

Probability of Detection (POD) as a Statistical Model for the Validation of Qualitative Methods

PAUL WEHLING

General Mills, Inc., 330 University Ave SE, Minneapolis, MN 55414

ROBERT A. LABUDDE

Least Cost Formulations, Ltd, 824 Timberlake Dr, Virginia Beach, VA 23464

SHARON L. BRUNELLE and MARIA T. NELSON

AOAC INTERNATIONAL, 481 N. Frederick Ave, Gaithersburg, MD 20877

A statistical model is presented for use in validation of qualitative methods. This model, termed Probability of Detection (POD), harmonizes the statistical concepts and parameters between quantitative and qualitative method validation. POD characterizes method response with respect to concentration as a continuous variable. The POD model provides a tool for graphical representation of response curves for qualitative methods. In addition, the model allows comparisons between candidate and reference methods, and provides calculations of repeatability, reproducibility, and laboratory effects from collaborative study data. Single laboratory study and collaborative study examples are given.

European validation organizations utilize ISO standard 16140:2003(E), *Protocol for the Validation of Alternative Methods* (1), as a guideline for the validation of qualitative and quantitative microbiological methods. In the United States, AOAC INTERNATIONAL uses the *AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Qualitative and Quantitative Food Microbiological Official Methods of Analysis* (hereafter referred to as “AOAC guidelines”), published in 2002 (2). The ISO standard and the AOAC guidelines are largely harmonized with respect to the types of studies and the study designs required for validation of microbiological methods, although statistical analyses differ. This paper will focus on the validation of qualitative methods.

Since publication of the ISO standard and the AOAC guidelines, technical and statistical issues with both documents have arisen. Chief among them is the fact that neither document addresses qualitative method validation with the use of unpaired or independent test portions. Many developers of proprietary methods are deviating from the traditional reference method enrichment schemes in favor of

either proprietary enrichment media or, in the case of highly specific detection technology such as PCR, nonselective enrichments. Such validations require the use of independent test portions, while the current statistical model applies only narrowly to the circumstance of proprietary methods using the reference method enrichment. Formulas for McNemar's ² analysis, sensitivity, specificity, false positive, and false negative have little utility outside this narrow scope.

Furthermore, method developers are devising detection methods for emerging pathogens for which there are no established regulatory reference methods. Since the statistical models contained in the current ISO standard and AOAC guidelines depend on comparison of candidate methods to regulatory reference methods, there is no guidance on establishing independent method performance parameters. As a result of this, methods cannot be compared between studies, only within a study using the same set of samples.

It is our intention to develop independent measures of method performance that are more broadly applicable with the goal of developing a statistical model for qualitative methods that will be as useful and successful as the quantitative model. Further, we intend to leverage the recommendations of the AOAC INTERNATIONAL Presidential Task Force on Best Practices for Microbiological Methods (<http://www.fda.gov/Food/ScienceResearch/LaboratoryMethods/ucm124900.htm>) to develop a new paradigm for validation of qualitative methods.

Historical Perspective

The system of validating qualitative methods with the parameters of sensitivity, specificity, false positive, and false negative has been used with greatest success for clinical tests. In clinical testing, determining the true clinical state of a patient is also a binary state (i.e., you are either pregnant or not, or you either have cancer or not). Eventually, the true state is revealed, either by birth (you really were pregnant) or by biopsy or autopsy (you never had cancer).

In qualitative chemical and microbiological testing, method sensitivity depends on the concentration of the analyte. For most chemical and microbiological methods, the

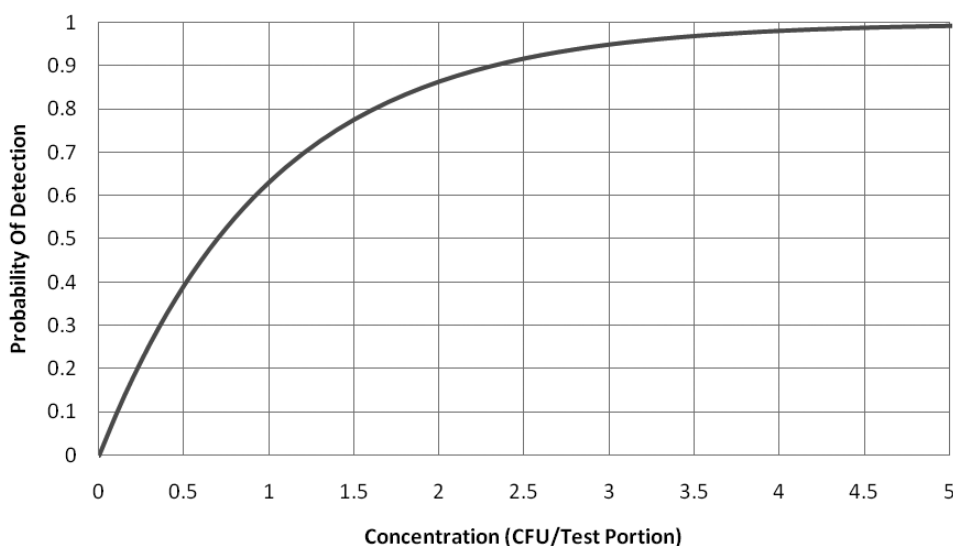


Figure 1. POD response curve of a microbiological method where Poisson sampling variation dominates.

probability of positive response may be very low at low (but non-zero) concentrations and very high at much higher concentrations (Figures 1 and 2). In effect, probability is dependent on concentration. A drawback of the sensitivity/specificity model is its difficulty of inclusion of concentration as a continuous variable for modeling probability response.

The validation schemes for qualitative microbiological assays as described in ISO 16140 were methods designed to alleviate this problem by treating each unknown level as a separate experiment and changing the probability definition of sensitivity from conditional on presence or absence of the analyte to conditional on the response of a reference method to an aliquot of the same test portion.

ISO 16140 was limited in scope to proprietary microbiological tests intended to be used as rapid substitutes for slower, well-validated reference methods that generally take days to perform. The methods of ISO 16140 require a valid reference method. We hope to develop a new model for qualitative method validation that will not require a reference method comparison, but will allow for one if there is such a need.

