TECHNICAL SPECIFICATION

ISO/TS 27878

First edition
2023-01

Reproducibility of the level of detection (LOD) of binary methods in collaborative and in-house validation studies

Reproductibilité de la limite de détection (LD) des méthodes binaires pour des études de validation internes et collaboratives

Reference number
ISO/TS 27878:2023(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Symbols **2**
5 General principles **3**
5.1 General considerations 3
5.2 Considerations for the conventional approach 3
5.3 Considerations for the factorial approach 3
6 Conventional approach **4**
6.1 Experimental design 4
6.2 Statistical model for methods for continuous measurands 4
6.3 Statistical model for methods for discrete measurands 7
6.4 Reliability of precision estimates 10
7 Factorial approach **10**
8 In-house validation **13**
9 Software **13**
Bibliography **15**

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

An appropriate approach for the validation of binary methods will often differ considerably from that of quantitative methods. Nevertheless, core concepts from the validation of quantitative methods can be successfully carried over to binary methods. In particular, the precision of a method – a performance characteristic usually associated with quantitative methods – can be determined for the level of detection (LOD) of binary methods.

In analytical chemistry, one of the fundamental indicators of method performance is the reproducibility of quantitative test results as described in ISO 5725 (all parts)[1]. This aspect of method performance is not usually taken into consideration in the validation of binary methods. However, in the last few years, novel validation approaches have been proposed in which the reproducibility of a binary method can be determined and meaningfully interpreted.

Why is it important to determine a method's reproducibility? In order to answer this question, consider an example from the field of microbiology. Take the case that, in the validation study, a method's LOD is determined as 3 CFU/ml (CFU = colony forming unit), but that the LOD is sometimes much higher depending on the laboratory or on the test conditions. Failing to detect the occasional unreliability of the method could lead to mistakes in routine laboratory determinations. On the other hand, if an LOD of 300 CFU/ml is obtained in the validation study, the method will not be validated even though this excessive LOD is not representative of its average performance. Accordingly, both the average LOD value and the reproducibility parameter – describing the variability of the LOD across laboratories or test conditions – capture important information about the performance of the method and should be determined in the course of the validation process.

In order to accomplish this, a suitable approach should be identified for the conversion of the binary results into quantitative ones. In this standard, two parametric models for the calculation of the LOD will be used: one model for methods for discrete measurands, e.g. microbiological and Polymerase Chain Reaction (PCR) methods, and one model for methods for continuous measurands, e.g. chemical methods.

Two different study designs will be applied. In the conventional approach, test conditions vary randomly from one laboratory to the other, whereas in the factorial approach, at least to some extent, test conditions are controlled. The factorial approach makes it possible to assess different sources of errors such as the variability arising in connection with different analysts, different instruments, different lots of reagents, different elapsed assay times, different assay temperatures etc. Such an approach also allows a reduction in workload and fewer participating laboratories.

Reproducibility of the level of detection (LOD) of binary methods in collaborative and in-house validation studies

1 Scope

This document provides statistical techniques for the determination of the reproducibility of the level of detection for

a) binary (qualitative) test methods for continuous measurands, e.g. the content of a chemical substance, and

b) binary (qualitative) test methods for discrete measurands, e.g. the number of RNA copies in a sample.

The reproducibility precision is determined according to ISO 5725 (all parts).

Precision estimates are subject to random variability. Accordingly, it is important to determine the uncertainty associated with each estimate, and to understand the relationship between this uncertainty, the number of participants and the experimental design. This document thus provides not only a description of statistical tools for the calculation of the LOD reproducibility precision, but also for the standard error of the estimates.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1 and ISO 5725-1 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
factor
binary or quantitative parameter within the method that can be varied at two or more levels within the limits of the specified method

EXAMPLE Technician.

3.2
factor level
value of the *factors* (3.1) within the experimental design

EXAMPLE Technician 1, Technician 2, etc.

3.3
level of detection
LOD
concentration from which on the *POD* (3.4) is not below a specified limit, e.g. 0,5 or 0,95 ($LOD_{50\%}$ or $LOD_{95\%}$).

Note 1 to entry: This definition is mathematically equivalent to the definitions for "level of detection" in ISO 16140-1[2], ISO 16140-2[3] and ISO 16140-4[4]. It differs from the definition used for chemical and physical methods for which a "limit of detection" is defined as the lowest quantity of an analyte that can be distinguished from the absence of that analyte with a stated confidence level.

Note 2 to entry: In this document, the term concentration (or concentration level) is used as a generic term to mean not only the actual concentration in the case of a measurand that can be quantified on a continuous scale, but also the number of colony forming units or DNA copies per aliquot in the case of measurands which are quantified on a discrete scale.

