

Validation Scheme for Qualitative Analytical Methods

(possible alternative title: "Performance characteristics and validation of qualitative measurement methods")

Contents

Foreword.....	1
Introduction.....	2
1. Scope	2
2. Normative References.....	3
3. Terms and Definitions (to be incorporated in the definitions document eventually)	3
4. Practical implications of the definitions for accuracy experiments.....	3
4.1. Standard measurement method	3
4.2. Validation experiment	3
4.3. Nature of test items	4
4.4. Identical test items	4
4.5. Short intervals of time.....	4
5. Planning and conducting a validation experiment.....	5
5.1. Participating laboratories.....	5
5.2. Number of laboratories	5
5.3. Number of levels.....	5
5.4. Number of replicates per level.....	5
5.5. Observation conditions.....	6
5.6. Calculation of POD and dPOD Values from Qualitative Single-Laboratory Data	6
5.7. Calculation of POD and dPOD Values from Qualitative Multi-Laboratory Validation Data.....	7
5.8. Results of a validation experiment.....	10
6. Statistical model	10
6.1. Basic model	10
6.2. Relationship of qualitative model to the quantitative model.....	11
7. Literature.....	12

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is

normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO NNNNN-1 was prepared by Technical Committee ISO/TC 34/SC 16,

ISO NNNNN-1 consists of one part

Introduction

A qualitative analytical method must be validated as much as possible in the same way as it is intended to be used for routine analyses. Response of a method must be characterized with respect to concentration of the analyte. Concepts that have been used in the past have focused on parameters such as sensitivity, specificity, false positive, false negative, and are based on presence/absence of the analyte in the test sample. The limitation of this approach is that it assumes that the analyte in the non-zero concentration has a predictable response. In practice, a non-zero concentration may result in a variable probability of a positive response. Concentration of analyte as a continuous variable is a better predictor of measurement response than a two-state, zero/non-zero variable.

This standard combines sensitivity, specificity, false positive and false negative parameters into one single parameter, Probability of Detection (POD). This allows for comparison of probabilities across concentrations and further allows for a simple graphical representation of validation data as a POD response curve graphed by concentration with associated error bars. This approach expresses the probability of detection as dependent on concentration, as opposed to dependent on whether the sample contains or does not contain the analyte. The goal of validation will then be to characterize the response probability curve as a function of analyte mass or concentration.

While for most methods the probability of detection will remain at a nominal 100% once the concentration of analyte is sufficient, it must be recognized that greater concentrations may lead to reduced POD at higher concentrations with certain technologies. Examples of this effect may be "prozone" or saturation of antibodies in antibody-based applications, and inhibition by excess nucleic acid.

1. Scope

The purpose of ISO NNNN-N:NNNN is to outline the general principles to be used when characterizing response of qualitative methods of analysis.

This standard applies to methods that yield a binary result whose response can be characterized with respect to the concentration of some analyte.

2. Normative References

ISO 5725-1:1994

ISO 5725-2:1994

16140:2003(E)

Wehling et. al.,: Journal of AOAC International V94, No. 1, 2011 1-13

3. Terms and Definitions (to be incorporated in the definitions document eventually)

For the purposes of this standard, the definitions given in ISO 3534-1 and ISO 5725-1 apply except as specifically defined below.

Qualitative Method: Method of analysis which yields a binary result, and whose response can be characterized with respect to the concentration of some analyte.

Probability of Detection (POD): The probability of a positive response of a qualitative method at a given concentration, or POD.

4. Practical implications of the definitions for accuracy experiments

4.1. Standard measurement method

- 4.1.1. In order that the measurements are made in the same way, the measurement method shall have been standardized. All measurements shall be carried out according to that standard method. This means that there has to be a written document that describes in full detail how the measurement shall be carried out.

4.2. Validation experiment

- 4.2.1. The method performance parameters should be determined from a series of test results reported by the participating laboratory or laboratories, organized under a panel of experts established specifically for that purpose. Such an experiment is called a "validation experiment". The estimates of performance parameters derived from such an experiment should always be quoted as being valid only for tests carried out according to the standard measurement method.
- 4.2.2. A validation experiment can often be considered to be a practical test of the adequacy of the standard measurement method. One of the main purposes of standardization is to eliminate differences between users (laboratories) as far as possible, and the data provided by an accuracy experiment will reveal how effectively this purpose has been achieved. Pronounced differences between the laboratory means may indicate that the measurement method can possibly be improved.