Our concept is to simplify this model and combine sensitivity, specificity, false-positive, and false-negative parameters into one single parameter, Probability of Detection (POD), that will cover all contingent ranges of concentration, both zero and non-zero. This simplified model allows the ability to compare probabilities across concentrations and

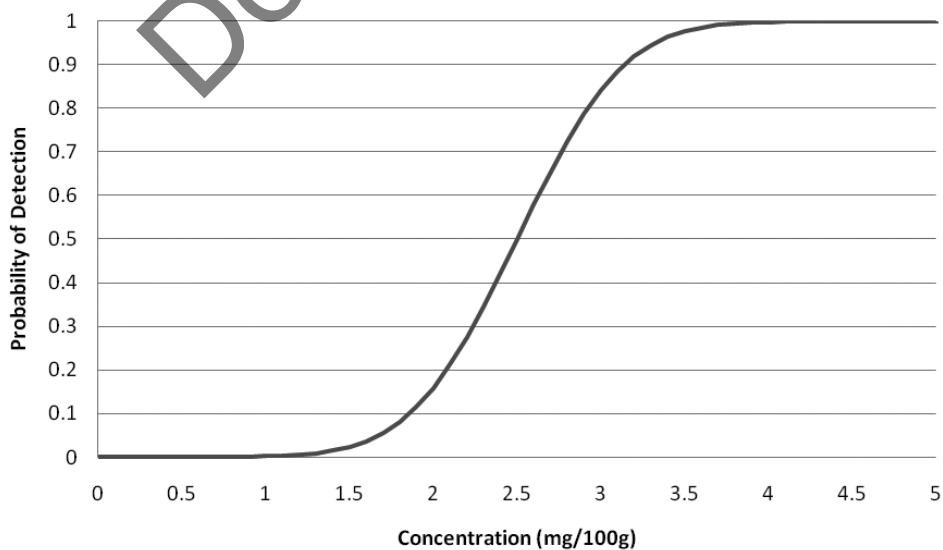


Figure 2. Theoretical POD response curve for a method that uses a threshold value on some normally distributed output.

further allows for a simple graphical representation of validation data as a POD curve graphed by concentration with associated error bars. Comparison of the response curves of two or more methods, the comparison to a hypothetical idealized response curve based on sampling probabilities, the comparison of the response curves for two or more foods, and the ability to calculate repeatability and reproducibility precisions are major advantages of this model over the traditional statistical parameters. The intention is that this model will be general enough to be a framework for method validation applicable to all chemical and microbiological methods, whether or not a reference method exists for the analyte.

POD Model for Qualitative Method Validation

Probability of Detection

Consider a qualitative method, which for the purposes of this discussion shall be limited to a method of analysis that is restricted to a binary result. For example, the method may give a positive or negative response, a 1 or a 0, or a TRUE or FALSE as the output. In this paper we will consider only methods with two possible response outcomes.

For such methods, we can describe the probability of the method giving a positive (or 1 or TRUE) result. This probability is generally considered to be dependent upon the concentration or amount of analyte present in the sample. For most well-designed qualitative methods, the probability of a positive response is near zero when the analyte is not present, and the probability should approach 1 as the analyte concentration or mass increases. The exact shape of this response curve will depend on the analyte type, the sampling procedures, and the detection principles involved in the method. Characterization of the POD parameter with respect to concentration is the fundamental task facing the experimenter who wishes to validate a qualitative method.

Figures 1 and 2 are graphical representations of theoretical POD response curves for two different types of qualitative methods. Figure 1 is a theoretical POD response curve of a microbiological method where Poisson sampling variation ("1-hit Poisson model"; 3) dominates. The POD here is dependent on the probability of obtaining at least one viable colony-forming unit (CFU) in the test portion. The curve is based entirely on sampling probabilities. Figure 2 is a theoretical POD response curve for a method that uses a threshold value on some normally distributed output. For example, an ELISA method may have a plate reader that gives output as OD, but for ease of use, the reader software prints a qualitative result of "+" if the OD is greater than some threshold value. If OD is normally distributed, the POD response curve will be of the form shown in Figure 2. See Finney (4) for further elaboration of probit response curves.

In this model, POD is a conditional probability, with concentration as the conditioning variable. It is our contention that the critical aim of a method validation study should be to characterize the shape of this POD response curve with respect to concentration. An understanding of method

response is critical to interpreting results, as well as being a fundamental criterion for choosing one method over another for a given intended use.

Estimates for POD

The estimation method for qualitative POD will utilize the same averaging techniques as used for quantitative methods (5–7), with the coding of results as 1 for a positive response and 0 for a negative response. The mean of all responses for a given concentration will be the proportion of positive responses. In this way, POD can be considered to be the qualitative analog of the mean parameter of a quantitative method. *Appendix A* contains an explicit elaboration of a statistical model for binary data, as derived from the qualitative model of ISO 5725-1. *Appendix B* contains detailed descriptions of POD estimation techniques and confidence intervals for those estimators.

Confidence Intervals for POD

At a given concentration, once the POD is estimated, all individual analytical responses are assumed to be Bernoulli trials at the probability $p = \text{POD}$. Confidence intervals for the POD estimate, therefore, should be based on binomial probabilities given the POD estimate and the total number of trials in the study.

For POD estimates, LaBudde (8) recommends a binomial 95% confidence interval calculated by the "modified" Wilson method, which he introduces. This method is a modification to improve coverage accuracy of the popular Wilson score interval, which is based on a normal approximation for the binomial probabilities, and is very efficient at $N > 20$ or so. For the sample sizes usually encountered in qualitative collaborative studies, this method will work well. It should be noted that the use of the normal approximation for these binomial intervals in no way implies a parametric assumption about the shape of the underlying POD curve.