3.4
probability of detection
POD
probability of a positive analytical outcome of a qualitative test method at a given concentration for a specific sample type

Note 1 to entry: This definition is based on the two slightly different definitions for "probability of detection" in ISO/TS 16393[6] and ISO 16140-1, ISO 16140-2 and ISO 16140-4.

Note 2 to entry: The POD is a measure of the probability of a positive analytical result and thus a theoretical value which can be approximated by a mathematical model.

3.5
rate of detection
ROD
proportion of positive analytical outcomes in a test series, when a qualitative method is performed several times with a specific sample

Note 1 to entry: The ROD is not a theoretical value based on a mathematical model [like the *POD* (3.4)] but the result of a series of repeated tests performed on a given sample.

4 Symbols

p	number of participating laboratories
σ_L^2	between-laboratory variance
POD = P	probability of detection
x	concentration level (see Note 1 to entry 3.3) at which the POD is calculated
ROD	rate of detection
$LOD_{50\%} = L_{50}$	50 % of the level of detection
$LOD_{95\%} = L_{95}$	95 % of the level of detection
L, H, B, C	global model parameters for the four-parameter sigmoid curve

a_i laboratory-specific correction of laboratory *i* for the global inflection point C

$N\left(\mu,\sigma^2\right)$ normal distribution with mean μ and variance σ^2

5 General principles

5.1 General considerations

In order to ensure that tests are conducted in the same manner in all participating laboratories, the test method should be standardized. All tests forming part of an experiment within an individual laboratory or of an interlaboratory experiment shall be carried out according to the corresponding standardized protocol.

The statistical methods described in this document are applicable for binary test methods which yield a yes/no result (e.g. the substance of interest is present or absent). For such test methods, one of the main criteria of the method's fitness for purpose is the level of detection (e.g. $LOD_{50\%}$ or $LOD_{95\%}$), i.e. the (concentration) level required to ensure a POD of 50 % or 95 %. The aim is thus to determine LOD values for the individual laboratories as well as an overall LOD across laboratories. The precision of the method can then be evaluated in terms of the variability to which the laboratory-specific LOD values are subjected.

The laboratory-specific LOD values and the mean LOD across laboratories can be computed based on a mathematical model for the relationship between level, *x*, and probability of detection $\mathrm{POD}_i(x)=P_i(x)$ for laboratory *i*: The $LOD_{95\%}$ of laboratory *i* is then the lowest level, *x*, for which $\mathrm{POD}_i(x)=P_i(x)\geq 0{,}95$.

5.2 Considerations for the conventional approach

The conventional approach is based on the assumption that, according to the design used in ISO 5725-2, all tests are performed under repeatability conditions in each of the laboratories involved. In particular, all tests in the laboratory are performed by the same technician, with the same equipment, under the same conditions and directly one after the other. Test results are considered to have been obtained from different laboratories under reproducibility conditions, i.e. many factors contribute to observed variability, e.g. differences in equipment, environmental conditions, reagent batches or technician.

NOTE Validation protocols according to the conventional approach based on LOD and POD can be found in ISO 16140-2, ISO 16140-4 and ISO/TS 16393 and AOAC Guidelines[7]. Examples and further protocols are discussed e.g. in References [8][9][10][11][12] and [13].

5.3 Considerations for the factorial approach

Compared to the conventional approach, in which tests are made under repeatability conditions in each of the laboratories, the factorial approach systematically varies one or more factors. For instance, half the tests are conducted with reagents from batch A, and the other half with reagents from batch B. Thus, the factorial approach makes it possible to ensure the full spectrum of test conditions is covered in the validation study and assess contributions to variability from separate sources of error. This approach translates to more efficient and reliable estimation of the total variability.

NOTE Validation protocols based on LOD for microbiological methods according to the factorial approach are given in ISO 16140-4[4] and ISO 16140-5[5].

6 Conventional approach

6.1 Experimental design

Results from at least 8 participants, 4 concentration levels, and 8 replicates per level and laboratory are required to obtain a statistically reliable POD curve. However, with such a design, the reliability of the results may not be sufficient and will need to be checked. For more reliable estimation of the LOD and the corresponding variability, it is recommended that results from at least 8 participants, 5 concentration levels, and 12 replicates per level and laboratory are available. If the number of participants is increased, the number of replicates can be reduced.

The lowest concentration level should be selected so that no further reduction in POD is expected, even if the concentration level is further reduced. The highest concentration level should be selected in such a way that no further increase in POD is to be expected even if the concentration level is further increased. The expected proportions of positive test results across laboratories should be between 20 % and 80 % for at least two concentration levels.

