Food Microbiology 84 (2019) 103163

Contents lists available at ScienceDirect

Food Microbiology

journal homepage: www.elsevier.com/locate/fm

Assessing the equivalence of *Vibrio parahaemolyticus* MPN and PCR quantification methods in oyster samples: A seven-year study

Samantha Lindemann[a], Bertrand Colson[b], Robert Newkirk[a], Steffen Uhlig[b], Ravinder Reddy[a,*]

[a] U. S Food and Drug Administration, Division of Food Processing Science and Technology, 6502 S. Archer Road, Bedford Park, IL, 60501, USA
[b] QuoData – Quality & Statistics, Prellerstr. 14, 01309, Dresden, Germany

ARTICLE INFO

Keywords:
Analytical methods
Equivalent
Method performance
Oysters
Shellfish

ABSTRACT

Data collected from FDA proficiency tests (PT) during 2012–2018 was used to evaluate the performance of most probable number (MPN) and polymerase chain reaction (PCR) methods used to enumerate *Vibrio parahaemolyticus* in oyster samples. The primary aim was to establish whether the MPN and PCR methods can be considered equivalent. The following criterion for equivalence was applied: the absolute value of mean bias and between-sample standard deviation must both be less than 0.1 (log10). Final calculations showed mean bias and between-sample standard deviation (SD) were 0.031 and 0.117 (log10) respectively. The between-sample SD criterion was slightly relaxed because with close to 700 results, the data set was large and overall mean bias was low. It was concluded that the two methods can be considered equivalent. The use of PT data for the assessment of method rather than laboratory performance is a secondary topic addressed in this paper. Important requirements for this use of PT data include availability of sufficient results for both methods and use of real food matrices. Ultimately, the results presented here provide an example of how PT data can be used to monitor method performance across many laboratories and samples as well as to assess method equivalence.

1. Introduction

Since 2006, three major *Vibrio parahaemolyticus* outbreaks across 13 states resulting in approximately 284 reported cases of illness have been documented (Centers for Disease Control and Prevention, 2017). *V. parahaemolyticus* has been identified as the leading cause of human gastroenteritis associated with raw seafood consumption (Kaysner and DePaola, 2001). Because of this, it is crucial that analytical laboratories can detect and quantify this pathogen, especially in seafood samples. The programs monitoring bivalve mollusc production worldwide include the EU Community Reference Laboratory for Monitoring of Bivalve Mollusc Production and the United States National Shellfish Sanitation Program (NSSP) (Lee and Murray, 2010). The NSSP is a cooperative program between state and federal agencies that is recognized by the U.S. Food and Drug Administration (FDA) and whose overall objective is to support the interstate exchange of sanitary shellfish including oysters, clams, mussels and scallops (FDA, 2015).

Laboratories servicing the NSSP are required to participate in the annual shellfish proficiency test (PT) administered by the FDA. In the framework of this PT, laboratories receive artificially contaminated oyster samples and the aim is to assess the analysts' ability to detect a variety of bacterial and coliform counts as well as *Vibrio* species. These samples are designed to simulate the samples regularly tested by NSSP laboratories. Commercially canned oysters are artificially inoculated with *V. parahaemolyticus* at levels ranging from 500-16,000 CFU/g. The range of inoculation levels reflects the FDA guidance for *V. parahaemolyticus* in seafood, which recommends that levels not exceed 10,000 viable cells per gram (Center for Food Safety and Nutrition, 2005). The samples provided for the shellfish PT can be tested using any analytical method routinely used by the participants. Laboratories participating in this PT typically enumerate *V. parahaemolyticus* using most probable number (MPN) or polymerase chain reaction (PCR) methods.