Comparing Method Responses

The POD for any two methods can be compared by difference at a given analyte concentration. Likewise, the POD for presumptive and confirmed results of a rapid method can be compared by difference. This difference in POD values is termed dPOD. The statistical significance of dPOD is determined by its confidence interval. If the confidence interval on dPOD includes zero, then the difference between the methods being compared is not significant. LaBudde (9) presents recommended confidence intervals for a POD difference.

Comparing the responses of methods as a difference on the probability scale is a direct analog to the quantitative parameter of bias between methods as the difference of means. Similarly, the response of a method could be compared to an accepted reference value, or to a theoretical response value such as a Poisson sampling curve. In all cases, the comparison should be made as a difference in probability.

Table 1. Analogous parameters between quantitative and qualitative validation models

Method attribute	Quantitative parameter	Quantitative estimate	Qualitative parameter	Qualitative estimate
General mean or expectation	Mean (μ)	Mean (\bar{x})	POD	POD or LPOD
Repeatability variance	s_r^2	S_r^2	s_r^2	S_r^2
Reproducibility variance	s_R^2	S_R^2	s_R^2	S_R^2
Laboratory variance	s_L^2	S_L^2	s_L^2	S_L^2
Expected difference between two methods ^a	Bias (B)	$\bar{x}_1 - \bar{x}_2$	dPOD	POD ₁ - POD ₂

^a Or difference between method response and an accepted reference value.

Estimating Method Variance

Table 1 shows how the POD statistical model for qualitative methods allows for a unified statistical approach for qualitative and quantitative methods whether for chemical or microbiological analytes. Method variance from a collaborative validation study can be modeled as:

$$s_R^2 = s_r^2 + s_L^2$$

where s_R is the SD of reproducibility, s_r is the SD of repeatability, and s_L is the laboratory effect. Calculations for qualitative methods are based on the POD values at each concentration, whereas calculations for quantitative methods are based on the mean values at each analyte concentration.

The estimation method for qualitative POD variance will use the analysis of variance (ANOVA) model as defined for the quantitative model given in ISO 5725-1 and will use the same ANOVA calculation methods, but instead of entering a qualitative result, results will be coded as 1 for a positive response and 0 for a negative response.

Both quantitative and qualitative methods have, in practice, commonly observed systematic dependencies between mean analyte level and variance. For the binary qualitative case, at concentrations where observed POD values are close to 0 or 1, very little variation will be observed in the data sets, as observations will be either all positive or all negative. At concentrations where POD values fall in the fractional range (e.g., between 0.15 and 0.85), more variation

will be observed within each laboratory. This is primarily a consequence of the binomial nature of the response of the method.

Just as in the case for quantitative methods, much of this variation is due to inherent variation between test portions as prepared for the collaborative study. Some of the variation will also be due to variation of the method of detection in this fractional range. The relative magnitude of these two sources of variation is usually not determined in a validation experiment, and both sources, regardless of their relative size, are included in the variance calculations as repeatability, or within-laboratory variance (see ISO 5725-1:1994 Section 6.4.5). As long as the conditions of the validation experiment conform to the ISO definition of repeatability as stated in ISO 5725-1:1994 (identical test materials, same method, same laboratory, same operator, and same equipment in a short period of time; 10), the within-laboratory SD estimate obtained from the ANOVA procedure described in *Appendix C* is correctly called "repeatability SD."

Method developers will need to accustom themselves to the expected magnitudes of SD estimates in this qualitative model. At POD values with little or no variation, SDs will be expected to be nearly zero. At fractional POD values, the SDs will increase, with a maximum at POD = 0.5, where the reproducibility SD will be expected to be very close to 1/2. *Appendix C* contains detailed descriptions of estimation techniques for the variance parameters.

Table 2. Example of single-laboratory data for detection of *E. coli* O157:H7 in apple juice

Concn, MPN/25 g ^a	N	Candidate method			Reference method				
		x^b	POD	95% CI ^c	x	POD	95% CI	dPOD	95% CI
0	5	0	0.00	(0.00, 0.43)	0	0.00	(0.00, 0.43)	0	(-0.43, 0.43)
1.05	20	12	0.60	(0.39, 0.78)	10	0.50	(0.30, 0.70)	0.10	(-0.19, 0.370)
2.30	20	20	1.00	(0.84, 1.00)	19	0.95	(0.76, 1.00)	0.05	(-0.12, 0.24)

^a MPN = Most probable number.

^b x = Number of positive trials.

^c CI = Confidence interval.