The proportion of positive test results expected at the beginning of the collaborative trial usually differs from the final POD. This may mean that the proportion of positive test results actually determined in the collaborative trial does not meet the above requirements. In this case, the results of the evaluation and, in particular, the calculated reproducibility of the LOD can only be regarded as an estimate.

NOTE These recommendations for the experimental design are based on simulation studies in which the standard error of the estimate of the laboratory standard deviation was evaluated.

6.2 Statistical model for methods for continuous measurands

The calculation of the LOD is based on a generalized linear mixed-effects model (GLMM) together with a four-parameter sigmoid curve given by Formula (1):

$$\mathrm{POD}_i = P_i = \frac{L-H}{1+\left(\frac{x}{a_i C}\right)^B} + H \tag{1}$$

where

i denotes the laboratory (i = 1, 2,..., p);

$\mathrm{POD}_i = P_i$ denote the probability of detection for laboratory i;

x denotes a given concentration level;

L, H, B, C are global model parameters (i.e. they are valid across all laboratories);

a_i denotes the laboratory-specific correction of laboratory i;

C denotes the global inflection point C.

It is assumed that the parameters, L (lowest probability of detection), H (highest probability of detection), and B (slope) are the same for all laboratories. The product $a_i C$ describes the location of the inflection point of the curve for laboratory i; for L = 0 %, H = 100 %, it corresponds to the concentration at which a POD of 50 % is reached. The value of this product is thus a direct measure of the performance of the specific laboratory. The parameter, C, corresponds to the performance of an average laboratory.

The a_i values are modelled as realizations of a random variable: It is assumed that the ln a_i values follow a normal distribution with

$$\ln a_i \sim N\left(0, \sigma_L^2\right)$$

The parameters L, H, B, C and σ_L^2 can be provided by maximum likelihood estimation, e.g. in mathematical-statistical software package. The variance σ_L^2 characterizes the variability of sensitivity between laboratories.

NOTE 1 Although there is no guarantee that the distribution of ln a_i values actually follows a normal distribution, the log transformation usually leads to a better approximation of the normal distribution. If the method displays poor precision, then the prediction range of the LOD values without log transformation could include infeasible negative values.

NOTE 2 It is assumed that the parameters L, H, C and B are the same for all laboratories, i.e. that the shape of the curve is sigmoidal and the same across laboratories. It should be checked whether this assumption is justified, e.g. through a graphic check of laboratory-specific POD curves.

The interpretation of the parameters will be explained with an example, see Reference [13]. A collaborative study of a method for the binary analysis of gluten in corn products was conducted to demonstrate that the binary test method can detect gluten contaminations below the threshold of 20 mg/kg gluten. A total of four corn sample lots with different gluten concentrations was submitted to 18 laboratories to evaluate the sensitivity and reproducibility of the test method. Each of the 18 laboratories conducted 10 tests for each of four concentration levels. Table 1 provides the corresponding numbers of positive results per laboratory and concentration level.

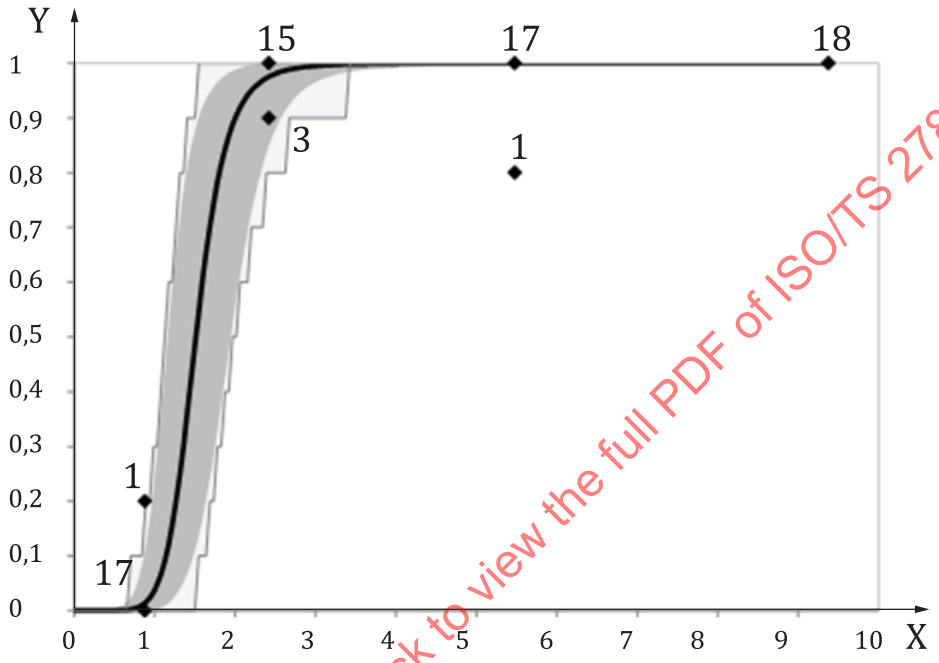
Table 1 — Number of positive test results per concentration level and laboratory (10 replicates)