Among both commercial and regulatory laboratories, there is increased interest in the use of rapid or alternative methods like PCR due to the large number of samples analyzed daily. A rapid or alternative method is defined as any method or system that reduces the time taken to obtain a microbiological test result (Fung, 1994). Appealing characteristics of these methods often include partial or total automation, faster reading of results, and improved result interpretation (Glynn et al., 2006; Jasson et al., 2010). NSSP laboratories choosing to use alternative PCR methods over MPN benefit from simplified interpretation of results and decreased testing time. The *V. parahaemolyticus* MPN method typically involves using thiosulfate citrate bile salts sucrose

* Corresponding author.
E-mail address: Ravinder.Reddy@fda.hhs.gov (R. Reddy).

https://doi.org/10.1016/j.fm.2019.02.006
Received 2 November 2018; Received in revised form 8 February 2019
Available online 13 February 2019

(TCBS) agar plates, which can be overgrown with competing microflora in oyster homogenate resulting in increased difficulty when interpreting results (Nordstrom et al., 2007). This overgrowth is often due to heavy growth of sucrose-fermenting bacteria that appear as yellow colonies, while *V. parahaemolyticus* appears as round, opaque blue or green colonies on TCBS (Letchumanan et al., 2014). Using MPN methods to identify and quantify *V. parahaemolyticus* can take 7–10 days, whereas PCR methods can be completed within two days (Letchumanan et al., 2014). Completing this testing in a shorter time frame is not only beneficial for laboratories, but it can also expedite decision-making that may impact the need to close or open shellfish growing areas (Yeung and Boor, 2004).

Since many methods already approved for regulatory use have limited method validation data available, it is beneficial to use PT data in order to monitor the performance of methods across many laboratories and samples. Well-designed PT studies can be used for this purpose while simultaneously fulfilling quality control requirements of analytical laboratories. We propose that PT samples used to monitor the performance of analytical methods should have some key characteristics including that they must be prepared in real food matrices and have realistic inoculation levels. Moreover, although both MPN and PCR methods are approved for use in the quantification of *V. parahaemolyticus* in shellfish, it has never been established whether these two methods can be considered equivalent. For these reason, the decision was made to determine whether both objectives – long-term method performance monitoring and equivalence assessment – could be achieved based on the annual FDA shellfish PT rounds. Accordingly, the suitability of the design of the PT scheme was appraised. It was found that the PT rounds were conducted with real food matrices inoculated at realistic levels; that the experimental design allows both repeatability and reproducibility estimates to be computed; and that sufficient test results were available for each of the two methods. It was concluded that both the monitoring of long-term performance and the assessment of method equivalence could be reliably conducted based on the annual FDA shellfish PT data from the period 2012–2018. The present study thus constitutes an example of purposefully using PT data for method performance characterization.

It is important to emphasize that, in general, PT data is collected in order to assess laboratory performance – not method performance. Nevertheless, under certain circumstances, it is justifiable to use PT data for purposes of method characterization, as in the present paper. The conditions under which this use of PT data is permissible can be derived from the usual fitness-of-purpose considerations for the analytical methods under consideration. Here, the fact that real food matrices inoculated at realistic levels were used in the PT rounds played a crucial role. Indeed, taking these considerations a step further, certain aspects of the experimental design (such as the number of sample replicates) and the statistical analysis can be modified to allow the simultaneous assessment of both method and laboratory performance based on a single PT study data set (Horwitz, 1994).

Monitoring the performance of the two methods was done based on sample-specific precision parameters. Precision is defined as the closeness of agreement between independent test results obtained under stipulated conditions (ISO, 1994; ISO, 2002). In collaborative studies, the experimental design often allows the computation of reproducibility and repeatability precision estimates. The corresponding measures of precision are repeatability and reproducibility standard deviations (Taverniers et al., 2004). The repeatability standard deviation (s_r) measures the variability within one laboratory, while the reproducibility standard deviation (s_R) characterizes the total variability between measured values – including the between-laboratory variability component.

The equivalence test was conducted with the MPN method as the reference. For the assessment of equivalence, the following criteria were applied: the absolute value of the mean method bias[1] (across samples), and the between-sample standard deviation should both be less than 0.1 (log 10) (Lindemann et al., 2016). In addition, a Bland-Altman plot and an accuracy profile were used to compare the performance characteristics of PCR and MPN methods in the quantification of *V. parahaemolyticus* in the various shellfish samples.

2. Materials and methods

2.1. Study design

In each annual shellfish PT exercise, pureed oyster meat was artificially inoculated with a combination of *Vibrio parahaemolyticus*, *Vibrio cholerae*, *Vibrio vulnificus*, *Vibrio mimicus*, and *Vibrio fluvialis* (Table 1). For the purpose of this study, only those samples inoculated with *V. parahaemolyticus* are shown in Table 1. A total of 8 oyster shellfish samples – four pairs of blind duplicates – were sent to each lab. Participant numbers for analytical testing of *V. parahaemolyticus* in the shellfish PT rounds from 2012 to 2018 were 30, 35, 38, 45, 49, 32 and 38 analysts, respectively. Detailed instructions for test portion preparation, analytical methods and data collection were provided to the laboratories prior to each PT exercise.