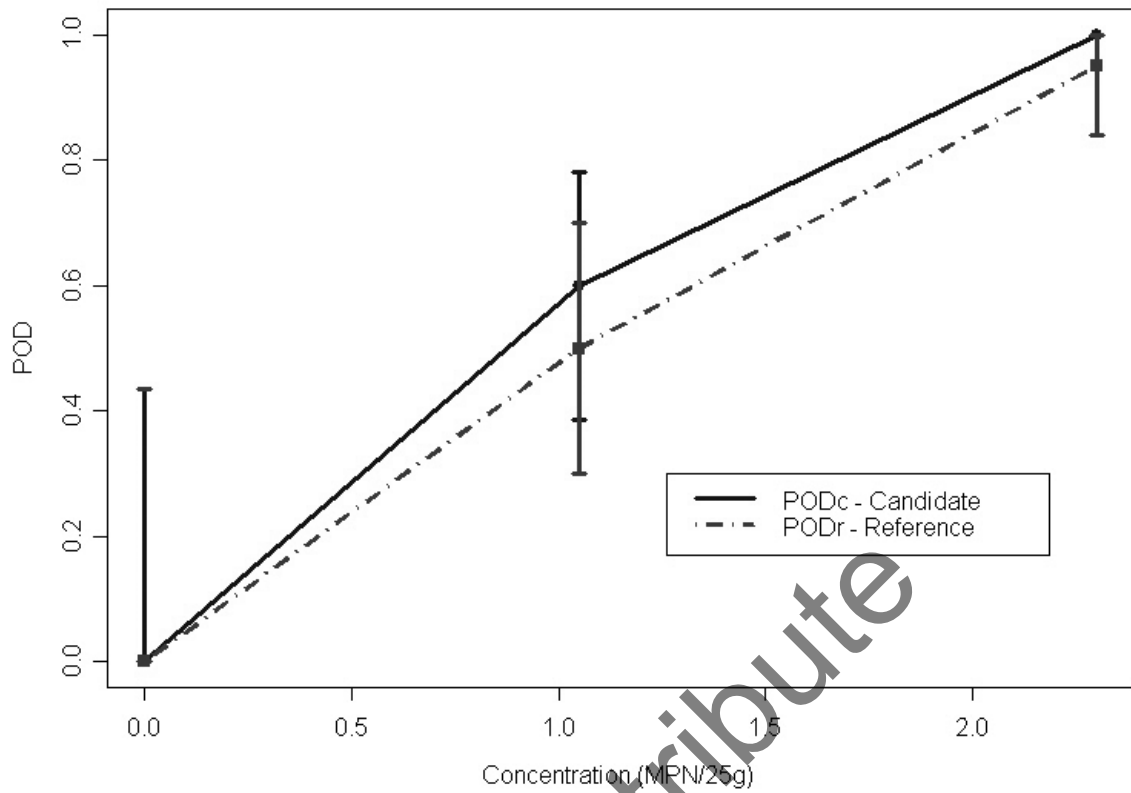


Figure 3. POD curve for *E. coli* O157:H7 in apple juice.

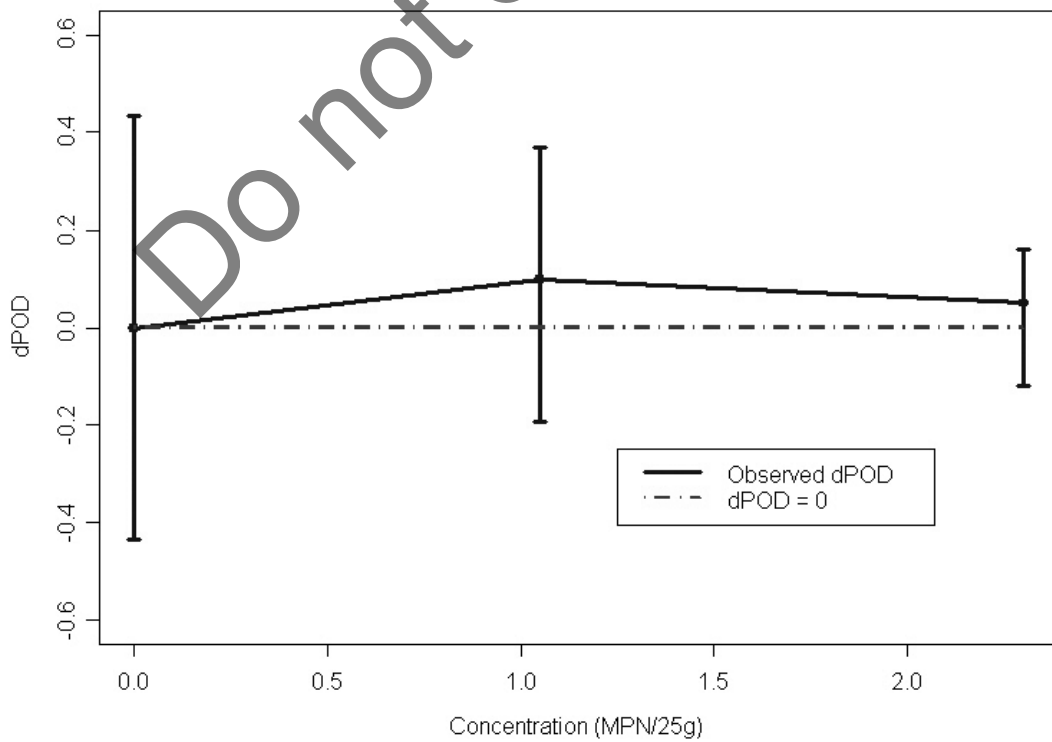


Figure 4. dPOD plot for *E. coli* O157:H7 in apple juice.

Examples

Table 2 shows an example of a single-laboratory study of *E. coli* O157:H7 in apple juice (personal communication; unpublished data). The difference between the candidate method response and the reference method response is calculated as dPOD. It is evident from the confidence intervals that the two methods are not statistically different at any level tested. The data are plotted in Figure 3, and the lack of a difference between methods is suggested by the overlapping confidence intervals of the candidate method POD (POD_c) and reference method POD (POD_r) plots as well as the dPOD plot (Figure 4), whose confidence intervals straddle the dPOD = 0 line.

Table 3 contains data from a collaborative study for detection of *Salmonella* in ground beef (11). The study produced data from 11 laboratories. The observed proportion positive response from Laboratory 6 at the high level (concentration = 10.75 CFU/25 g) was unusually low compared to the other laboratories. For the purpose of illustration, we have not included results from this laboratory in the final analysis.

Table 4 includes the summary statistics for the collaborative study, which are designated as LPOD and dLPOD, with associated 95% confidence intervals. The designator L before the POD signifies that this is a composite POD pooled across laboratories and includes between-laboratory variation in addition to variation inherent in the binomial nature of the binary probabilities. LPOD is an estimate of the average POD across laboratories. Figures 5 and 6 are graphical representations of the collaborative study statistics, LPOD versus concentration, and dLPOD versus concentration. The dLPOD plot shows a significant difference between the candidate and reference methods at the middle (concentration = 0.75 CFU/25 g) level.

Discussion

The advantages of the POD model for qualitative methods over the current paradigm include the ability to plot qualitative data as a function of concentration and the unification of statistical parameters between qualitative and quantitative method validation. The POD model solves the statistical problems with unpaired test portions, is easy to understand and implement by the nonstatistician, and provides calculations of statistical parameters, such as reproducibility, that were not possible with the traditional clinical model.

