Proposed Use of a 50 % Limit of Detection Value in Defining Uncertainty Limits in the Validation of Presence-Absence Microbial Detection Methods

Background

A 50 % endpoint Limit of Detection (LOD_{50}) procedure can be used to calculate the absolute performance efficacies, and their associated uncertainties, of presence/absence methods for microbial detection in foods (3, *5*, *8*, *10*).

Validation of methods for microbial detection in foods or other matrices involves determining microbe recoveries. Recoveries are expressed qualitatively as presence-absence data, which are obtained from quantitative spiking experiments. Replicate samples of foods are spiked with the microbe of interest, generally at several concentration levels. Usually three different levels are used. However, only the data from the level which gives partial recovery are considered relevant. Such data are most reflective of a method's detection endpoint but a limit of detection is rarely estimated. The calculation discussed here maximizes the use of the data from such trials by using data from more than one spiking level to calculate an LOD_{50}.

Performance efficacies of new microbial detection methods are usually determined by comparison to recognized standard methods. This comparison is only strictly valid when common samples are used for the new and the standard methods. Then the methods are being compared at an equal microbial concentration. The situation is more complicated when comparisons involve a non-paired sample experimental design. Nevertheless, to a first approximation, comparative method validations always have the advantage of not really needing to determine the exact spiking concentration and thus virtually side-steps the fundamental problem of microbial enumeration variability at low concentrations. However, it is difficult to compare the results from different trials because the variability of the proportions of positive recoveries can be at least partly due to the technical difficulty of standardizing the spike levels from trial to trial. Also, sometimes only a single method may be validated so no intra-study comparison is possible.

The LOD_{50} method *normalizes* the results of such studies by estimating the spiking concentration (cfu/analytical portion size), which would correspond to 50 % recovery. Importantly, it also provides a measure of the uncertainties in terms of confidence intervals (at the 95% level) of the estimated LOD_{50}. A 50 % endpoint is used because the low concentration region of the recovery-concentration relationship is theoretically a sigmoid curve, it being governed by the Poisson distribution. In the case of *Listeria* methods, at least, recovery-concentration curves are clearly describable by the Poisson relationship (6). The confidence intervals of asymptotic region estimates are somewhat narrower than those of estimates in the mid-region. Nevertheless, the concentration corresponding to the midpoint of a sigmoid recovery curve can be more precisely determined than for a point in the one of the asymptotic regions tending toward either 0 or 100% recovery.

Methods for calculating an LOD_{50}

The LOD_{50} calculation could potentially employ one of several mathematical tools (Table 1). These are used to calculate the dose corresponding to a 50% response value (ID_{50} and LD_{50}) from the log-normal dose-response curve observed in an animal infection and mortality study. Thus in the LOD_{50} determination, the proportion of replicates at a given spiking level that is culture negative (nominally uninfected) is treated just as would be the proportion of uninfected or surviving animals at a given challenge dose. Conversely, a test culture positive result is analogous to an animal infection or death. These calculation methods have been reviewed (*1*, *4*, *7*). They have various limitations and advantages (Table 1). The calculations are often laborious but this is not a major factor given the appropriate

computer application software. The methods differ statistically but appear to give endpoint estimates that differ only by a few percent. This variation is insignificant relative to the imprecision of spiking level estimation (7).

Table 1. Estimation methods for LD_{50} and ID_{50} values

Name	Characteristics
Probit analysis	High efficiency; reiterated interpolation; replicates/spike & spiking intervals can vary
Reed & Muench	Lowest efficiency
Spearman-Kärber	Symmetrical doses; 0 and 100 % response values needed
Moving Average	Simple interpolation; curve shape not presumed

The lack of clear statistical superiority of the other calculation methods to the Spearmann- Kärber method along with its previous application to LOD_{50} calculations in studies of foods spiked with pathogens (5, *8, 9, and 10*) is the reason for its use in the present proposal. Also, an Excel version of the generalized Spearman-Kärber LOD_{50} calculation for 3, 4, and 5-level spiking protocols (*2;* Anthony.Hitchins@cfsan.fda.gov), now makes it more easily circulated and PC-user friendly. The accompanying Excel file provides a 3-level spike example, a trial worksheet, a back-up copy, a revealed code version, and the generalized Spearman-Kärber formula.

The LOD_{50} Determination

Foods are quantitatively spiked in replicate (*at least* in triplicate) with the test microbe at several inoculum levels (*at least* three). The proportion of replicates in which the microbe is detected at each spiking level is used to calculate the LOD_{50} by the generalized version of the Spearman-Kärber method . The confidence interval of the estimate narrows with increasing replication. The spiking level enumerations have their own confidence limits, which can be quite broad as in a 3-replicate MPN, but the overriding effect of any one MPN value is more or less ameliorated by the use of 3 or more enumeration levels in estimating the LOD_{50} value. Furthermore, the number of replicates can be increased to reduce the confidence interval of the MPN.