2.2. Strains

Samples were artificially inoculated with a variety of *Vibrio* isolates in each annual round of the shellfish PT. Isolates were acquired from the FDA's Gulf Coast Seafood Laboratory (GCSL) on Dauphin Island, Alabama. These isolates were stored at −80°C, resuscitated in 10 mL alkaline peptone water (APW) (prepared in-house) and incubated at 35°C for 24–28 h prior to beginning transfers to prepare working stocks.

2.3. Oyster sample preparation

For each round of the shellfish PT, a 1:2 dilution of whole canned oysters to sterile simulated sea water (pH 8.2 and salinity 28–30 ppt) was pureed using a high capacity blender set to high for 3 min. Eight to 10 L of the oyster puree was then weighed into sterile, 16 L vats prior to the inoculation of blind duplicate samples. The 8–10 L bulk oyster puree samples were artificially inoculated using working stocks of liquid cultures which were 3 days old (Table 1). Working stocks of *V. parahaemolyticus*, *V. cholerae*, *V. vulnificus*, *V. mimicus*, and *V. fluvialis* were prepared by transferring 1 mL of resuscitated culture to 10 mL of APW. The APW was prepared in-house by mixing 10 g peptone with 10 g NaCl in 1000 mL distilled water, adjusting pH to 8.5 ± 0.1 and autoclaving. The high concentration of NaCl helps to inhibit growth of other bacteria (DePaola and Kaysner, 2004).

Blind replicate, 100 g test portions were dispensed from inoculated 8–10 L bulk samples for this PT annually. Bulk samples were inoculated according to Table 1, and all samples were prepared as blind duplicates excluding those from 2012. There were no blind replicate samples for 2012 because only four oyster puree samples were sent to participants prior to 2013. During inoculation, bulk samples were maintained in the temperature range of 8–10 °C while spinning on magnetic stir plates. Bulk samples were spun for at least 2 h after inoculation for homogenization of bacterial inocula. Prior to aseptically dispensing 100 g of sample to sterile testing bottles for distribution, homogeneity of bulk samples was tested by analyzing three sub-samples per vat in duplicate for total aerobic plate counts. Each of the participating laboratories received a 100 g test portion for samples 1–8 dispensed from bulk sample vats from 2012 to 2018. All samples were shipped chilled

[1] In ISO 5725-1 (Accuracy (trueness and precision) of measurement methods and results), trueness is defined as the closeness of agreement between a test result and the accepted reference value (ISO, 1994). Bias, or the difference between the test results and accepted reference values, is one measure of trueness (ISO, 1994).

Table 1
Summary of bacterial composition for Shellfish PT across 6 annual rounds.

Sample no.[a]	Year	Bacterial Composition	Total bacteria[b]	Total *VP*[b,c]
4	2012	*V. parahaemolyticus, V. vulnificus, V. cholerae, V. mimicus*	4.1	3.7
5	2012	*V. parahaemolyticus, V. mimicus*	3.3	3.0
1/7	2013	*V. parahaemolyticus, V. vulnificus, V. cholerae, V. mimicus*	4.3	3.6
3/5	2013	*V. parahaemolyticus, V. cholerae,*	3.8	2.9
4/8	2013	*V. parahaemolyticus, V. vulnificus*	4.0	3.6
1/7	2014	*V. parahaemolyticus, V. vulnificus, V. cholerae, V. mimicus, V. fluvialis*	3.7	3.1
3/5	2014	*V. parahaemolyticus, V. cholerae*	3.3	3.1
4/8	2014	*V. parahaemolyticus, V. vulnificus, V. fluvialis*	3.3	2.9
2/6	2015	*V. parahaemolyticus, V. mimicus, V. fluvialis*	3.5	3.1
3/5	2015	*V. parahaemolyticus, V. cholerae*	3.3	3.1
4/8	2015	*V. parahaemolyticus, V. vulnificus, V. fluvialis*	3.5	2.9
1/6	2016	*V. parahaemolyticus, V. vulnificus, V. cholerae, V. fluvialis*	4.2	3.0
2/8	2016	*V. parahaemolyticus, V. cholerae, V. mimicus*	4.3	4.2
3/5	2016	*V. parahaemolyticus, V. vulnificus, V. fluvialis*	4.1	3.9
1/6	2017	*V. parahaemolyticus, V. vulnificus, V. cholerae, V. fluvialis*	4.0	2.7
2/8	2017	*V. parahaemolyticus, V. cholerae, V. mimicus*	4.1	3.9
3/5	2017	*V. parahaemolyticus, V. vulnificus, V. fluvialis*	4.0	3.6
1/6	2018	*V. parahaemolyticus, V. cholerae, V. mimicus*	4.1	3.9
3/8	2018	*V. parahaemolyticus, V. vulnificus, V. cholerae, V. fluvialis*	4.0	2.7
4/7	2018	*V. parahaemolyticus, V. vulnificus, V. fluvialis*	3.9	3.6