Our primary goal in the development of this validation scheme was to devise a system that is general enough to use on all qualitative (binary output) methods. The system as presented is general enough to be used on a variety of methods, chemical or microbiological, with or without a reference method comparison. Whenever possible, we have retained flexibility to include method comparisons, if required, as part of a validation. Because of this flexibility, the POD model can be used as a replacement model for the

current AOAC validation guidelines and ISO 16140, and can also be used as a validation model for any other qualitative method validation.

In addition to flexibility, we have tried to simplify parameters as much as possible. The combination of information inherent in sensitivity, specificity, false positive, and false negative as the POD parameter is certainly a simplification. In addition, treating POD as conditional on concentration as opposed to conditional on presence/absence relieves the model of any inconsistencies inherent in the expression of sensitivity.

Modeling qualitative method validation as POD conditional on concentration is consistent with method end-user requirements. Usually method users will specify requirements for method detection as relative to analyte concentration, which may relate to toxicity, allergenic sensitivity, or other important parameters. The qualitative requirement may be specified as a certain POD at a given concentration, sometimes at a regulatory action level. Acceptance criteria for method validations can be developed based on actual end-use requirements as opposed to strict adherence to reference method performance. At the same time, the special case where method users are concerned about response relative to a reference method can be accommodated within the general model.

The POD model as described here can be further modified to include other parameters. We have developed a base model whereby POD is modeled as probability conditional on concentration. The next logical step is to use binomial regression techniques to relate observed POD across experimental concentration levels. The potential advantage of this would be to interpolate method response at other concentration levels not specifically tested in the validation experiment. Other parameters could be defined and estimated. For example, an LOD parameter could be defined from the POD parameter. LOD₅₀ can be defined as the concentration where POD is equal to a probability of 1/2. Estimation techniques modeling POD across concentration levels using binomial regression techniques can be developed to estimate LOD at POD = 1/2 or any other probability. The scope of this paper is to define the base model, as this type of regression estimation is not currently used in the ISO standard. Our intention was to develop the base model to replace the current system in its traditional applications, but one advantage of the POD model is that it opens up the possibility of this type of future development, which is not available in the current model.

The quantitative method validation parameters and estimation techniques as outlined in AOAC INTERNATIONAL guidelines and ISO standards have been developed over the course of 125 years of chemical and microbiological methods development. These techniques have proven to be widely successful as a tool for validation of quantitative methods. Our goal in developing a qualitative validation model is to develop a system that will be as successful as the quantitative system. To achieve this goal, we have borrowed extensively from parameters and statistics as

Table 3. Example of collaborative data for detection of *Salmonella* in ground beef

Concn, MPN/25 g ^a	Laboratory	n	Candidate method			Reference method			dPODc or dLPODc	95% CI
			x ^b	PODc or LPODc	95% CI ^c	x	PODr or LPODr	95% CI		
0	1	6	0	0		0	0	0.00		
0	2	6	0	0		0	0	0.00		
0	3	6	0	0		0	0	0.00		
0	4	6	0	0		0	0	0.00		
0	5	6	0	0		0	0	0.00		
0	6	6	0	0		0	0	0.00		
0	7	6	0	0		0	0	0.00		
0	8	6	0	0		0	0	0.00		
0	9	6	0	0		0	0	0.00		
0	10	6	0	0		0	0	0.00		
0	11	6	0	0		0	0	0.00		
0	All	60	0	0	(0.0, 0.060)	0	0	0.00	(-0.060, 0.060)	
0.75	1	6	1	0.17		2	0.33	-0.17		
0.75	2	6	1	0.17		1	0.17	0.00		
0.75	3	6	0	0.00		3	0.50	-0.50		
0.75	4	6	1	0.17		3	0.50	-0.33		
0.75	5	6	3	0.50		5	0.83	-0.33		
0.75	6	6	0	0.00		1	0.17	-0.17		
0.75	7	6	1	0.17		2	0.33	-0.17		
0.75	8	6	5	0.83		4	0.67	0.17		
0.75	9	6	0	0.00		4	0.67	-0.67		
0.75	10	6	2	0.33		2	0.33	0.00		
0.75	11	6	0	0.00		2	0.33	-0.33		
0.75	All	60	14	0.23	(0.06, 0.41)	28	0.47	-0.233	(-0.45, -0.014)	
10.75	1	6	4	0.67		6	1.00	-0.33		
10.75	2	6	5	0.83		4	0.67	0.17		
10.75	3	6	5	0.83		5	0.83	0.00		
10.75	4	6	5	0.83		6	1.00	-0.17		
10.75	5	6	6	1.00		6	1.00	0.00		
10.75	6	6	0	0.00		2	0.33	-0.33		
10.75	7	6	6	1.00		6	1.00	0.00		
10.75	8	6	6	1.00		6	1.00	0.00		
10.75	9	6	6	1.00		5	0.83	0.17		
10.75	10	6	4	0.67		6	1.00	-0.33		
10.75	11	6	4	0.67		6	1.00	-0.33		
10.75	All	60	51	0.85	(0.76, 0.94)	56	0.93	-0.083	(-0.18, 0.048)	

^a MPN = Most probable number.^b x = Number of positive trials.^c CI = Confidence interval.

Table 4. Statistical summary for collaborative study for *Salmonella* in ground beef

	Candidate method		
	Low level	Mid level	High level
Concentration	0.00 MPN/25 g	0.75 MPN/25 g	10.75 MPN/25 g
Number laboratories (reported/used)	(11/10)	(11/10)	(11/10)
N total replicates	60	60	60
LPODc	0.00	0.233	0.850
LPODc 95% CI	(0.00, 0.060)	(0.040, 0.384)	(0.757, 0.943)
S_r^a	0.00	0.3568	0.3606
S_L^b	0.00	0.2144	0.000
S_R^c	0.00	0.4162	0.3606
Reference method			
Concentration	0.00 MPN/25 g	0.75 MPN/25 g	10.75 MPN/25 g
Number laboratories (reported/used)	(11/10)	(11/10)	(11/10)
N total replicates	60	60	60
LPODr	0.00	0.467	0.933
LPODr 95% CI	(0.00, 0.060)	(0.331, 0.602)	(0.841, 0.974)
S_r	0.00	0.4954	0.2449
S_L	0.00	0.0711	0.0598
S_R	0.00	0.5005	0.2522
Comparison of candidate method to reference method			
Concentration	0.00 MPN/25 g	0.75 MPN/25 g	10.75 MPN/25 g
Number laboratories (reported/used)	(11/10)	(11/10)	(11/10)
N total replicates	60	60	60
dLPOD	0.00	-0.233	-0.083
dLPOD 95% CI	(-0.060, 0.060)	(-0.452, -0.014)	(-0.184, 0.048)

^a S_r = Repeatability SD.

