Set up multiple enrichments for each dilution for each method being evaluated.

2. Dry Dilution Method

    a. Prepare serial dilutions of seed culture by blending/mixing into the uninoculated portions.
    b. Analyze multiple samples (minimum of 5 per dilution) of each dilution by the reference method and method being evaluated.
    c. Analyze results and determine MPN based on the number of positive and negative tubes at the highest usable dilution according to the MPN rules.

Limit of Detection Methodology to Determine Low Level Sensitivity: DTE as an Alternative to Use of Reference Culture Method Comparisons

Certain situations may arise wherein a direct comparison to the reference culture method may not be the best microbiological practice. One such situation is where preliminary data indicate that the reference culture method may not be as sensitive as the alternative method. A second potentially problematic situation occurs when the alternative method and the reference method employ different primary enrichment media. In such a situation in which there is also a need to reach a fractional endpoint to determine the limit of detection, the incidence of positives and negatives would be expected to be random, assuming that proper homogenization of the matrix was accomplished. In this situation the performance data, as expressed as false positive and false negative results, will be very high for both methods, thereby rendering methods performance statistics of questionable utility to the analyst. Presently, the results are reviewed subjectively for "reasonableness." It is possible to employ an alternative study design to eliminate this anomaly in the data. This approach may be termed Dilution to Extinction. In this

approach, a concentrated inoculum is stabilized in a small amount of the matrix of interest. Subsequent dilutions are made in the matrix itself until the recovery by the alternative method becomes, at first, fractional and then progresses to all negative results. This approach also lends itself to validation of virus and toxin measurement assays where comparison to reference methods at low contamination levels may be problematic.

Using this technique, the sample is assayed by the alternative method but there is no unpaired companion sample run with the reference culture method. Instead the enriched sample is confirmed using the appropriate isolation and confirmation technique defined in the reference method. By proceeding in this manner there can be no disagreement in the results that is attributed to only the variability in uniformity of inoculum dispersion at the limit of detection. In such a scheme the sample size per level may be reduced from 5 replicates, but at least 3 replicates and as many as 5 levels may be run to reach fractional and finally completely negative determinations. The number of levels included in the methods comparison study may be increased from the current 2 levels to 5 and the number of replicates per level reduced to 4. For the collaborative study, 2 or 3 levels plus uninoculated controls should be run using 4 samples per level. Data generated by this protocol design are suitable for analysis by the LOD_{50} method, which is described elsewhere in this document.

For methods chosen to be validated using the Single Laboratory Validation (SLV) or a Multi-Laboratory Validation (MLV) involving 2 or more laboratories but not a full Harmonized Collaborative Validation (HCV), the use of the LOD_{50} analysis is an appropriate statistical methodology. This technique may be used in concert with the DTE methodology previously described, but only when an appropriate reference method is not available or different primary enrichment broths are specified. For HCV methods, the DTE approach may be employed in the methods comparison study. It may also be used in the full collaborative study, but only when an appropriate reference method is not available. For the collaborative study the LOD_{50} method may be used to evaluate the data.

Validation against a Less Sensitive Reference Method: General Approaches

The need for highly specific and sensitive diagnostic methods for pathogens and toxins has always been critical to food safety, public health, and national security. For the past 50 years, the gold standard methods used to detect bacterial or viral pathogens has been culture-based analyses. Culture-based methods allow for non-directed analyses (i.e., can isolate/detect multiple pathogens from a single plate), are cost effective, have a true limit of detection (LOD) of one viable/culturable organism per sample size, and have been developed so as to be performed by a variety of well-trained professionals, including medical technologists, sanitarians, bacteriologists, and virologists who perform the detection and identification of the infectious agents. With the emergence of alternative, technically more sophisticated diagnostic methods like the polymerase chain reaction (PCR), a system is needed to compare the results from the two distinct platforms to validate or confirm results.

Where applicable, the use of an established reference method is recognized as a preferred means to confirm the results of an alternative method. More recently, however, it has become increasingly apparent that, in some circumstances, the alternative method may be more sensitive that the traditional reference method(s) which are available for confirmation. In such instances it is the opinion of the Working Group that it is appropriate to employ alternative methodology to resolve discrepant results. Such additional efforts would only be required when there was a difference manifest between the alternative and reference methods for an individual sample. Possible approaches could include re-assay of discrepant samples by both methods to confirm the validity of the preliminary determinations, use of a third assay that is based on a different detection technology and for which the performance characteristics of that third method are known, or use of molecular markers if they exist that could confirm the presence of the microorganism in the growth medium. Another attractive alternative is the limit of detection validation presented above, as it eliminates the mandatory use of a reference method.

