

Efficient estimation of the limit of detection and the relative limit of detection along with their reproducibility in the validation of qualitative microbiological methods by means of generalized linear mixed models

Steffen Uhlig¹  · Petra Gowik²

Received: 2 June 2017 / Accepted: 11 September 2017 / Published online: 27 October 2017
© The Author(s) 2017. This article is an open access publication

Abstract The reproducibility of measurement results is a core performance characteristic for quantitative methods. However, in the validation of qualitative methods it is not clear how to characterize a method's reproducibility. One approach for determining a qualitative method's reproducibility is presented for microbiological methods, where the distribution of colony forming units (CFU) follows a Poisson distribution. The method's reproducibility is defined in terms of the variability of the limit of detection (LOD) values. For a better estimation of reproducibility precision, our proposed approach is using an orthogonal factorial plan. Since an exact determination of absolute contamination levels is often not possible, following the ISO 16140-2:2016 [Microbiology of food and animal feed—method validation—part 2: protocol for the validation of alternative (proprietary) methods against a reference method, 2016], an approach is proposed which is based on the ratio of the LOD values of a reference and an alternative method. This approach is illustrated on the basis of an example.

Keywords Limit of detection · Relative limit of detection · Validation · Microbiology · Methods · Factorial study

1 Introduction

An appropriate approach for the validation of qualitative 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 qualitative methods. This paper shows how the reproducibility of a method—a performance characteristic usually associated with quantitative methods—can be determined in collaborative studies for qualitative methods in microbiology.

In analytical chemistry, one of the fundamental indicators of the performance of a quantitative method is the reproducibility of test results, as described in ISO 5725 (ISO 1994). While the concept of reproducibility is easily interpreted for qualitative methods in terms of consistent test results across laboratories for samples with the same level of contamination, it is not clear at all how to describe or characterize a qualitative method's reproducibility in such a way as to make possible a comparison to criteria or other methods. In the last few years, however, novel validation approaches have been proposed for the characterization of the reproducibility of a qualitative method (Uhlig et al. 2011, 2013, 2015; Grohmann et al. 2015).

Why is it important to determine a method's reproducibility? In order to answer this question, consider the case that a level of detection (LOD) of 3 colony forming units (CFU) per mL is determined in the validation study of a qualitative microbiological method, but that the LOD is sometimes much higher depending on the laboratory or measurement conditions. In such a case, failing to detect the occasional

✉ Steffen Uhlig
uhlig@quodata.de
Petra Gowik
petra.gowik@bvl.bund.de

¹ QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany

² Federal Office of Consumer Protection and Food Safety, Mauerstr. 3942, 10117 Berlin, Germany

unreliability of the method could lead to mistakes in routine laboratory determinations. On the other hand, if a LOD of 300 CFU/mL is obtained in the validation study, the method will not be accepted even if 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 measurement conditions—capture important information about the performance of the method and should be determined in the course of the validation process.

In the case of microbiological methods, an exact determination of absolute contamination levels is often not possible. For this reason, the ISO 16140-2 (ISO 2016) proposes an approach which is based on the ratio of the LOD values of a reference and an alternative method. Just as in the case of the LOD, both average and reproducibility precision parameters can be calculated for this relative LOD (RLOD) value.

In order to determine the reproducibility of a qualitative method, a suitable approach must be identified for the conversion of the qualitative results into quantitative ones. In this paper, the case will be considered where the distribution of CFU contamination levels follows a Poisson distribution. The reliability and robustness of the validation can be enhanced by means of a systematic study of the effect of influence factors. Such an approach also allows a reduction in workload, with reliable validation parameters with as few as 5 participating laboratories.

2 Materials and methods

The approach presented here is based on the computation of a power curve, which plots the probability of detection POD (probability that the target microorganism is detected) as a function of the contamination level x (in CFU/mL). The limit of detection $LOD_{95\%}$ or $LOD_{50\%}$ is then defined as the contamination level corresponding to $POD(LOD_{95\%}) = 0.95$ or $POD(LOD_{50\%}) = 0.5$.