4.3. Nature of test items

- 4.3.1. Validation of qualitative methods requires the use of known positive (Low and high POD) and negative (effectively zero POD) materials. For some bioanalytical methods the provision of this material may be challenging to achieve. Special challenges arise when biological material is being tested, and pure reference material (CRM traceable back to SI units) may therefore not be available as it is with small molecules. In such cases, the use of naturally incurred samples may be the best source of materials for the validation of such methods. Concentrations that may lead to reduced POD at higher concentrations should also be tested.

4.4. Identical test items

- 4.4.1. In a validation experiment, samples of a specific material or specimens of a specific product are typically sent from a central point to a number of laboratories. The definition of repeatability conditions stating that the measurements in these laboratories shall be performed on identical test items refers to the moment when these measurements are actually carried out. To achieve this, two different conditions have to be satisfied:
- a) the samples have to be identical when dispatched to the laboratories;
 - b) they have to remain identical during transport and during the different time intervals that may elapse before the measurements are actually performed.

In organizing accuracy experiments, both conditions shall be carefully observed.

4.5. Short intervals of time

- 4.5.1. According to the definition of repeatability conditions (xxxxx), measurements for the determination of repeatability have to be made under constant operating conditions. In particular, the equipment should not be recalibrated between the measurements unless this is an essential part of every single measurement. In practice, tests under repeatability conditions should be conducted in as short a time as possible in order to minimize changes in those factors, such as environmental, which cannot always be guaranteed constant.
- 4.5.2. There is also a second consideration which may affect the interval elapsing between measurements, and that is that the test results are assumed to be independent. If it is feared that previous results may influence subsequent test results (and so reduce the estimate of repeatability variance), it may be necessary to provide separate samples coded in such a way that an operator will not know which are replicates. Instructions would be given as to the order in which those specimens are to be measured, and presumably that order will be randomized so that all the "replicate" items are not measured sequentially. This might mean that the time interval between repeated measurements may appear to defeat the object of a short interval of time unless the measurements are of such a nature that the whole series of measurements could all be completed within a short interval of time. Common sense must prevail.

5. Planning and conducting a validation experiment

Guidance on conducting a validation experiment is given in ISO 5725-1 sections 6.1 and 6.2 (and in ISO TC 69/SC 6/WG 1: ISO/WD 15725-1, Date: 2010-12). Information particular to the organization of collaborative studies for qualitative methods is described in this standard. For information on performance of collaborative studies refer to ISO15725 sections xx.xx.

5.1. Participating laboratories

(from ISO TC 69/SC 6/WG 1: ISO/WD 15725-1, Date: 2010-12)

Laboratories participating in any validation study for qualitative methods should have experience in performing the type of qualitative study being trialed. However, the participating laboratories should not consist exclusively of those that have gained special experience during the process of standardizing the method. Neither should they consist of specialist "reference" laboratories in order to demonstrate the accuracy to which the method can perform in expert hands.

5.2. Number of laboratories

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.

Some of the variation will be due to variation of the method of detection in this fractional range. The model is a modification to improve coverage accuracy of the 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. However, Wehling et al. has shown that 10 laboratories are sufficient to carry out a collaborative study for a qualitative method.

5.3. Number of levels

The number of concentrations of the analyte to be tested(to be completed after discussion)

5.4. Number of replicates per level

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 $n \geq 12$ replicates/laboratory, the linear additive model approximates very well under the assumption of normality and provides for adequate parameter estimates in practice. The repeatability SD, s_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.

5.5. Observation conditions

- 5.5.1. The factors which contribute to the variability of the observed values obtained within a laboratory are listed in ISO 5725_____. They may be given as time, operator and equipment when observations at different times include the effects due to the change of environmental conditions and the recalibration of equipment between observations. Under repeatability conditions, observations are carried out with all these factors constant, and under reproducibility conditions observations are carried out at different laboratories; i.e. not only with all the other factors varying but also with additional effects due to the difference between laboratories in management and maintenance of the laboratory, stability checking of the observations, etc.
- 5.5.2. It may be useful on occasion to consider intermediate precision conditions, in which observations are carried out in the same laboratory but one or more of the factors time, operator or equipment are allowed to vary. In establishing the precision of a measurement method, it is very important to define the appropriate observation conditions, i.e. whether the above three factors should be constant or not. Furthermore, the size of the variability arising from a factor will depend on the measurement method. For example, in chemical analysis, the factors “operator” and “time” may dominate; likewise with microanalysis the factors “equipment” and “environment”, and with physical testing “equipment” and “calibration” may dominate.