Matrix Inhibition of Molecular Methods

When developing molecular based methods, specific consideration must be given to the potential for matrix related inhibition which may lead to both invalid and/or false negative determinations. A series of matrix addition experiments is in order to define at what, if any level, the matrix from which the isolation is attempted may be capable of inhibiting the amplification reaction itself. Furthermore, it is strongly suggested that, when using molecular based methodology in a routine testing environment, the method must contain an appropriate internal control which will fail to amplify in the presence of a matrix interference event. When validating molecular based methods, specific attention should be paid to results obtained using the proper internal controls to validate that no matrix inhibition has occurred.

Objective 7:

How is the statistical basis of a validated method influenced if the homogeneity of the sample cannot be assumed, particularly at a very low CFU level? How does this influence the performance parameter of a method? How can samples be prepared to minimize this effect? Define the optimum procedures for sampling.

Summary of Recommendations

Procedures for Analysis of Data

The Working Group supports long established methodologies for statistical analysis of test results contained in the guidelines developed by AOAC and contained in ISO16140, with one clarification about the appropriate use of Chi Square statistics. Chi Square calculations should include confirmed positives from presumptive positive samples. In other words, a negative Test Method result, even if confirmed positive, remains classified

as a negative result, and is not considered as a Confirmed positive result when calculating the Chi Square.

It may also be useful to consider an alternate method of compiling data that would allow for the presentation of confidence intervals for the alternative method as an additional statistic. One such technique is proposed below.

Alternate technique: LOD_{50}

An alternate approach is the LOD_{50} Analysis. To augment and summarize the results of methods comparison and collaborative studies of qualitative microbiology methods, 50% detection endpoint values can be added to the result presentations. These can be calculated from the usual data obtained in such studies by the generalized Spearman-Karber method and would not require additional laboratory work. Some examples are presented. This procedure is also adaptable to experimental designs that differ from that of the traditional AOAC validation study. When analyzing data using LOD_{50} Analysis, the study designs for the methods comparison and the collaborative studies may differ. For example, combining the limit of detection approach with LOD_{50} Analysis, it is preferable to employ 4 or 5 levels of inoculation and 4 replicates per level for the methods comparison study. For the collaborative study, the existing study design of 2 levels plus uninoculated control with 6 replicates per level or, as an option, 3 levels of contamination plus controls with 4 replicates per level, may be preferred. The actual number of levels and number of replicates per level would be determined based on the fit-for-purpose concept and the level of confidence required for the intended use.

Introduction

When an analyte is at the level of 1 particle per sample, heterogeneous distribution of analyte, as described by the Poisson distribution equation, becomes significant. In microbiology, the particle is either a single autoreplicative organism or a sub-cellular particle (virion or naked nucleic acid) capable of being replicated *in vivo* or *in vitro*. In chemistry, the analyte particle is an atom or molecule, generally in homogeneous solution, and routine qualitative chemical analysis is not performed at the Poisson level.

The limit of detection for qualitative microbiological methods is theoretically 1 organism per analytical portion of a sample (or 0.04 cfu per g in the typical 25-g portion). The endpoint is not a sharp cut-off of the % positive samples versus concentration (MPN/g or cfu/g) curve due to the Poisson distribution effect. As a consequence, the limit of detection curve has a sigmoid-like shape so that at the concentration level of 1 organism per sample only about 63 % of tested samples will test positive. That is, at least about 36% of samples will be true negatives because of failure of incurred or artificial inoculation (spiking) due to the Poisson distribution effect. This effect kicks in significantly below about 3 cfu per g. The region of the curve at about 100% is of particular interest because a significant deficit in positive samples by a test method, relative to an ideal percent positive value of 100%, represents mainly false-negatives, that is a failure to detect positive samples. As the curve approaches 100% asymptotically this

region is difficult to define experimentally. Therefore, it is easier to work at the 50% region of the curve where the curve is steepest and close to linear. Thus the performance of a test method can be defined as the concentration (MPN or cfu per g ± confidence limits) at which 50 % positive samples are observed. This will, allowing for the confidence limits of the estimate, not be less than about 0.028 cfu per g with a 25-g sample analytical portion. If it is significantly larger, it means the method is performing less than ideally because as well as true negatives there are false negatives.