^b S_L = Laboratory SD.

^c S_R = Reproducibility SD.

defined in ISO 5725. Every parameter developed in the POD model, as shown in Table 1, is directly analogous to traditional parameters in the quantitative scheme. The POD model as described here will provide a way to bridge the differences that previously existed between qualitative and quantitative method characterizations.

Conclusions

The POD model is a simple, broadly applicable model for qualitative method validation that harmonizes with the statistical parameters for quantitative method validation, yielding a unified statistical approach for all method

validation. Our vision is that when adopted, the POD model for qualitative method validation will be as successful as the AOAC/ISO/International Union of Pure and Applied Chemistry (IUPAC) harmonized quantitative system.

Acknowledgments

We would like to thank Scott Coates of AOAC INTERNATIONAL Research Institute for support on this project, Darryl Paulson for ideas, and the AOAC INTERNATIONAL Presidential Task Force on Best Practices for Microbiological Methods for starting the process moving forward.

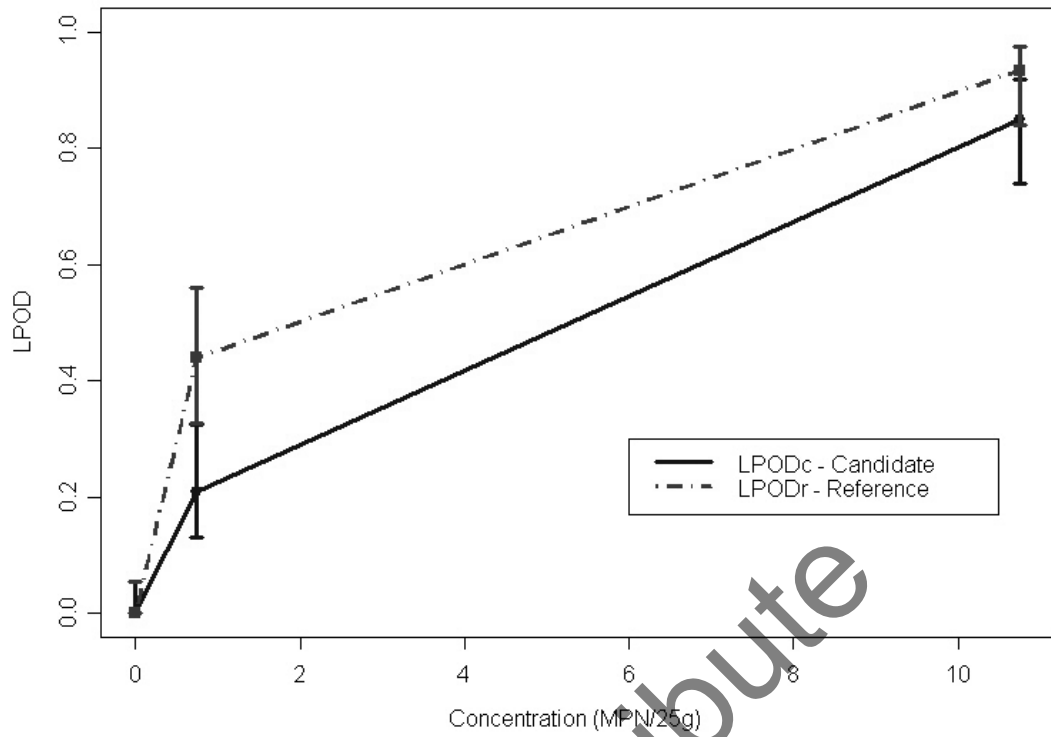


Figure 5. LPOD curve for *Salmonella* in ground beef collaborative study.

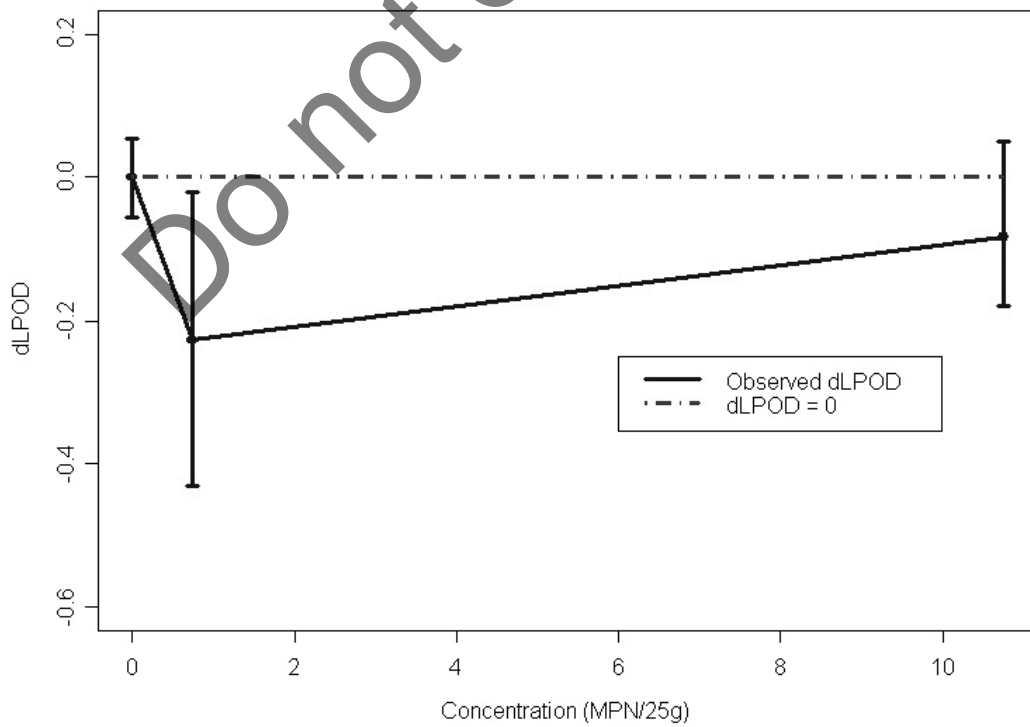


Figure 6. dLPOD plot for *Salmonella* in ground beef collaborative study.