In the case of the detection of target microorganisms, it cannot be assumed that, for a particular dilution level, the CFU contamination level is the same from one test sample to the next. In the context of a collaborative method validation, it is thus necessary to distinguish between the theoretical or nominal CFU contamination level and the unknown actual CFU contamination level in a given test

sample. The fundamental assumption is that, for a given nominal CFU contamination level, the actual contamination level in a particular test sample is subject to random variation and follows a Poisson distribution. More specifically, with x denoting the nominal CFU contamination level in CFU/mL, the probability that a test sample has a contamination level of k CFU/mL is

$$p_k = \frac{x^k}{k!} \exp(-x), \quad \text{for all } k = 0, 1, 2, 3, 4, \dots \quad (1)$$

On the assumption that every colony is detected, the probability of detection is thus

$$POD = 1 - \exp(-x). \quad (2)$$

This model is refined by introducing an extra parameter $0 \leq a \leq 1$ (referred to as the *sensitivity parameter*) to account for unsuccessful detection:

$$POD = 1 - \exp(-a \cdot x). \quad (3)$$

As can be seen, the *POD* increases with a . The value $a = 0$ corresponds to $POD = 0$ no matter the nominal number of copies (i.e. the method is useless), while, at the other extreme, the value $a = 1$ corresponds to $POD = 1 - \exp(-x)$ (i.e. the method is perfect).

Taking consecutive logarithms and rearranging, one obtains

$$\ln(-\ln(1 - POD(x))) = \ln a + \ln x. \quad (4)$$

This equation will now be expanded in two directions. First, in the framework of a collaborative validation study, it will be assumed that different laboratories have different sensitivities a_i and that $\ln a_i$ follows a normal distribution with

$$\ln a_i \sim N(\mu, \sigma_{lab}^2). \quad (5)$$

The parameter μ represents the average (log) sensitivity parameter across laboratories and the variance σ_{lab}^2 characterizes the variability of (log) sensitivity across laboratories. Accordingly, the model can now be written:

$$\ln(-\ln(1 - POD_i(x))) = \ln a_i + \ln x, \quad (6)$$

where the subscript i represents the laboratory. This POD model is known as a GLMM (generalized linear mixed model) with “complementary log–log” link function and is similar to the one described in (Uhlig et al. 2015).¹ For further information on generalized linear mixed models, the reader is referred to Nelder

¹ The model described here does not include the slope parameter, see ISO 16140-2 (ISO 2016). Indeed, it has been observed that, in the case of culture methods, the slope parameter can usually be omitted.

Table 1 Study design in the case of five factors for each participating laboratory

Setting	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
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 2 Summary of ROD values for each participating laboratory

Contamination level	Setting/run							
	1	2	3	4	5	6	7	8
L_0								
L_1								
L_2								

Table 3 Positive results for one method

Laboratory	Contamination level		
	L_0	L_1	L_2
Laboratory 1	k_{10}/n_0	k_{11}/n_1	k_{12}/n_2
Laboratory 2	k_{20}/n_0	k_{21}/n_1	k_{22}/n_2
Laboratory 3	k_{30}/n_0	k_{31}/n_1	k_{32}/n_2
Etcetera
All

k_{ic} denotes the number of positive results for Laboratory i and contamination level L_c , and n_c denotes the number of replicates for level L_c

and McCullagh (1983), McCulloch and Searle (2001) and Jiang (2007).

In practice, it may occur that a_i values greater than 1 are observed (or, equivalently, $\ln a_i \geq 0$). Theoretically, such an occurrence is not compatible with

the Poisson distribution assumption, since, for a given nominal concentration x , the corresponding POD would be greater than $1 - \exp(-x)$. Accordingly, it may seem desirable to constrain the sensitivity parameter estimates to values $a_i \leq 1$. However, $a_i > 1$ can be interpreted as an indication that the average target microorganism concentration is greater than the nominal concentration or that the number of false positives is too large. In the framework of a validation study, this constitutes useful information and, for this reason, it was decided not to build in an extra constraint (note that $a_i > 0$ is ensured by applying the exponential function to the $\ln a_i$ estimate).