[a] Sample numbers 1–8 are repeated annually for each round of the Shellfish PT. Therefore, samples are differentiated based on Sample No. and Year.
[b] All values are $\log_{10}$ CFU/g.
[c] *VP* = *V. parahaemolyticus*.

overnight in insulated containers. Each package contained a "temperature control" with 100 g of uninoculated oyster puree and laboratories were asked to take the temperature of this control upon arrival and report the results.

2.4. Analytical methods

This PT is method-independent, meaning participants are allowed to use those methods routinely employed in their laboratories. Participants report which analytical methods were used for screening, detection and enumeration. For this study, test results obtained via most-probable-number (MPN) and polymerase chain reaction (PCR) methods for the enumeration of *V. parahaemolyticus* were grouped together for the statistical analysis and method comparison. Laboratories reporting an MPN enumeration method typically reported use of the Bacteriological Analytical Manual (BAM) Chapter 9: *Vibrio* MPN method or MFLP-37 Detection and Enumeration of *Vibrio* species in Food. The BAM method involves a ten, five or three-fold serial dilution in APW with overnight enrichment and culturing turbid tubes to TCBS (DePaola and Kaysner, 2004). TCBS is a selective agar with 0.8% ox bile, 1% NaCl and pH 8.6 designed to inhibit the growth of other gram-positive organisms (DePaola and Kaysner, 2004). PCR methods for *V. parahaemolyticus* quantification included the National Shellfish Sanitation Program (NSSP) qPCR for *Vibrio* method as well as BAX *Vibrio* real time PCR. Reported confirmation methods included API 20E, VITEK, and PCR.

All participants were instructed to dilute 20 g oyster puree from each sample into 80 mL of phosphate buffered saline to prepare a 1:10 dilution (oyster puree is received by participants as a 1:2 dilution). After preparing the 1:10 oyster puree dilution, participants were instructed to go forward with a three-tube, five-dilution MPN or proceed with the selected PCR method. The NSSP qPCR method is applied in an MPN format (Nordstrom et al., 2007). Instructions also included details on incubating samples at 35 ± 2 °C for 18–24 h and request for participants to use the BAM-MPN calculator.

2.5. Statistical evaluation

Prior to the statistical analysis, all data were $\log_{10}$ transformed. Outliers were removed if an explanation of the deviating results was available. The Q/Hampel algorithm was used to calculate sample-specific mean values, reproducibility standard deviations (s_R), and repeatability standard deviations (s_r) (ISO, 2015a). Q/Hampel is described in ISO 13528:2015 as the robust algorithm with the best performance in terms of breakdown point and efficiency (ISO, 2015a). In this algorithm, the mean value estimate is obtained with the Hampel estimator. When using the Hampel estimator, the influence which a particular value has on the computed assigned value is downweighted by applying a rule that compares laboratory deviations to the reproducibility standard deviation. The latter is obtained with the Q method, in which the dispersion estimate is computed based on the 25th percentile of the pairwise differences between the individual measured values.

The statistical analysis was performed using 672 sets of PT results – each consisting of two replicate measured values (for all but the two 2012 samples) – corresponding to 1287 data points. The aim was to compare and monitor the performance of the MPN and PCR methods for the quantification of *V. parahaemolyticus*. As far as the comparison of the two methods is concerned, different statistical approaches were applied: the Bland-Altman plot, an equivalence test and the accuracy profile. In all three approaches, PCR was considered the alternative method while MPN was defined as the reference method.