References

- (1) ISO 16140:2003(E) *Microbiology of Food and Animal Feeding Stuffs—Protocol for the Validation of Alternative Methods*, International Organization for Standardization, Geneva, Switzerland
- (2) AOAC INTERNATIONAL *Methods Committee Guidelines for Validation of Qualitative and Quantitative Food Microbiological Official Methods of Analysis* (2002) *J. AOAC Int.* **85**, 1187–1200 (Also reprinted in OMA Program Manual as “Appendix X”)
- (3) LaBudde, R.A. (2008) *Statistical Analysis of Interlaboratory Studies, XX, Measuring the Performance of a Qualitative Test Method, TR290*, Least Cost Formulations, Ltd, Virginia Beach, VA, <http://www.lcfltd.com/Documents/tr290%20stat%20anal%20inter%20XX%20meas%20perf%20qual%20test%20metho.pdf>
- (4) Finney, D.J. (1979) *Statistical Method in Biological Assay*, 3rd Ed., Oxford University Press, Oxford, UK
- (5) ISO 5725-2:1994, *Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method*, International Organization for Standardization, Geneva, Switzerland
- (6) AOAC INTERNATIONAL *Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis* (1995) *J. AOAC Int.* **78**, 143A–160A (Also reprinted in *Official Methods of Analysis of AOAC INTERNATIONAL* (2000) 17th Ed. and (2005) 18th Ed.)
- (7) Horwitz, W. (1995) *Pure Appl. Chem.* **67**, 331–343
- (8) LaBudde, R.A. (2009) *Coverage Accuracy for Binomial Proportion 95% Confidence Intervals for 12 to 100 Replicates, TR297*, Least Cost Formulations, Ltd, Virginia Beach, VA, <http://www.lcfltd.com/Documents/tr297%20coverage%20accuracy%20binomial%20proportions.pdf>
- (9) LaBudde, R.A. (2009) *Statistical Analysis of Interlaboratory Studies, XXII, Statistical Analysis of a Qualitative Multicollaborator Study as a Quantitative Study Under the Large Sample Approximation, TR296*, Least Cost Formulations, Ltd, Virginia Beach, VA, <http://www.lcfltd.com/Documents/tr296%20stat%20anal%20inter%20XXII%20qual%20study%20char%20v3.pdf>
- (10) ISO 5725-1:1994, *Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 1: General Principles and Definitions*, International Organization for Standardization, Geneva, Switzerland
- (11) Mozola, M.A., Peng, X., & Wendorf, M. (2007) *J. AOAC Int.* **90**, 738–755

Do not distribute

Appendix A

Statistical Model for Binary Qualitative Methods of Analysis

For a particular material/level combination, it is useful to assume that every test result y is the sum of three components:

$$y = m + B + e$$

where y is the test result (limited to the values 0 or 1), m is the overall mean expected response, B is the laboratory component of bias under repeatability conditions, and e is the random error occurring in every measurement under repeatability conditions.

Note that there is a special case constraint to y for the binary case:

$$y = 0, 1$$

In the special case, because of the constraint placed on y , the practical implication is that m , B , and e will also be constrained for an individual replicate.

$$\begin{aligned} 0 &= m + B + e \\ 1 &= m + B + e \\ 1 &= m + e \end{aligned}$$

The following defining equations for the variance parameters will still apply as given in the general model of ISO 5725-1:

$$\begin{aligned} \sigma_L^2 &= \text{var}(B) \\ \sigma_W^2 &= \text{var}(e) \\ \sigma_r^2 &= \frac{\text{var}(e)}{r} \\ \sigma_R &= \sqrt{\frac{\text{var}(e)}{r}} \\ \sigma_L &= \sqrt{\frac{\sigma_L^2}{r}} \end{aligned}$$

The implication of this is that the “qualitative model” is not a separate model distinct from the “quantitative model,” but a special case subset of the general quantitative model. A convenient way of visualizing the qualitative case is to assume the quantitative model followed by a threshold detector that transforms the response to 0 or 1 values.

For quantitative methods, if m is in units of concentration, it is generally expected that $m = c$. If m is not a concentration (or amount) of analyte, m and c can be related by a calibration function.

For qualitative binary methods, this calibration cannot be easily achieved without replication, so the mean, m , has a special connotation in the binary model. With the coding convention of $y = 0, 1$ (i.e., 0 = “Negative” and 1 = “Positive”), the mean is the probability of a positive response at that concentration tested. This probability is the probability of a positive response at a given concentration, or POD:

$$m = \text{POD} = P(c)$$

Results of a validation experiment can be graphically presented as a plot of POD as a function of concentration, with 95% confidence intervals.

Variance component estimation via ANOVA with an additive model is not strictly correct for random laboratory variation adding to binary within-laboratory variation for a single replicate. However, with replication of $n = 12$ replicates/laboratory, the linear additive model approximates very well under the assumption of normality (as we have verified by extensive simulation) and provides for adequate parameter estimates in practice. The repeatability SD, σ_r , is estimated under the binomial model, so it is “exact” in the distribution sense. In the limits where POD approaches 0 or 1, all variance disappears, so the issue of accuracy becomes irrelevant.

Appendix B

Calculation of POD and dPOD Values from Qualitative Single-Laboratory Data

Calculate the POD as the ratio of the number positive (x) to total number tested (N):

$$\text{POD} = \frac{x}{N}$$

where POD is POD_c , POD_R , etc.