The second model expansion consists in the implementation of a factorial experimental design. In this approach, different influence factors are identified as probable sources of variability, e.g. different operators or reagent batches. These factors are then systematically varied in the design. Typically, each factor is varied across 2 levels, e.g. 2 operators or 2 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. An example for an orthogonal design with 8 settings is provided in Table 1. For

Table 4 Design with five factors and eight settings to be implemented within each laboratory and for each contamination 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

further information on orthogonal designs, the reader is referred to Tamhane (2009).

Typically, for each setting, replicate measurements are carried out at 3 different contamination levels. The nominal contamination levels can be selected for instance as $L_0 = \text{Blank}$, $L_1 = 0.8 \text{ CFU/mL}$ and $L_2 = 10 \text{ CFU/mL}$. The term run will be used to refer to the performance of all the measurements at the 3 contamination levels for one particular setting. In case that n replicate measurements are performed at each contamination level, there are thus $3n$ test results per run. On the basis of the replicates, ROD (rate of detection) values are calculated. For instance, if k of the n results are positive (i.e. the organism was detected), then the ROD value is k/n . The results can be entered in a table such as Table 2, where each empty cell corresponds to one ROD value.

Alternatively, an overview of the data for one method across laboratories can conveniently be displayed in a table such as Table 3.

Taking into account the different runs, the model described by Eq. (6) is now expanded as follows:

$$\ln(-\ln(1 - \text{POD}_{ij}(x))) = \ln a_i + \ln x + \eta_{ij}, \quad (7)$$

where the subscript j represents the run, and the laboratory-specific run effect η_{ij} actually consists of a sum of factor effects $\eta_{ij} = \gamma_{i11} \cdot Z_{j11} + \gamma_{i12} \cdot Z_{j12} + \dots + \gamma_{iq1} \cdot Z_{jq1} + \gamma_{iq2} \cdot Z_{jq2}$, where γ_{ikl} is the effect of factor k ($k = 1, \dots, q$) in laboratory i for factor level l and Z_{jkl}

is the design matrix² element (0 or 1) for run j , factor k and factor level l (it is assumed that every factor has two levels).

Note that in a validation study, the design matrix elements are constants, i.e. they are not subject to random variation. They are systematically selected in order to reflect the spectrum of measurement conditions in the laboratory. However, in routine measurements no such deliberate control is exercised over measurement conditions, and the Z_{jkl} values can be seen as independent realizations of a random variable with zero mean and unit variance.

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

On the basis of the model described in Eq. (7), the variance components σ_{lab}^2 and σ_k^2 ($k = 1, \dots, q$) can be estimated in standard software such as R. Once they have been calculated, the total variance is obtained as

² The design matrix codifies which factor levels are associated with a particular test result. Thus, if there are 2 levels per factor, the design matrix contains zero and one ("0" for the one level and "1" for the other level). Note that one could also use a different coding strategy, such as coding the one factor level with "-1" and the other factor level with "1". The same results would be obtained, but some of the calculations would require slight adjustments [e.g. Eq. (8)].

Table 5 Data for example

Contamination level (CFU/mL)	Setting	Laboratory 1				Laboratory 2				Laboratory 3				Laboratory 4				Laboratory 5							
		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	-	-	-	0	-	-	-
	2	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	3	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	4	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	5	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	6	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	7	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-	0	-	-	-
	8	0	-	-	-	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	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	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	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	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	1	0	0	0
	6	1	1	1	0	1	1	1	0	0	0	1	1	0	1	0	1	0	0	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	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	1	1	1	1
10	1	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	2	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	0	-	-	-	0	-	-	-
	3	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	4	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	5	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	6	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	7	1	-	-	-	1	-	-	-	0	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-
	8	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-

Alternative method. Refer to Table 4 for the factor levels corresponding to each setting. R1, ..., R4 denote the test results. While only one test result is obtained at contamination levels L_0 and L_2 , 4 replicates are obtained at contamination level L_1