In the Bland-Altman plot, the focus lies on the sample-specific bias and variability across samples or testing conditions (Bland and Altman, 2007). For a given sample, the difference between the two mean values (across laboratories) corresponding to the two methods is referred to as the bias (Hanneman, 2008). A positive bias shows that the alternative method yields higher results than the reference method (Hanneman, 2008). In the Bland-Altman plot, the sample-specific bias values are plotted against mean concentration values. In addition, limits of agreement (LoA) are displayed, corresponding to the mean bias ± 1.96 times the corresponding standard deviation (Bland and Altman, 2007). It is expected that 95% of all results will fall within the LoA (Bland and Altman, 2007). Values lying outside the LoA can thus be seen as displaying an excessive bias. However, it should be noted that the LoA values do not constitute a criterion for the equivalence of the two methods. Instead, the LoA values simply reflect the variability present in the data. In the "classical" Bland-Altman plot, the standard deviation (across samples) of the bias values serves as the basis for the LoA values. This standard deviation reflects two main error components:

Table 2
Summary statistics by method for *Vibrio parahaemolyticus* in oyster samples for six annual PT rounds (2012–2018). Estimates for means and standard deviations were computed with the Q/Hampel method. More than 10% of MPN results of the samples for the year 2013 lie above the quantification range, so that mean and precision could not be determined.

Sample[a]	MPN					PCR			
	Year	n	Mean	S_R	S_r	n	Mean	S_R	S_r
4	2012	22	2.966	0.977		5	2.872	0.668	
5	2012	24	3.280	1.110		6	3.350	0.704	
1/7	2014	33	2.449	0.590	0.284	6	2.188	1.279	0.304
3/5	2014	31	2.243	0.722	0.366	6	1.975	1.227	0.332
4/8	2014	31	2.253	0.458	0.294	6	2.098	1.191	0.354
2/6	2015	36	2.599	0.614	0.453	9	2.562	0.715	0.499
3/5	2015	36	2.329	0.982	0.512	9	2.399	0.540	0.264
4/8	2015	36	2.343	0.811	0.352	9	2.289	0.522	0.331
1/6	2016	33	2.167	0.956	0.342	14	2.481	0.761	0.261
2/8	2016	35	3.080	0.759	0.453	14	3.675	0.498	0.371
3/5	2016	34	3.295	0.494	0.494	14	3.439	0.590	0.261
1/6	2017	26	2.484	0.698	0.597	5	2.252	1.110	0.350
2/8	2017	26	3.445	0.591	0.309	5	2.868	0.816	0.265
3/5	2017	27	3.353	0.568	0.374	5	2.713	1.401	0.454
1/6	2018	36	2.777	0.866	0.434	2	3.372	0.018	0.000
3/8	2018	35	1.929	1.086	0.269	2	2.447	0.595	0.121
4/7	2018	36	2.610	0.547	0.497	2	3.005	0.785	0.785

All values are $\log_{10}$ CFU/g.
[a] Sample numbers 1–8 are repeated annually for each round of the Shellfish PT. Therefore, samples are differentiated based on Sample No. and Year.

reproducibility error and between-sample deviations. Since the focus lies on the investigation of variability between samples in the Bland-Altman plot, it is sensible to take only the between-sample error component into consideration in the computation of the LoA values (note that the standard deviation corresponding to the between-sample error component is less than the standard deviation across samples in the Bland Altman plot). Doing so yields a "corrected" Bland-Altman plot which makes it possible to check whether the uncertainty intervals (corresponding to the reproducibility error) of the individual bias values intersect the agreement range. The classical Bland-Altman plot was created using Analyse-it for Microsoft Excel (Analyse-it, Leeds, UK). The corrected Bland-Altman plot was created based on the results of a random effects model – details will be provided in a subsequent publication.