Fifty Percent Endpoints

Expressing the limit of detection as the concentration corresponding to 50% positive samples ± confidence limits (usually 95%) is simply a shorthand way of summarizing the performance results for a method. A single number with its limits is used to express the result for a given food matrix. The current tabulations of results more or less nicely compare test and control results statistically but do not clearly tell us how well and with what degree of confidence the methods approach the theoretically maximum possible performance parameter of 1 organism detected per 25 g sample. Neither do they clearly distinguish between true negatives and false negatives. Nor does the current way allow us to easily compare the detection limits for different food matrices and/or analyte strains

Fifty percent endpoint values make efficient use of all the data from control, low and high inoculation levels as well as providing confidence limits to make significance comparisons. In the current method of presenting results, attention is generally focused on the results from only the one of the inoculation levels used that gives the lowest apparent false positive rate.

Some typical and hypothetical examples are given in Table 3. They were generated with an Excel spreadsheet program <Anthony.Hitchins@cfsan.fda.gov>. Approximate endpoint estimates are generally still possible with unusual positive response data patterns that sometimes happen with the 3-inoculation level study design (Table 1) or when a sample with a single level of naturally incurred contamination is used.

As with the traditional treatment of the validation study results, the endpoint calculation is also dependent on the accuracy of the enumeration of the sample contamination. The confidence limits of the 3-tube MPN enumeration typically used are quite broad. This aspect will be addressed in a future document.

The method requires just an extra calculation with the conventional validation data so its adoption would not be a dramatic change. It is interesting to note that the current AOAC experimental design already allows for a gradual detection limit cutoff. Thus, two levels of inoculation are specified so as to try to ensure at least one set of detection results that are not all positive or negative.

Confidence limits will usually be broader, but still tolerable, for the endpoints from pre-collaborative studies than for those from collaborative studies, which have a greater

number of replicates. For example, compare the 5 replicate and 75 replicate confidence limit ranges calculated for proportionate positive responses in Table 3.

Table 3. Examples of the Application of the LOD_{50} Calculation to Qualitative Microbiology Detection Method Data

No. replicates / level (labs x reps / level)	No. positives at control, low and high inoculation levels (MPN/25 g)[a] <1[b] MPN	2.5 MPN	10 MPN	50% Endpoint (cfu/25g) and 95% confidence limit range[c]
A. Collaborative study type data				
40(10 x 4)	0	16	40	1.33 (1.05-1.65)
60 (15 x 4)	0	50	60	1.93 (1.75-2.10)
60 (15 x 4)	1[d]	25	58[e]	3.3 (2.90-3.75)
B. Pre-collaborative study type data				
5 (1 x 5)	0	2	5	3.15 (2.03-5.75)
4(1 x 4)	0	1	4	2.80 (0.78-10.32)
3 (1 x 3)	0	2	3	2.33 (1.28-4.23)

[a]Levels of MPN/25g determined on the day analysis initiated.

[b] The method requires a definite concentration value for a zero positives response. Therefore 1 MPN/25g is chosen for the controls (uninoculated samples) rather than inoculating replicated sample sets with inoculum so dilute as to virtually ensure zero positive-responses. This is fair as the proportion of positive responses approach zero asymptotically according to the Poisson equation.

[c] Divide by 25 to obtain the 50 % detection endpoint expressed as cfu/g of the 25 g sample.

[d] This value exemplifies natural or accidental background contamination at a very low level. The calculated mean background concentration per 25 g, $m = -\text{Ln } P_0 = -\text{Ln } (60-1)/60 = 0.016$ MPN/25 g. Since this is <<< 1, a value of zero positive was used for the control level. For a significant level of naturally incurred contamination either do not calculate a 50% endpoint value or estimate it by assuming appropriate concentrations for 0 and 100 % positive responses as illustrated else where in these footnotes.

[e] The calculation requires a 100 % response value. In this case, a 100% response was very conservatively assumed to be at a 100 MPN/25 g inoculum level.