Table 6 ROD values for laboratory 1

Factor	Setting/run							
	1	2	3	4	5	6	7	8
Technician	1				2			
Culture medium	1	2	1	2	1	2	1	2
Thawing process	1	2	1	2	2	1	2	1
Incubator	1	2	2	1	1	2	2	1
Background flora	1	2	2	1	2	1	1	2
Contamination level (CFU/mL)								
Blank	0	0	0	0	0	0	0	0
0.8	0.75	0.75	0	1	0.25	0.75	1	0.25
10	1	1	1	1	1	1	1	1

Table 7 Data for example

Contamination level (CFU/mL)	Setting	Laboratory 1				Laboratory 2				Laboratory 3				Laboratory 4				Laboratory 5				
		R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	
Blank	1	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	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	
	2	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	1
	3	0	1	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	1	1	1	1
	4	1	0	0	0	1	0	1	1	1	0	1	1	0	0	1	1	0	0	0	0	1
	5	0	0	1	1	0	0	1	0	1	1	1	1	1	0	1	0	0	0	1	1	1
	6	0	1	0	1	0	1	1	0	1	0	1	0	1	1	0	0	1	0	1	1	1
	7	0	0	0	1	0	1	0	0	1	0	1	0	1	1	1	1	0	1	1	1	1
	8	1	1	0	1	0	0	0	0	0	1	1	1	0	0	0	0	1	0	1	1	1
10	1	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	
	2	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	
	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	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	
	8	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	1	-	-	-	

Reference method. Refer to Table 4 for the factor levels corresponding to each setting. R1, ..., R4 denote the test results. While only one test result is obtained at contamination levels L_0 and L_2 , 4 replicates are obtained at contamination level L_1

Table 8 Positive results for the alternative method

Laboratory	Contamination level		
	$L_0 = \text{Blank}$	$L_1 = 0.8 \text{ CFU/mL}$	$L_2 = 10 \text{ CFU/mL}$
Laboratory 1	0.000	0.594	1.000
Laboratory 2	0.000	0.688	1.000
Laboratory 3	0.000	0.250	0.875
Laboratory 4	0.000	0.281	1.000
Laboratory 5	0.000	0.344	0.875
All	0.000	0.431	0.950

$$\sigma_{total}^2 = \text{Var}(\ln a_i + \ln x + \eta_{ij}) = \sigma_{lab}^2 + \sigma_1^2 + \dots + \sigma_q^2 \tag{8}$$

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

As far as the interpretation of the sensitivity parameter a [see Eq. (3)] is concerned, note that by

definition of $LOD_{95\%}$ (and using $\ln 0.05 \cong -3$), it follows that

$$LOD_{95\%} = -\frac{\ln 0.05}{a} = \frac{3}{a} \tag{9}$$

This establishes a direct relationship between the average sensitivity a [calculated as e^μ , see Eq. (5)] and $LOD_{95\%}$. Thus, in the ideal case ($a = 1$), we obtain

Table 9 Positive results for the reference method

Laboratory	Contamination level		
	$L_0 = \text{Blank}$	$L_1 = 0.8 \text{ CFU/mL}$	$L_2 = 10 \text{ CFU/mL}$
Laboratory 1	0.000	0.406	1.000
Laboratory 2	0.000	0.344	1.000
Laboratory 3	0.000	0.625	1.000
Laboratory 4	0.000	0.406	1.000
Laboratory 5	0.000	0.563	1.000
All	0.000	0.469	1.000

$LOD_{95\%} \cong 3$. On the other hand, if the sensitivity parameter a drops to $1/2$, $LOD_{95\%}$ increases to $\cong 6$.