When there is comparison with a standard method the spiking level can be determined from the standard method result, since the MPN enumeration would be done with the standard method anyway. Thus the proportion of negative culture at a given spiking level yields, by the Poisson equation, the mean spiking level. In this method, the number of replicates should be preferably 10 or more. Replication values of 40 or more are easily achievable in multilaboratory experiments.

Incidentally, in multilaboratory experiments the LOD 50% can be calculated from the pooled data or it can be estimated as the mean of the individual laboratory LOD 50% values. In the latter case an estimate of interlaboratory uncertainty can be made.

Table 2 shows a simulated LOD_{50} experiment.

Table 2. Example of an LOD_{50} experiment using hypothetical data for a 4-level spike

Spiking Level (cfu/25 g)	Microbe Recovery			
	No. replicates	No. positive	No. negative	LOD_{50} (CI)[a]
0[b]	10	0	10	.
1	10	5	5	1.26 (0.53 – 3.03) cfu/25-g
10	10	9	1	.
100	10	10	0	.

[a] Calculated by the Spearman-Kärber method. CI = 95 % confidence interval.
[b] A value of 0.1 was assumed for the calculation.

An LOD_{90} value can be calculated from the LOD_{50} value in Table 2: it is 2.87 cfu/25-g test portion. This calculation assumes that the LOD endpoint curve is described by the Poisson equation even when the observed LOD_{50} value is different from the theoretical Poisson-based minimum LOD_{50} value of 0.307 cfu/25-g test portion. This assumption is reasonable for the majority of published *Listeria* method validation studies (6).

Typically collaborative qualitative microbiology method validations involve 3 spiking levels and 5 replicate determinations per level for each of 10 or more laboratories. This provides 150 or more data points (10 laboratories x 3 levels x 5 replicates). Intuitively, the LOD_{50} estimate by mathematical interpolation will be more accurate the greater the number of data points comprising the curve in the zone around the LOD_{50} point. Increasing the number of concentration levels does not require maintaining the same level of replication in order to sustain a given confidence level interval with a constant number of laboratories. This is illustrated in Table 3.

Table 3. Confidence Intervals for Two Spike-Level:Replicate Trade-off Scenarios with Similar LOD_{50} Results

10 lab x 7 level x 3 rep - 20[a] = 190 data points			10 lab x 3 level x 7 rep - 20[a] = 190 data points		
Mean Level[b] (cfu/25g)	Replicates per level	Positive Replicates[c]	Mean Level[b] (cfu/25g)	Replicates per level	Positive Replicates[c]
4.6	30	30	4.6	90	90
2.3	30	27	.	.	.
1.15	30	21	.	.	.
0.625	30	13	0.625	90	39
0.313	30	6	.	.	.
0.157	30	1	.	.	.
0(<0.075)	10[c]	0	0(<0.075)	10[c]	0
LOD_{50} = 0.760 cfu /25-g analytical portion			LOD_{50} = 0.700cfu /25-g analytical portion		
95% confidence interval = 0.575– 0.875			95% confidence interval = 0.550-.875		

[a] The number of replicates at the zero level can be less than at the other levels, say 1 per laboratory, since their purpose here is to provide a zero positives data point as well as the usual assurance of a negligible natural contamination rate.

[b] Level intervals based on 1:2 dilutions as in R. Flowers's dilution to extinction method. Not all levels used in the 3-level scenario. More levels in the LOD concentration zone could be set-up with a lower dilution rate, e.g. 1in 1.5.

[c] Common levels of the two scenarios have equal proportions of positive replicates.

Discussion and Recommendations

The method is broadly applicable (3) to all published AOAC collaborative studies except that in a proportion of the results it has been necessary to resort to dummy values for the required 100% positive response data points. The dummy concentration value for 100% positive response is currently over conservatively set at 10x the experimental concentration that yielded the highest proportion of positives. In a planned revision of the Spearman-Kärber LOD_{50} program, the 100% positive dummy concentration will be calculated by multiplying the highest concentration giving positives by the reciprocal of the proportion positive at that concentration. Of course, this necessity for a 100% positive dummy concentration can be largely avoided by increasing the number of concentration levels studied from the usual 3 levels to 4 or more concentration levels. The process of preparing concentrations that give partial positives is somewhat chancy and so it is likely that analysts are preparing levels giving 100% positive responses but are not presenting them since the current study protocol does not require them. So increase of the number of levels is unlikely to be onerous especially since the number of replicates per level can be correspondingly reduced (i.e. the product of the number of replicates per level and the number of levels need not be changed).