Laboratory No.	Concentration level			
	0,88 mg/kg	2,42 mg/kg	5,48 mg/kg	9,38 mg/kg
01	0	10	10	10
02	0	10	10	10
03	0	10	10	10
04	0	10	10	10
05	0	10	10	10
06	0	10	10	10
07	0	10	10	10
08	0	9	10	10
09	0	10	10	10
10	0	9	8	10
11	0	10	10	10
12	0	10	10	10
13	0	10	10	10
14	0	10	10	10
15	0	9	10	10
16	0	10	10	10
17	0	10	10	10
18	2	10	10	10

Figure 1 shows the POD curve of a laboratory with average performance (solid line) along with 95 % prediction range of laboratory-specific POD (dark grey zone) and 95 % prediction range of laboratory-specific RODs (light grey step-functions). The numbers adjacent to the diamonds indicate the laboratory numbers having obtained the corresponding ROD.

For instance, at the concentration level 0,88 mg/kg, one laboratory has an ROD of 0,2, and 17 laboratories have an ROD of 0. Comparison with Table 1 shows that the laboratory with the ROD of 0,2 is laboratory 18. The light grey step-functions show the 95 % prediction range for the ROD values, obtained from simulation runs performed on the basis of the parameter estimates (Monte Carlo simulation). Figure 1 can be read as follows: a POD of 80 % is reached by a laboratory with an average performance at a concentration of about 1,7 mg/kg (solid line), whereas a top-performing laboratory will reach this POD at 1,3 mg/kg (upper dark grey zone) and a low-performing laboratory will need a concentration of about 2,2 mg/kg (lower dark grey zone).

NOTE 3 None of the selected concentration levels is within the 20 % to 80 % interval; therefore, the calculated reproducibility data can only be considered as an inaccurate estimate.



Key

X concentration, in mg/kg Y POD and ROD

Figure 1 — Mean POD curve, laboratory-specific RODs and prediction ranges

A special case of the model in 6.2 with $L = 0$, $H = 1$ and constant a_i value is equivalent to the logit model for $x > 0$. In other words, the logit model is already included in the model in 6.2. In practical terms, this statement also holds for the probit model, since it is very similar to the logit model, see e.g. Reference [14].

If continuous test results are available, the validation study should be based on these rather than on the corresponding binary results. In other words, insofar as binary results are obtained by comparing continuous test results to a threshold, the laboratories should submit the original continuous results, and the comparison with the threshold should be conducted as part of the validation study.

In many cases, the original continuous results will not be available, of course. In particular, in many cases, the assay yields a binary result, even though it is based on a continuous response.

Finally, it should be noted that the estimate of the between-laboratory variance σ_L^2 obtained from the binary results on the basis of the model described above is closely related to the between-laboratory standard deviation σ_L from ISO 5725-2. Indeed, if p laboratories each submitted two replicate LOD

values in a collaborative study, it would be possible to consider the σ_L estimate computed according to ISO 5725-2 to be equivalent to the σ_L estimate as computed here.

NOTE 4 Given an estimate for a variance (such as the between-laboratory variance estimate σ_L^2 mentioned above), the corresponding standard deviation is obtained by taking the square root.

6.3 Statistical model for methods for discrete measurands

In the case of measurands quantified on a discrete scale (e.g. microbiological culture methods or PCR methods), the four-parameter model discussed in 6.2 is no longer appropriate. The reason is the difference in distributional assumptions regarding the concentration, *x*. In 6.2, *x* denotes a nominal concentration level *per se*, and differences between the actual concentration of a test portion and the nominal concentration level can be assumed to be negligible. In the case of discrete measurands, *x* denotes e.g. the number of colony-forming units or DNA copies per test portion. For the sake of terminological convenience, these discrete quantities are referred to as concentration levels (see Note 1 to entry to definition 3.3) but, in the case of the discrete measurands considered here, differences between the *actual* concentration in a test portion and the nominal concentration can no longer be assumed to be negligible; rather, for a given nominal concentration level, the actual concentration levels of test portions are assumed to be subject to random variability and to follow a Poisson distribution. This assumption will be referred to in the following as the "Poisson assumption". For this reason, the cloglog (complementary-log-log) model is appropriate for the calculation of the LOD and its variability in the case of discrete measurands; accordingly, the following generalized mixed linear model (GLMM) is applied:

$$\ln\{-\ln[1-\mathrm{POD}_i(x)]\} = \ln a_i + b\ln x$$

where

- i denotes the laboratory (i = 1,2,...,p);
- x denotes a given concentration level;
- b is a global positive parameter that models the dependence of the sensitivity on the concentration level;
- a_i denotes the sensitivity corresponding to laboratory i.