By the same token, one obtains

$$LOD_{50\%} = -\frac{\ln 0.5}{a} \cong \frac{0.69}{a}. \tag{10}$$

As far as the interpretation of the reproducibility parameter σ_{total}^2 is concerned, it is noted that Eq. (9) implies that $\ln LOD_{95\%} = \ln 3 - \ln a$. For the upper and lower limits of the 95 % prediction interval of the LOD estimate, it follows that

$$\ln LOD_{95\%,upper} = \ln 3 - (\ln a - 1.96 \cdot \sigma_{total})$$

and

$$\ln LOD_{95\%,lower} = \ln 3 - (\ln a + 1.96 \cdot \sigma_{total}).$$

Accordingly, one obtains

$$\ln \frac{LOD_{95\%,upper}}{LOD_{95\%,lower}} = 3.92 \cdot \sigma_{total}. \tag{11}$$

One obtains the same result with $LOD_{50\%}$ instead of $LOD_{95\%}$.

Thus, the (log) reproducibility variability of the $LOD_{95\%}$ (or $LOD_{50\%}$), defined as the logarithmic ratio between upper and lower 95 % confidence limits, is proportional to σ_{total} .

Simulation studies were conducted in order to assess the reliability of the σ_{total} estimate. With 5 participant laboratories, a relative standard error of less than 30 % was observed for the σ_{total} estimate. It can thus be concluded that reliable reproducibility estimates are achieved with as few as 5 laboratories.

Finally, it is important to take into account the fact that, in the case of microbiological methods, sufficient sample stability is difficult to achieve. As a result, the reliability of sensitivity and reproducibility estimates can be compromised. This difficulty can be overcome by including test results from a reference method in the validation study. Indeed, if, for each sample and laboratory, test results from both the

alternative method (i.e. the method being validated) and the reference method are obtained, then it is reasonable to expect that the instability of the samples will affect both methods in the same manner. In order to assess the performance of the alternative method, the study of the relative level of detection (in accordance with ISO 16140-2 (ISO 2016), Section 5.1.4) needs to be conducted. In this approach, a reliable indicator of the performance of the alternative method is obtained by determining the ratio of the LOD values corresponding to the 2 methods:

$$RLOD_{50\%} = \frac{LOD_{50\%,alt}}{LOD_{50\%,ref}}. \tag{12}$$

There are 2 approaches for the determination of the $RLOD_{50\%}$. If the contamination levels are not known, only a direct estimation of $RLOD$ is possible, see Section 5.1.4.2 of ISO 16140-2 (ISO 2016).

The mathematical model for the determination of $RLOD$ is derived from the model for LOD described above. In accordance with this model, we have

$$\ln LOD_{50\%,alt} \sim N(\ln 0.69 - \ln a_{alt}, \sigma_{alt}^2) \tag{13}$$

and

$$\ln LOD_{50\%,ref} \sim N(\ln 0.69 - \ln a_{ref}, \sigma_{ref}^2). \tag{14}$$

This implies

$$\ln RLOD_{50\%} \sim N(a_{ref} - a_{alt}, \sigma_{RLOD}^2), \tag{15}$$

where

$$\sigma_{RLOD}^2 = \sigma_{alt}^2 + \sigma_{ref}^2 - 2 \cdot \rho \cdot \sigma_{alt} \cdot \sigma_{ref} \tag{16}$$

and where ρ denotes the correlation between the two methods. If the 2 methods are independent, then we have $\rho = 0$, corresponding to the case of an “unpaired study” in the wording of ISO 16140-2 (ISO 2016). Thus, in the case of an unpaired study, σ_{alt}^2 can be obtained from σ_{RLOD}^2 if σ_{ref}^2 is available (e.g. from an earlier validation study).