The generalized Spearman-Kärber method also requires a data point giving the concentration corresponding to zero positives. MPN limits of detection vary from <0.003 to <3 MPN/g depending on the maximum MPN sample size in the range from 100g down to 0.1g. There is no precise non-zero spike concentration (zero is not compatible with the logarithmic Spearman-Kärber calculation) corresponding to the controls used in AOAC studies. A value of 0.004 per g has been chosen as the concentration corresponding to the negative controls. This value is close to the minimum MPN likely to be encountered in spiking studies but more importantly is the extrapolation to zero% positive point of the midpoint region of the response curve, which is approximately linear and, which contains the LOD_{50} point of interest. While one can interpolate the LOD_{50} value from the experimental data, using the Spearman-Kärber method to obtain the LOD_{50} also provides the confidence limits.

A proportion of published AOAC study results were not readily amenable to the LOD calculation. The use of 3-tube MPN sometimes gives sequential concentrations that are equal even though they should be different and even gives sequential values that are different but appear as if they have been inadvertently reversed. These problems can be solved by using a better MPN enumeration with more tubes per level or by using the standard method positive responses in a one level multi-tube MPN calculation (if a standard method is available) or by using the method suggested by R. Flowers. Nevertheless, retention of the conventional statistical tests used currently would be advisable for rare instances were the LOD_{50} cannot be calculated.

Conclusions

The calculation of an LOD_{50} value by the generalized Spearman-Kärber method provides a convenient way to condense virtually *all* of the raw data from a multi-level food spiking trial into one readily comprehensible *absolute* value of performance efficacy. In addition, it provides the estimate's *uncertainty*, given as the 95% confidence interval. The breadth of the confidence interval will depend inversely on the number of replicates at each level. The replication at each level need not be constant in this generalized version of the Spearman-Kärber calculation. More sophisticated calculation methods may become available in the future but meanwhile the generalized Spearman-Kärber method is already available to do the job of calculating detection limits and moreover it has the advantage of not requiring complex computations. The problem with modeling an empirical response curve from all available study data is that each data point from study to study involves so many variables and the plot of % positive versus concentration is highly scattered.

In study designs where a new method and the standard method are compared, LOD_{50} values do not just augment the conventional relative performance parameters with absolute performance parameters; in addition, they also provide estimates of the uncertainties of the method's performances. LOD_{50} values for one-method study designs can be compared with previously published values for that and other methods and also with the theoretically expected minimal recovery value for a particular analytical portion size.

It is clear that the generalized Spearman-Kärber method will be most useful if the AOAC collaborative study design is adjusted appropriately by innovations such as Russ Flowers's dilution to extinction method.

REFERENCES

1. Armitage, P. and Allen, I. 1950. Methods of estimating the LD_{50} in quantal response data. J. Hyg., Camb. **48**:298.
2. Encyclopedia of Statistical Sciences. 1982. **1**:355. John Wiley & Sons, New York.
3. Burney, A. A., and Hitchins, A. D. 2003. Determination of the limits of detection of AOAC validated qualitative microbiology methods. Manuscript in preparation.
4. Finney, D. J. 1952. *Probit Analysis*, 2nd ed. Cambridge University Press, London and New York.
5. Hitchins, A. D. 1989. Quantitative comparison of two enrichment methods for isolating *Listeria monocytogenes* from inoculated ice cream. J. Food Protect **52**:898.
6. Hitchins A. D. 1998. Retrospective interpretation of qualitative collaborative study results: Listeria methods. AOAC Intl. Annual Meeting **112**: 102, Abstract J-710.
7. Meynell, G. G., and Meynell, E. 1965 *Theory and Practice in Experimental Bacteriology*, pp 179-182. Cambridge University Press, London and New York.
8. Thunberg, R.L., Tran, T.T., and Walderhaugh, M. O. 2000. Detection of thermophilic *Campylobacter* spp. in blood-free enriched samples of inoculated foods by the polymerase chain reaction. J. Food Protect. **63**:299.
9. Tran, T. T., Stephenson, P. and Hitchins, A. D. 1990. The effect of aerobic mesophilic microfloral levels on the isolation of inoculated *Listeria monocytogenes* strain LM82 from selected foods. J. Food Safety **10**: 267.
10. Twedt, R. M., Hitchins, A. D., and Prentice, G. 1994. Determination of the presence of Listeria monocytogenes in milk and dairy products: IDF collaborative study. J. AOAC INTERNATIONAL **77**:395.