The computations carried out to create the corrected Bland-Altman plot are also useful for the equivalence test itself. This test follows the procedure elaborated within the Proficiency Testing Online Platform (www.quodata.de) and a related version of this test is implemented in PROLab Plus software (QuoData, Dresden, Germany). In the equivalence test, the focus shifts from the sample-specific bias values to the overall bias across matrices/samples. For the equivalence of PCR (alternative) to MPN (reference), two criteria were specified: 1) the absolute value of the overall bias across samples should be significantly less than 0.1 $\log_{10}$ CFU/g and 2) the between-sample standard deviation (excluding the reproducibility variance component, as explained above) should be less than 0.1 $\log_{10}$ CFU/g – at least for high CFU/g levels. If both criteria are met, it is expected that the bias of the alternative method will lie within ± 0.296 $\log_{10}$ CFU/g for 95% of the samples. In the case of lower CFU/g levels, higher random variability may be observed for the reference method as well as the alternative method, therefore these criteria may be relaxed.

Finally, the accuracy profile is a "fit for purpose" graphical, decision making tool used to check the requirement that the prediction range for the method bias is less than a pre-specified acceptability limit (Feinberg, 2007; ISO, 2016). Unlike the Bland-Altman plot, where the focus lies on between-sample variability, the accuracy profile is intended to assess each sample-specific bias against a pre-defined criterion. According to ISO 16140-2:2016, the accuracy profile is plotted by first calculating the following: medians and SDs for each method, bias, β-expectation tolerance interval (ETI) limits (i.e. prediction limits for the bias between the two methods), and acceptability limits. The upper and lower β-ETI limits are calculated according to the following formula: $\beta_i \pm T \cdot s_{alt}\sqrt{1 + 1/n}$ where β_i is the bias, T is the 10% quantile of a Student-t distribution with 24 degrees of freedom (following ISO 16140-2:2016 β is set at 80%), s_{alt} is the standard deviation of the alternative method and n is the number of replicates. The acceptability limits are calculated using the following: ± 3.3 s_{ref} where s_{ref} is the standard deviation of the reference method. When plotted, the β-ETI limits should lie within the acceptability limits on the accuracy profile. It should be noted, however, that in order to achieve more reliable results, the robust Q/Hampel estimator was used in the computation of the statistical parameters for accuracy profile shown in this paper. The computations were performed by means of the PROLab software package (QuoData, Dresden, Germany).

2.6. Data exclusion

While no formal outlier tests were used, the Bland-Altman plot and accuracy profile showed results from 2013 displayed a higher degree of variability than those results from the other years analyzed. Additionally, more than 10% of MPN results from 2013 samples 1/7 and 3/5 lay above the quantification range. These inconsistent results from 2013 may have been due to variable shipping conditions. Previous data from this annual PT have indicated oyster samples spiked with *Vibrio* species need to be maintained as close to 9 °C as possible during shipping. In 2013, some laboratories indicated the temperature of their samples exceeded 9 °C and these temperature fluctuations may have impacted overall analytical results. Since 2013, no major shipping issues have been documented. For these reasons, all the results from 2013 were excluded from the equivalency analysis. Although the use of robust statistics like the Q/Hampel method often circumvent the need to apply outlier tests, it is still important to examine the data in order to make an informed decision regarding which data to include in the analysis.

3. Results and discussion

3.1. Summary statistics

A total of 537 analysts submitted data for MPN and 119 analysts submitted data for PCR (Table 2). Summary statistics provided in Table 2 including mean value, reproducibility standard deviation (s_R), and repeatability standard deviation (s_r), were used to monitor the performance of the two methods over time and across samples separately for each method. The difference between s_r of MPN and PCR across samples ranged from −0.434 to 0.288 $\log_{10}$ with a mean s_r difference of −0.072 $\log_{10}$ CFU/g. For s_R, the difference ranged from −0.848 to 0.833 $\log_{10}$ CFU/g with a mean s_R difference of 0.035 $\log_{10}$ CFU/g. These results indicate that the repeatability (s_r), or variability within one laboratory, was slightly lower for PCR compared to MPN. The reproducibility (s_R), or variability across all laboratories, was slightly higher for PCR compared to MPN. These results indicate that measured values from a single analyst for PCR are more precise than MPN, but the variability from analyst to analyst is somewhat higher for PCR than MPN. More PCR variability from analyst to analyst, or lab to lab, may be attributed to the use of different DNA extraction methods or differences in experience across analysts.