Table 10 $\log_{10}RLOD_{50\%}$ values for each laboratory and setting

Laboratory	Setting	$LOD_{50\%,alt}$	$LOD_{50\%,ref}$	$\log_{10}RLOD_{50\%}$
1	1	0.69	4.60	-0.82
	2	0.69	0.80	-0.06
	3	4.60	0.80	0.76
	4	0.69	1.77	-0.41
	5	1.77	0.80	0.35
	6	0.69	0.80	-0.06
	7	0.69	1.77	-0.41
	8	1.77	0.69	0.41
2	1	0.80	0.80	0.00
	2	0.80	0.80	0.00
	3	0.69	4.60	-0.82
	4	0.80	0.69	0.06
	5	0.69	1.77	-0.41
	6	0.69	0.80	-0.06
	7	0.69	1.77	-0.41
	8	0.69	4.60	-0.82
3	1	1.77	1.77	0.00
	2	0.80	0.80	0.00
	3	1.77	0.69	0.41
	4	1.77	0.69	0.41
	5	1.77	0.69	0.41
	6	0.80	0.80	0.00
	7	4.60	0.80	0.76
	8	4.60	0.69	0.82
4	1	1.77	0.80	0.35
	2	0.80	4.60	-0.76
	3	4.60	1.77	0.41
	4	0.80	0.80	0.00
	5	1.77	0.80	0.35
	6	0.80	0.80	0.00
	7	1.77	0.69	0.41
	8	4.60	4.60	0.00
5	1	4.60	1.77	0.41
	2	2.75	0.80	0.54
	3	1.77	0.69	0.41
	4	0.69	1.77	-0.41
	5	4.60	0.80	0.76
	6	4.60	0.69	0.82
	7	4.60	0.69	0.82
	8	0.69	0.69	0.00

Note that the 5 $LOD_{50\%}$ values correspond to the number of positive results at the middle concentration level. Refer to Table 4 for the factor levels corresponding to the settings

3 Results and discussion

Five laboratories take part in an interlaboratory validation study for a newly developed culture method, referred to as the alternative method. The

laboratories obtain yes/no test results at 3 contamination levels and 8 settings. Within each laboratory and for each contamination level, the design provided in Table 4 is implemented.

The 3 nominal contamination levels taken into consideration in this study are $L_0 = \text{Blank}$, $L_1 = 0.8 \text{ CFU/mL}$ and $L_2 = 10 \text{ CFU/mL}$. For the 2 contamination levels L_0 and L_2 , only one test result is obtained. For the contamination level L_1 , 4 replicates are obtained. The test results are provided in Table 5. Table 6 provides the corresponding ROD values for one of the laboratories (see Table 2).

The estimation of the model parameters is carried out in the statistical software R. Alternatively, the computations can be performed by means of an extended version of the software PROLab POD (Quo-Data). The mean sensitivity estimate is 0.61. It follows that $LOD_{50\%}$ is approximately 1.13 [see Eq. (10)]. Finally, σ_{total} is estimated as 0.76. It follows that $\ln \frac{LOD_{50\%,upper}}{LOD_{50\%,lower}} = 3.92 \times 0.76 = 2.97$ [see Eq. (12)].

The σ_{total} estimate of 0.76 is relatively high and may constitute sufficient ground to call into question the fitness of the method. It is important to note, however, that the high reproducibility may be due, to some extent, to a lack of sample homogeneity. In order to investigate this question, the method is compared to a reference method. The test results for the reference method are provided in Table 7. Table 8 and 9 provide overviews across laboratories for the two methods (see Table 3).

The LOD of the reference method is calculated as $LOD_{50\%,ref} = 0.88$. As can be seen, it is lower than that of the alternative method ($LOD_{50\%,alt} = 1.13$), i.e. the reference method is more sensitive. The corresponding $RLOD_{50\%}$ value is calculated as 1.28 [see Eq. (12)].

The possible effect of sample instability is offset by considering not the reproducibility with respect to the LOD, but rather with respect to the RLOD. In order to determine the 2 methods' reproducibility with respect to $RLOD_{50\%}$, in a first step, for each laboratory and for each setting, maximum likelihood estimates for $LOD_{50\%,alt}$ and for $LOD_{50\%,ref}$ are calculated along with corresponding $\log_{10}RLOD_{50\%}$ values. In order to avoid implausible sensitivity estimates in case that all the test results are negative for a particular laboratory and setting, a minimum of 0.15 is stipulated for the sensitivity parameter a [see Eq. (3)] (this minimum corresponds to the upper confidence limit). The $\log_{10}RLOD_{50\%}$ estimates are provided in Table 10 (note that these estimates only take on five values, depending on the number of positive results at the middle concentration level). In a second step, a