NOTE 1 The Poisson assumption requires that $\mathrm{POD}_i(x) = P_i = 0$ for $x = 0$. This means that the above model should only be used if the number of false-positive results is negligible. Another consequence of the Poisson assumption is that $\mathrm{POD}_i(x) = P_i(x)$ will approach 1 with increasing x; in other words, the model is also susceptible to false negatives. This assumption constitutes an important difference to the four-parameter model discussed in 6.2, which admits both false positives and false negatives.

NOTE 2 The complementary log-log model is a standard model for microbiological methods and qualitative PCR. The model establishes a relationship between the probability of a positive result and the concentration, when a test portion is taken from a homogeneous sample. It is assumed that the probability of detecting an individual cell or DNA (RNA) copy does not depend on the concentration level. The probability of a positive result is then simply derived from the Poisson distribution: a qualitative result is positive if at least one cell or DNA (RNA) copy is detected.

It is assumed that the $\ln a_i$ values follow a normal distribution with

$$\ln a_i \sim N\left(\mu, \sigma_L^2\right), \qquad \mu = \ln a.$$

The three parameters a, σ_L^2 and b can be determined by maximum likelihood estimation in standard statistical software such as R. The parameter a represents the average sensitivity parameter (at x = 1) across laboratories and the variance σ_L^2 characterizes the variability of sensitivity between

laboratories. Note that predicted values for the ln a_i can be computed once the three curve parameters have been estimated.

For a derivation of the model, see Reference [15]. For further details (e.g. regarding the role of the parameter *b*), see Reference [16].

It is assumed that the parameters *a* and *b* are the same for all laboratories, i.e. that the shape of the curve is the same across laboratories. It should be checked whether this assumption is justified, e.g., by means of a graphic check of laboratory-specific POD curves.

Once the estimates for the three model parameters and the predictors have been computed, an overall POD curve, $\mathrm{POD} = P = 1 - \exp\left(-a \cdot x^{b}\right)$, can be plotted.

Consider the following example in Reference [17]. A collaborative study of a test method for the qualitative (binary) detection of genetic modifications in rice products was conducted in order to demonstrate that the binary test method can detect genetic modifications at a very low concentration level of the genetically modified rice (transgenic rice) in rice products. A total of six rice sample lots with different DNA copy numbers were submitted to 17 laboratories to evaluate the sensitivity and reproducibility of the test method. Each of the 17 laboratories conducted six tests for each of six concentration levels. Table 2 displays the corresponding numbers of positive results per laboratory and concentration level.

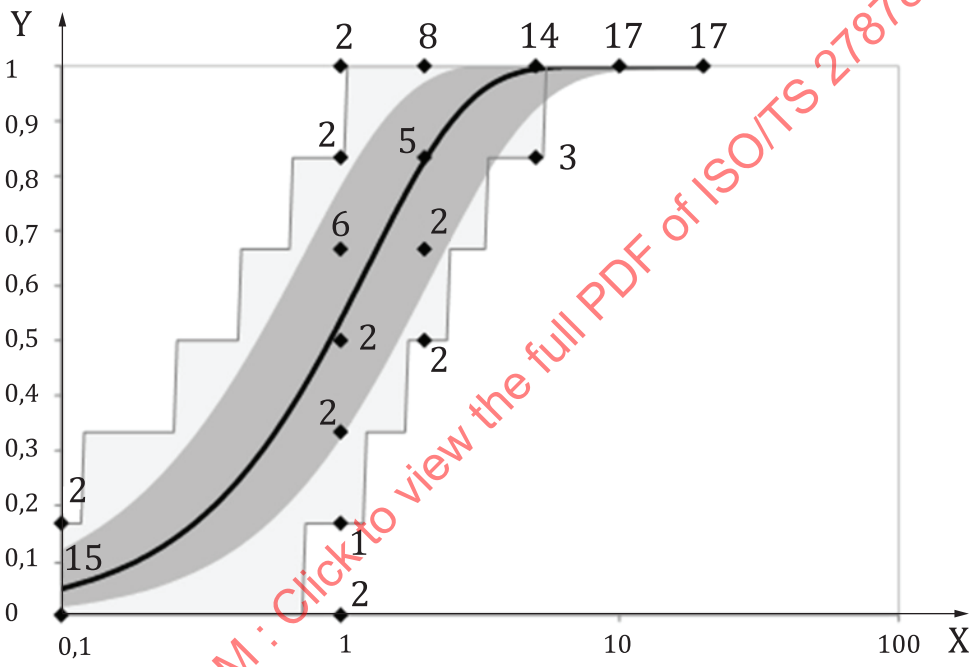
Table 2 — Number of positive test results per concentration level (number of copies per test portion) and laboratory (six PCR replicates)