3.2. Bland-Altman plot

The Bland-Altman plot displayed in Fig. 1 shows 20 data points representing results from all samples analyzed across the seven years of data. The mean bias (0.0647 $\log_{10}$ CFU/g) as well as 95% LoA are also displayed. One of the 20 data points lies outside the LoA. This

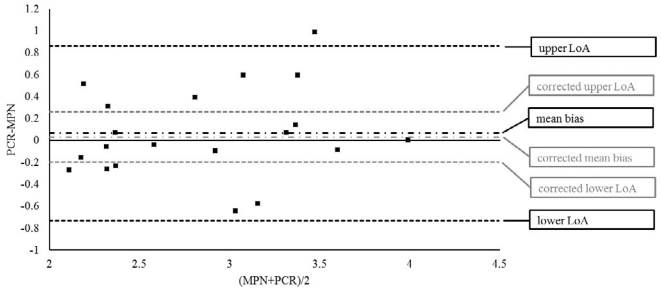

Fig. 1. Bland-Altman plot for *V. parahaemolyticus* quantitative MPN and PCR for the annual Shellfish PT from 2012 to 2018. Mean bias and Limits of Agreement (LoA) using the classical Bland Altman approach are shown in black while corrected mean bias and LoA values are shown in grey.

corresponds to the expectation of the classical Bland-Altman plot that no more than 1 in 20 data points lie outside the LoA (Bland and Altman, 2007).

Also shown in the Bland-Altman plot in Fig. 1 are corrected values for mean bias as well as upper and lower LoA. (Please refer to the Materials and Methods Section for an explanation of these corrected values). It should be noted that in the classical Bland-Altman plot, the LoA are much wider than in the corrected plot and the mean bias of 0.0647 $\log_{10}$ CFU/g is higher than the corrected value of 0.031 $\log_{10}$ CFU/g. Eleven data points fall outside the corrected LoA.

3.3. Analysis of equivalency

The mean bias across all samples analyzed was computed as 0.031 $\log_{10}$ CFU/g and the between-sample standard deviation was computed as 0.117 $\log_{10}$ CFU/g (Table 3). The mean bias value can be seen as an indication that there are no large deviations between the two methods within the population of samples analyzed. The between-sample standard deviation value exceeds the specified threshold. This indicates some risk of sample-specific bias. However, considering that the mean bias is very low, and the evaluation was based on six years' worth of data (excluding 2013), the results demonstrate a sufficient degree of equivalence between the two methods. In particular, the bias will lie within the limits −0.20 $\log_{10}$ CFU/g and 0.26 $\log_{10}$ CFU/g for 95% of samples.

Table 3
Analysis of equivalency for V. *parahaemolyticus* data.

Sample[a]	Year	Bias
4	2012	−0.094
5	2012	0.07
1/7	2014	−0.261
3/5	2014	−0.268
4/8	2014	−0.155
2/6	2015	−0.037
3/5	2015	0.07
4/8	2015	−0.054
1/6	2016	0.314
2/8	2016	0.595
3/5	2016	0.144
1/6	2017	−0.232
2/8	2017	−0.577
3/5	2017	−0.64
1/6	2018	0.595
3/8	2018	0.518
4/7	2018	0.395
Mean bias across samples		0.031
Sample standard deviation (corrected for random error)[b]		0.117

All values are $\log_{10}$ CFU/g.

[a] Sample numbers 1–8 are repeated annually for each round of the Shellfish PT. Therefore, samples are differentiated based on Sample No. and Year.

[b] This corrected sample standard deviation estimate was computed by means of a new approach which will be described in a subsequent paper. The confidence interval for this estimate is [0, 0.376].

3.4. Accuracy profile

The accuracy profile shows that the beta-ETI limits (prediction range for the bias) lie outside the acceptability range for 4 samples, 2 of which corresponds to the year 2013 (Fig. 2). This confirms that the samples from 2013 are anomalous, as previously explained. The two other samples for which the acceptability criterion was not met are 2017 3/5 and 2018 3/8. For these two samples, there were very low numbers of participants who submitted results for PCR; accordingly, the wider prediction range is not unexpected. All other sample results meet the expectations of the accuracy profile. It should also be mentioned that the samples from 2012 are not displayed due to the lack of replicate results.