linear mixed model is fitted to these $\log_{10}RLOD_{50\%}$ estimates. The resulting reproducibility is $\sigma_{total,RLOD} = 0.49$. This is considerably less than reproducibility estimate for the alternative method (calculated as 0.76). The $RLOD_{50\%}$ estimate obtained from fitting the linear mixed model to the $\log_{10}RLOD_{50\%}$ provided in Table 10 is 1.27, which matches well with the value 1.28 calculated directly from the $LOD_{50\%,alt}$ and $LOD_{50\%,ref}$ values according to Eq. (12) (see previous paragraph).

4 Conclusions

In this paper, a validation approach is presented for microbiological qualitative methods where the distribution of CFU contamination levels follows a Poisson distribution. In this approach, the method's reproducibility is a measure of the reproducibility of the LOD parameter across laboratories and measurement conditions. Since a microbiological qualitative method cannot be reliably validated without determining the variability of the LOD , the method's reproducibility—calculated as σ_{total} —provides essential information about the method's performance.

Moreover, the factorial design presented here constitutes a systematic approach to measurement conditions which, over and above ensuring the full range of measurement conditions is represented in the validation study, makes it possible to reduce the workload, with reliable reproducibility estimates with as few as 5 laboratories. In addition, the factorial approach also allows a quantitative analysis of the impact of different influence factors.

If, as is often the case for microbiological methods, sufficient stability of the samples is not ensured, then test results from a reference method should be taken into consideration, and the assessment of the reproducibility is carried out with respect to the two methods' relative level of detection. Since it can be expected that sample instability will affect both methods in the same manner, considering the ratio of the 2 LOD values should offset any bias in the estimate of reproducibility caused by sample instability. The reproducibility of the $RLOD$ parameter only provides information regarding the reproducibility of the LOD of the alternative and reference methods if the two measurement procedures can be

considered independent, e.g. involving different culture media, reagents and instruments.

Finally, it needs to be noted that the approach presented here can be adapted to in-house validation studies. The factor "Laboratory" can be replaced by the factor "Day" or "Week". The variability between the laboratories would then correspond to the variability between days or weeks.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Grohmann L, Reiting R, Mäde D, Uhlig S, Simon K, Frost K, Randhawa GJ, Zur K (2015) Collaborative trial validation of cryIAb/Ac and Pubi-cry TaqMan-based real-time PCR assays for detection of DNA derived from genetically modified Bt plant products. *Accredit Qual Assur* 20:85–96
- ISO 16140-2:2016 (2016) Microbiology of food and animal feed—method validation—part 2: protocol for the validation of alternative (proprietary) methods against a reference method
- ISO 5725 (1994) Accuracy (trueness and precision) of measurement methods and results—parts 1–6
- Jiang J (2007) Linear and generalized linear mixed models and their applications. Springer, New York
- McCulloch CE, Searle SR (2001) Generalized, linear, and mixed models. Wiley, New York
- Nelder JA, McCullagh P (1983) Generalized linear models. Chapman and Hall, Cambridge
- Tamhane AC (2009) statistical analysis of designed experiments: theory and applications. Wiley, Hoboken
- Uhlig S, Niewöhner L, Gowik P (2011) Can the usual validation standard series for quantitative methods, ISO 5725, be also applied for qualitative methods? *Accredit Qual Assur* 16:533–537
- Uhlig S, Krügener S, Gowik P (2013) A new profile likelihood confidence interval for the mean probability of detection in collaborative studies of binary test methods. *Accredit Qual Assur* 18:367–372
- Uhlig S, Frost K, Colson B, Simon K, Mäde D, Reiting R, Gowik P, Grohmann L (2015) Validation of qualitative PCR methods on the basis of mathematical-statistical modelling of the probability of detection. *Accredit Qual Assur* 20:75–83