Laboratory No.	Nominal number of DNA copies per test portion					
	0,1	1	2	5	10	20
01	0	3	5	5	6	6
02	0	4	6	6	6	6
03	1	0	5	6	6	6
04	0	4	3	6	6	6
05	0	0	5	6	6	6
06	0	4	6	6	6	6
07	0	5	6	6	6	6
08	0	5	6	6	6	6
09	0	6	4	6	6	6
10	0	2	5	6	6	6
11	0	1	6	6	6	6
12	0	4	6	6	6	6
13	0	4	4	6	6	6
14	0	3	3	5	6	6
15	0	6	5	6	6	6
16	0	2	6	6	6	6
17	1	4	6	5	6	6

Figure 2 shows the POD curve of a laboratory with average performance (solid line) along with predictions ranges for both the laboratory-specific POD curves and the concentration- and laboratory-specific ROD values. For the POD curves, the 95 % prediction range is displayed (dark grey zone), corresponding to the interval of typical variability for the POD curves. As far as the ROD values are concerned, 90 % prediction ranges are displayed (light grey step-functions). These ranges were obtained via simulation runs performed on the basis of the parameter estimates. The higher significance level for the ROD values (10 % rather than 5 %) reflects the fact that the focus lies on the identification of concentration-specific *conspicuous* ROD values, whereby "conspicuous" corresponds to a 5 % *one-sided* level of significance. Two-sided consideration is not possible as for low POD values the

lower prediction limit is 0, and for high POD values the higher prediction limit is 1; for low (high) POD it cannot be tested whether ROD values are significantly lower (higher) than expected. The use of the 95 % and 90 % prediction ranges (for POD curves and ROD values, respectively) is only a suggestion. Depending on the context, other values may be deemed more expedient.

The numbers adjacent to the diamonds indicate the laboratory numbers having obtained the corresponding ROD. For instance, at the concentration level 5 DNA copies per test portion, three laboratories have an ROD of 5/6 ≈ 8,33 and 14 laboratories have an ROD of 1. Comparison with Table 2 shows that the three laboratories with the ROD of 0,83 are laboratories 01, 14, and 17. Figure 2 can be read as follows: a POD of 50 % is reached by a laboratory with an average performance at a concentration level of about 1 DNA copy per test portion (solid line), whereas a top-performing laboratory will reach this POD at a concentration level of about 0,6 DNA copies per test portion (upper dark grey zone). A low-performing laboratory will need a concentration of about 1,2 DNA copies per test portion (lower dark grey zone).



Key

X number of DNA copies

Y POD and ROD

Figure 2 — Mean POD curve, laboratory-specific RODs and prediction intervals

The primary precision validation parameter is the standard deviation σ_L. This parameter has a simple interpretation: it represents the relative between-laboratory variability of the $LOD_{95\,\%}$. This standard deviation could also be calculated by means of a quantitative collaborative study according to ISO 5725-2 in which the measured characteristic would be the natural logarithm of the $LOD_{95\,\%}$. If each laboratory estimated its $LOD_{95\,\%}$ twice (using independent dilution series), the laboratory standard deviation of the $\ln(LOD_{95\,\%})$ values would be equal to the standard deviation σ_L derived from the complementary log-log approach described here. The advantage of the complementary log-log approach is its statistical efficiency: the number of required replicates is much lower.

Instead of characterizing the variability by σ_L, it would also be possible to provide the 95 % confidence interval for the mean LOD: first, compute the average sensitivity as $a = e^{\mu}$; then, compute

$\mathrm{LOD}_{95\,\%} = L_{95} = \frac{3}{a}$ (see NOTE below); finally, the 95 % confidence interval can be derived by applying the exponential function to both sides of the equation $\ln \frac{(\mathrm{LOD}_{95\,\%,\mathrm{upper}})L_{95,\mathrm{up}}}{(\mathrm{LOD}_{95\,\%,\mathrm{lower}})L_{95,\mathrm{lo}}} = 4 \cdot \sigma_L$.