5.6. 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):

$$\text{POD} = \frac{x}{N}, \quad \text{Where POD is } \text{POD}_C, \text{POD}_R, \text{ etc.}$$

The POD estimates and 95% confidence interval (LCL, UCL) estimates are given by:

For the case where $x = 0$,

$$\text{POD} = 0$$

$$\text{LCL} = 0$$

$$\text{UCL} = 3.8415 / (N + 3.8415)$$

For the case where $x = N$,

$$\text{POD} = 1$$

$$\text{LCL} = N / (N + 3.8415)$$

$$\text{UCL} = 1$$

For the case where $0 < x < N$,

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

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

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

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 $x \leq 1$, set $\text{LCL} = 0$. If $x \geq N-1$, 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. (See LaBudde[8].)

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

$$d\text{POD}_C = \text{POD}_C - \text{POD}_R$$

The associated 95% confidence interval (LCL, UCL) for the expected value of $d\text{POD} = \text{POD}_1 - \text{POD}_2$ is estimated by:

$$\text{LCL} = d\text{POD} - \sqrt{(\text{POD}_1 - \text{LCL}_1)^2 + (\text{POD}_2 - \text{UCL}_2)^2}$$

$$\text{UCL} = d\text{POD} + \sqrt{(\text{POD}_1 - \text{UCL}_1)^2 + (\text{POD}_2 - \text{LCL}_2)^2}$$

where $(\text{LCL}_1, \text{UCL}_1)$ is a 95% confidence interval for POD_1 and $(\text{LCL}_2, \text{UCL}_2)$ is a 95% confidence interval for POD_2 , as determined above.

5.7. Calculation of POD and dPOD Values from Qualitative Multi-Laboratory Validation Data

For a Multi-Lab trial where L = number of labs, R = replicates per lab, $N = LR$ = total replicates, LPOD estimate is given by

$$\text{LPOD} = \frac{x}{N}, \quad \text{Where } x \text{ is the number of positive results.}$$

5.7.1. 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]:

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

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.

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 , standard deviation due to lab effect as:

$$\sigma_R^2 = \sigma_L^2 + \sigma_r^2$$

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$$

$$s_L = \sqrt{s_R^2 - s_r^2}$$

Step 3: Calculate $s(\text{POD})$ as the standard deviation of the individual laboratory POD estimates.

$$s(\text{POD}) = \sqrt{\frac{\sum (\text{POD}_i - \text{LPOD})^2}{L-1}}$$

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

$$df = \frac{\left[\frac{s_L^2}{L} + \frac{s_r^2}{N} \right]^2}{\frac{\left[\frac{s_L^2}{L} \right]^2}{L-1} + \frac{\left[\frac{s_r^2}{N} \right]^2}{N-L}}$$

Step 5: Calculate 95% confidence limits for LPOD

If $0.15 \leq \text{LPOD} \leq 0.85$:

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

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

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:

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

$$UCL = 1$$

5.7.2. 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:

$$LCL = dLPOD - \sqrt{(LPOD_1 - LCL_1)^2 + (LPOD_2 - UCL_2)^2}$$

$$UCL = dLPOD + \sqrt{(LPOD_1 - UCL_1)^2 + (LPOD_2 - LCL_2)^2}$$

5.8. Results of a validation experiment

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

6. Statistical model

6.1. Basic model

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.

e is the random error occurring in every measurement under repeatability conditions.

6.1.1. Constraints in the model

In the qualitative model, there is a special case constraint to y for the binary case:

$$y \in \{0,1\}$$

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

$$0 \leq m \leq 1$$

$$-1 \leq B \leq 1$$

$$-1 \leq e \leq 1$$

6.1.2. General mean, m

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 $y \in \{0,1\}$ (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.

$$m = POD = P(+ | c)$$

6.1.3. Variance Parameters

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

$$\sigma_L^2 = \text{var}(B)$$

$$\sigma_W^2 = \text{var}(e)$$

$$\sigma_r^2 = \overline{\text{var}(e)}$$

$$\sigma_r = \sqrt{\overline{\text{var}(e)}}$$

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$$

6.2. Relationship of qualitative model to the quantitative model

The “qualitative model” is not a separate model distinct from the “quantitative model,” but a special case subset of the basic quantitative model (ISO 5725-1).

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. In this case the model follows a threshold detector that transforms the response to 0 or 1 values. With the coding convention of $y = \hat{1}\{0,1\}$ (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:

$$m = POD = P(+ | c)$$

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 $n \geq 12$ replicates/laboratory, the linear additive model approximates very well under the assumption of normality (Wehling et al., 2011.) 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.

7. Literature

ISO 5725-1:1994

ISO 5725-2:1994

ISO 16140:2003(E)

Wehling et. al., Journal of AOAC International V94, No.1, (2011) 1-13

DRAFT