The LOD_{50} method can be applied to quantitative methods when the Dilution to Extinction experimental protocol is used as described in the next section.

Illustration of the application of dilution-to-extinction to quantitative microbiology using enumeration method study data

The dilution to extinction method can be used to enumerate microbes in foods and other matrices. As a work in progress simulation, the natural micro-flora in five sub-samples taken from a homogenized food sample have been enumerated by dilution and plating of 0.1 ml amounts in duplicate on plate count agar. This simulation applies either to one laboratory enumerating 5 samples or to 5 laboratories enumerating one sample each. The colony counts obtained at the various dilutions are displayed in Table 4, Panel A. The calculated mean colony count and 95% confidence interval for the sample are presented in the second and third columns, respectively, of Panel D.

To apply the dilution to extinction method to this example, the actual colony counts for the 5 sub-samples presented in Panel A are transformed into presence and absence data in Panel B. To transform the data, the combined duplicated count for each subsample at a given dilution is designated a positive result if the combined colony count is > 0 and negative if it is 0. The number of positives per five sub-samples at each dilution is totaled (Panel B). The totals per dilution are processed by the Spearman-Kärber LOD_{50} calculation. The LOD_{50} result and its confidence interval represent the *limit of dilution* at which 50% of the diluted sub-samples are culture positive (Panel C). The reciprocal of the product of the LOD_{50} value, or each of its uncertainty limits, and the volume of dilution tested (0.2 ml in this example) multiplied by a constant m (m=0.69) give the dilution to extinction estimate of the mean count of the sub-samples (Panel D, columns 4 and 5). The constant m represents the mean count per test volume corresponding to 50 % negative (or positive) results. It is obtained from the Poisson relationship, $P_o = e^{-m}$, for the proportion of negative cultures. The data in Panel B can also be used to calculate a not significantly different 5-tube MPN result of 6500/g with a confidence interval of 1800 – 20,000. Thus the dilution to extinction method performs as well as the MPN. However, it may be technically superior, at least in some applications, because it is a plate as opposed to a broth culture method.

It can be seen from Panel D of Table 4 that the dilution-to-extinction and plate-count results are not significantly different. The dilution to extinction method was subjected to the experimental design of the plate count method in order to directly compare the two calculations. That design is not optimal for the dilution to extinction method but it can be modified appropriately for optimization.

Table 4. Enumeration by Dilution to Extinction using the 50% Extinction Value				
A. Raw data				
	Colony count (CFU/ 0.1ml of stated dilution)			
Sub-Sample	10^{-1} dilution	10^{-2} dilution	10^{-3} dilution	10^{-4} dilution
1	49;51	4;5	1; 0	0; 0
2	62;72	6;7	1;1	0; 0
3	40;43	4;4	0; 0	0; 0
4	62;65	6;7	1; 0	0; 0
5	31;40	3;4	0;2	0; 0
	B. Raw Data Processed for LOD_{50} Calculation			
	Positives (mean CFU > 0 per 0.2 ml) per 5 samples at stated dilution			
Sub-Samples	10^{-1} dilution	10^{-2} dilution	10^{-3} dilution	10^{-4} dilution
1-5	5	5	4	0
	C. LOD_{50} Result			
Sub-Samples	Dilution for 50% positive		95% confidence dilution interval	
1-5	5.013×10^{-4}		$(1.838 - 13.677) \times 10^{-4}$	
	D. Comparison of Counts by Plating and by LOD_{50}			
	Av. CFU/plate x $[1 / (0.1 \times 10^{-1})]$		0.69x1/[volume tested x LOD_{50}]#	
Sub-Samples	Av. CFU /g	±1.96SD CFU		
1-5	**5150**	2430 - 7816	**6900**	2518 - 18800

Application of LOD_{50} to existing methods

The LOD_{50} may best be applied when the study design includes several levels of inoculation as described above. In situations where previous studies exist which were conducted under protocols which required two levels of inoculation plus negative controls, this analysis technique may be applied retrospectively. The confidence intervals of the data may suffer somewhat from the reduced number of levels; however, valid interpretations may be drawn. This would be the situation with the many AOAC collaboratively studied microbiological Official Methods of Analysis. In these studies, generally two levels plus an uninoculated control level were run. The data from such studies may be retrospectively analyzed to calculate confidence intervals.