NOTE 3 The identity $LOD_{95\,\%} = L_{95} = \frac{3}{a}$ is obtained as follows. Starting with the basic equation $\mathrm{POD} = P = 1 - \exp\left(-a \cdot x^b\right)$, and assuming b = 1, we have $0{,}95 = 1 - \exp\left(-a \cdot L_{95} = -a \cdot \mathrm{LOD}_{95\,\%}\right)$. After rearranging and taking the log of both sides, this yields $a \cdot \mathrm{LOD}_{95\,\%} = a \cdot L_{95} = -\ln(0{,}05) \approx 3$.

The complementary log-log approach described in this clause can be used only if the number of false-positive results is negligible. This assumption should be examined by a test series of blank samples.

6.4 Reliability of precision estimates

In order to assess the reliability of the reproducibility precision estimate, a simulation can be run either on the basis of resampling (bootstrapping) from the available measured data (say, 1 000 resamples) or on the basis of a Monte Carlo simulation, whereby the parameters for the underlying distributions are estimated from the available measured data (again, of the order of 1 000 simulation runs). For each simulation run or each resample, a precision estimate is then obtained. A confidence interval for the theoretical reproducibility can then be obtained by means of the 2,5 % and 97,5 % percentiles of the resulting distribution of precision estimates.

7 Factorial approach

The two models described in 6.2 and 6.3 can be expanded in order to accommodate the implementation of a factorial experimental design. Different influence factors are identified as probable sources of variability in this approach, e.g., different operators or reagent batches. These factors are then systematically varied in the design. Typically, each factor is varied across two levels, e.g., two operators or two different reagent batches. If five factors are included in the design, each with two levels, there are thus 2^5 = 32 different combinations or settings. Particularly efficient designs called orthogonal designs make it possible to reduce the number of settings, e.g. from 32 to 8. For further information on orthogonal designs, see References [15] and [18].

Typically, for each setting, replicate tests are carried out at different concentration levels (see Note 1 to entry to definition 3.3.). Let m denote the number of concentration levels. The term *run* refers to the performance of all the tests at the m concentration levels for one particular setting. In the case that n replicate tests are performed at each concentration level, there are thus $m \cdot n$ test results per run.

Taking into account the different runs, the model described in 6.2 is now expanded as follows:

$$\ln a_{ij} = \ln a_i + \eta_{ij}$$

where, as above, the subscript, i, represents the laboratory and $\ln a_i$ denotes a laboratory-specific (log) sensitivity effect. The subscript, j, represents the run, and $\ln a_{ij}$ denotes a laboratory- and run-specific (log) sensitivity effect. The laboratory-specific run effect η_{ij} can be expressed as the sum of factor effects

$$\eta_{ij} = \gamma_{i11} \cdot z_{j11} + \gamma_{i12} \cdot z_{j12} + \cdots + \gamma_{iq1} \cdot z_{jq1} + \gamma_{iq2} \cdot z_{jq2}$$

where γ_{ikl} is the effect of factor, k ($k = 1, \ldots, q$) in laboratory, i, for factor level, l, and z_{jkl} is the design matrix element (0 or 1) for run j, factor k, and factor level, l. Note that, as mentioned above, it is assumed that every factor has two levels.

The within-laboratory effects γ_{ikl} values are modelled as independent normal random effects with $\gamma_{ikl} \sim N\left(0, \sigma_k^2\right)$.

The variance components σ_k^2 $(k = 1, \ldots, q)$ can be estimated with mathematical-statistical software packages. Once they have been computed, the total variance is obtained as

$$\sigma_{\text{tot}}^2 = \text{Var}\left(\ln a_i + \eta_{ij}\right) = \sigma_L^2 + \sigma_1^2 + \ldots + \sigma_q^2 .$$

The σ_{tot}^2 parameter thus characterizes the reproducibility of the method.

Similarly, the model described in 6.3 is now expanded as follows:

$$\ln\left\{-\ln\left[1 - \text{POD}_{ij}(x) = 1 - P_{ij}(x)\right]\right\} = \ln a_i + b \cdot \ln x + \eta_{ij}$$

The η_{ij} terms have the same structure as in the model for methods for a continuous measurand, and the reproducibility is computed in the same manner. Details can be found in References [15] and [18].

Consider the following example, taken from the field of microbiology. Since colony-forming unit numbers per test portion can be assumed to follow a Poisson distribution, the cloglog model from 6.3 is applied. Five laboratories submit replicate binary results at three concentration levels: blank, near the presumed LOD (0,8 CFU/ml), and high (10 CFU/ml). The number of replicates depends on the concentration level: one replicate at the blank and high levels (where information regarding method performance is deficient), and four replicates at the "near LOD" level (where information regarding method performance is high).