The POD estimates and 95% confidence interval [lower control limit (LCL) and upper control limit (UCL)] estimates are given by:

(1) For the case where $x = 0$:

$$\begin{aligned} \text{POD} &= 0 \\ \text{LCL} &= 0 \\ \text{UCL} &= 3.8415/(N - 3.8415) \end{aligned}$$

(2) For the case where $x = N$:

$$\begin{aligned} \text{POD} &= 1 \\ \text{LCL} &= N/(N - 3.8415) \\ \text{UCL} &= 1 \end{aligned}$$

(3) For the case where $0 < x < N$:

$$\begin{aligned} \text{POD} &= \frac{x}{N} \\ \text{LCL} &= \frac{x - 1.9207 \sqrt{x \frac{x^2}{N} - 0.9604}}{N - 3.8415} \\ \text{UCL} &= \frac{x + 1.9207 \sqrt{x \frac{x^2}{N} - 0.9604}}{N - 3.8415} \end{aligned}$$

where $1.9600 = z$, the Gaussian quantile for probability 0.975, $1.9207 = z^2/2$, $0.9604 = z^2/4$, and $3.8415 = z^2$.

Finally, if $x = 1$, set $\text{LCL} = 0$. If $x = N - 1$, set $\text{UCL} = 1$.

The confidence interval corresponds to the uncorrected Wilson score method, modified for $x = 1$ and $x = N - 1$ to improve coverage accuracy on the boundary (8).

The difference in proportions detected is estimated by LaBudde (9):

$$dPOD_c = POD_c - POD_R$$

The associated 95% confidence interval (LCL, UCL) for the expected value of $dPOD = POD_1 - POD_2$ is estimated by:

$$LCL = dPOD - \sqrt{\frac{POD_1}{L} \left(\frac{LCL_1}{L} \right)^2 + \frac{POD_2}{L} \left(\frac{UCL_2}{L} \right)^2}$$

$$UCL = dPOD + \sqrt{\frac{POD_1}{L} \left(\frac{UCL_1}{L} \right)^2 + \frac{POD_2}{L} \left(\frac{LCL_2}{L} \right)^2}$$

where (LCL_1, UCL_1) is a 95% confidence interval for POD_1 and (LCL_2, UCL_2) is a 95% confidence interval for POD_2 , as determined above.

Appendix C

Calculation of LPOD and dLPOD Estimates from Qualitative Multilaboratory Validation Data

For a multilaboratory trial where L = number of laboratories, R = replicates/laboratory, and $N = LR$ = total replicates, LPOD estimate is given by:

$$LPOD = \frac{x}{N}$$

where x is the number of positive results.

Method for Estimating LPOD 95% Confidence Intervals

Step 1: Analyze data as per quantitative statistical procedures given in ISO 5725-2 (5) with data coded as 1 for positive response and 0 for negative response. Record the mean LPOD, s_R , and s_r .

The repeatability variance is the within-laboratory variance pooled across all laboratories. For a given matrix/level experiment, the repeatability variance can be pooled with the following formula (5):

$$s_r^2 = \frac{\sum_{i=1}^p n_i (s_i^2 - \bar{s}^2)}{p-1}$$

where s_i^2 is the within-laboratory variance for the i th laboratory and is the mean squared deviation from the mean (POD) estimate for the i th laboratory, p is the number of laboratories, and n_i is the number of observations for the i th laboratory.

Reproducibility variance is the sum of laboratory variance and repeatability variance:

$$s_R^2 = s_r^2 + s_L^2$$

Step 2: Calculate S_L , SD due to the laboratory effect as:

$$S_L = \sqrt{\frac{\sum_{r=1}^R \sum_{l=1}^L (x_{lr} - \bar{x}_l)^2}{L-1}}$$

Step 3: Calculate $s(POD)$ as the SD of the individual laboratory POD estimates:

$$s(POD) = \sqrt{\frac{\sum_{i=1}^L (POD_i - LPOD)^2}{L-1}}$$

Step 4: Calculate degrees of freedom, df , for $s(POD)$ as follows:

$$df = \frac{\frac{s_L^2}{L} + \frac{s_r^2}{N}}{\frac{s_L^2}{L} + \frac{s_r^2}{N}}$$

Step 5: Calculate 95% confidence limits for LPOD:

If $0.15 < LPOD < 0.85$:

$$LCL = \max(0, LPOD - \frac{t_{0.975, df} s(POD)}{\sqrt{L}})$$

$$UCL = \min(1, LPOD + \frac{t_{0.975, df} s(POD)}{\sqrt{L}})$$

If $LPOD < 0.15$ or $LPOD > 0.85$:

$$LCL = \frac{x - 1.9207 + 1.9600 \sqrt{x \frac{x^2}{N} - 0.9604}}{N - 3.8415}$$

$$UCL = \frac{x + 1.9207 + 1.9600 \sqrt{x \frac{x^2}{N} - 0.9604}}{N - 3.8415}$$

where x is the number of observed positive outcomes and N is the total number of trials.

If $LPOD = 0$:

$$LCL = 0$$

$$UCL = 3.8415 / (N - 3.8415)$$

If $LPOD = 1$:

$$LCL = N / (N - 3.8415)$$

$$UCL = 1$$

Method for Estimating 95% Confidence Intervals for dLPOD

dLPOD is the difference between any two LPOD estimates, for example, to compare a candidate method to a reference method:

$$dLPOD_C = LPOD_C - LPOD_R$$

The associated 95% confidence interval (LCL, UCL) for the expected value of $dLPOD = LPOD_1 - LPOD_2$ is estimated by:

$$\begin{aligned} \text{LCL } dLPOD &= \sqrt{LPOD_1 - LCL_1^2 - LPOD_2 - UCL_2^2} \\ \text{UCL } dLPOD &= \sqrt{LPOD_1 - UCL_1^2 - LPOD_2 - LCL_2^2} \end{aligned}$$

Do not distribute