Even though the test results obtained on the basis of the blank samples appear to contribute nothing to the estimation of the LOD and its reproducibility, it is necessary to include the blank level in the experimental design. Indeed, doing so constitutes a precautionary measure to ensure that the samples were indeed blank prior to inoculation. More importantly, however, it is essential to monitor whether false-positive test results are obtained. The reason is that the Poisson assumption (on which the mathematical model is based) is only admissible if the number of false positives is negligible.

Returning now to the example: within each laboratory, tests are carried out at all three concentration levels for eight different settings. Each setting corresponds to a combination of factor levels for the five levels selected for the experimental design. Table 3 provides the factorial design, showing the relationship between settings and factors. For each factor, the two-factor levels are coded "1" and "2". Design and test results are from ISO 16140-5:2020, Annex C.

Table 3 — Design with five factors and 8 settings to be implemented within each laboratory and for each concentration level

Setting	Factor 1 (Technician)	Factor 2 (Culture medium)	Factor 3 (Thawing process)	Factor 4 (Incubator)	Factor 5 (Background flora)
1	1	1	1	1	1
2		2	2	2	2
3		1	1	2	2
4		2	2	1	1
5	2	1	2	1	2
6		2	1	2	1
7		1	2	2	1
8		2	1	1	2

Table 4 provides the results for the five laboratories performing tests at three concentration levels per setting, for a total of eight settings. R1,...,R4 denote the replicates. While only one test result is obtained at the blank and high levels, four replicates are obtained at the concentration level of 0,8 CFU/ml.

Table 4 — Binary test results from five laboratories performing tests at three concentration levels per setting, for a total of eight settings

		Laboratory 1				Laboratory 2				Laboratory 3				Laboratory 4				Laboratory 5			
Contamination level [CFU/ml]	Setting	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
Blank	1	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	2	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	3	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	4	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	5	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	6	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	7	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	8	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
0,8	1	1	1	1	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0
	2	1	1	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	0	1	1
	3	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0
	4	1	1	1	1	1	0	1	0	1	0	0	0	0	1	0	1	1	1	0	1
	5	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0
	6	1	1	1	0	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	0
	7	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0
	8	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1
10	1	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	2	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	0	-	-	-
	3	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	4	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	5	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	6	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	7	1	-	-	-	1	-	-	-	0	-	-	-	1	-	-	-	1	-	-	-
	8	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-

On the basis of these data, the factorial precision estimates provided in Table 5 are obtained:

Table 5 — Factorial precision estimates

Factor	σ_k^2
Technician:	0,004 8
Culture medium:	0,099 7
Thawing process:	0,048 6
Incubator:	0,039 8
Background flora:	0,248 2
Lab (σ_L^2):	0,133 8
Total variance (σ_{total}^2):	0,574 9
Reproducibility standard deviation:	0,758 2

As can be seen from Table 5, the background flora, culture medium, and between-laboratory variance components dominate. $\text{LOD}_{50\%}$ is computed as 1,13 CFU/ml.

For further details regarding this example, see Reference [5].

In this example, the cloglog model from 6.3 was applied. However, the factorial approach can equally be used in the case of continuous measurands and the four-parameter model from 6.2. It is important to point out that, for the four-parameter model, to modify the design with respect to the number of nominal concentration levels: due to the larger number of parameters estimated in the model, at least five concentration levels should be included.

8 In-house validation

For in-house validation studies, the approach described in Clause 7 can be applied. The only difference is that, since there is only one laboratory, the laboratory-specific (log) sensitivity terms are no longer included. Accordingly, for methods for continuous measurand, the model is now

$$\ln a_j = \ln a + \eta_j$$

where η_j now represents the run-specific effect (a sum of factorial effects). For methods for discrete measurand, the model is now

$$\ln\left\{-\ln\left[1-\text{POD}_j(x)=1-P_j(x)\right]\right\}=\ln a + b\cdot\ln x + \eta_j .$$

The σ_{tot}^2 now corresponds to the in-house reproducibility variance, i.e., an intermediate precision characteristic.

Alternatively, the same approach as described in Clause 7 can be applied, with the index *i* representing no longer the laboratory but a singled-out factor to which the remaining factors are subordinate. For instance, the index *i* could now denote the factor day, and the remaining factors are nested within the factor day, i.e. that the corresponding random effects characterize variability within a given day. For further details, see Reference [15]. Factorial designs for microbiological methods are presented in ISO 16140-4[4].

9 Software

The calculations described in this document are based on standard models of statistics and can be realized with software packages where these standard models are available. Among them are standard