4. Conclusion

Overall, the results from evaluating the Shellfish PT data indicate that the performance of MPN and PCR for quantifying *V. parahaemolyticus* in oyster samples can be considered equivalent. The Bland-Altman plot demonstrated that the bias between the two methods was relatively low. The majority of sample data also met the requirements for accuracy profile analysis. The equivalence test indicated that the mean method bias of 0.031 $\log_{10}$ CFU/g was acceptable. Although the between-sample standard deviation exceeded the specified threshold of 0.100 $\log_{10}$ CFU/g slightly, these two methods can be considered equivalent while acknowledging some risk of sample-specific bias. Because these two methods can be considered equivalent and are both already approved for use in the NSSP, method selection criteria should be applied by each participating laboratory. Selection criteria may include analysis time, limit of detection (LOD), and total cost. The MPN method takes 7–10 days to complete with a reported LOD of 3 MPN/g while the PCR method can be completed in two days and has an LOD of 1 CFU/reaction (Blackstone et al., 2003; Cruz et al., 2015; Duan and Su, 2005; Letchumanan et al., 2014). Overall costs for the MPN and PCR methods have been found to be very comparable (WHO, 2016). Therefore, analysis time and LOD values are the most suitable method selection criteria in this case and a method should be selected based on the individual requirements of the participating laboratories.

These results also illustrate how to assess method equivalency and analytical method performance by combining evaluation approaches from different international standards. Finally, this paper also shows how data originally collected in the framework of proficiency testing can be used for purposes of method performance characterization – such as long-term performance monitoring and equivalence assessment.

Declarations of interest

None.

Funding

This research was funded by the United States Food and Drug Administration.

Acknowledgements

We would like to extend a special thank you to Dr. Jessica Jones and Dr. William Burkhardt from FDA's Gulf Coast Seafood Laboratory in Dauphin Island for their helpful feedback. A special thank you is also owed to Linda Chandler who has provided expertise, advice and guidance about the NSSP for many years. We'd also like to express our gratitude to all laboratories that continue to participate in this proficiency test each year.

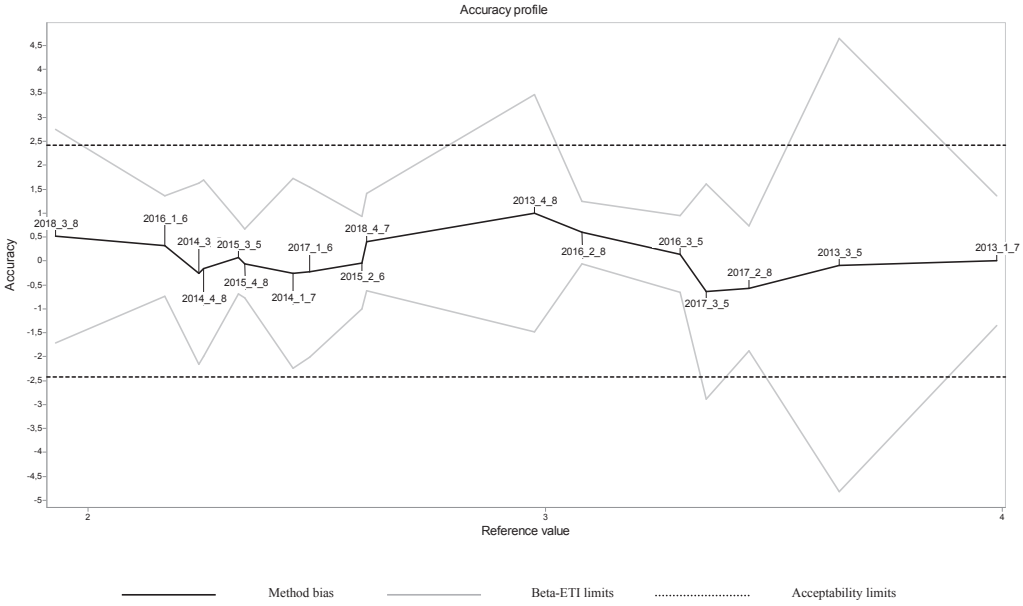

Fig. 2. Accuracy profile comparing *V. parahaemolyticus* PCR to MPN enumeration methods for the annual Shellfish PT from 2012 to 2018. Sample labels are displayed in the following format: "Year_blind replicate 1_blind replicate 2".

References

Blackstone, G.M., Nordstrom, J.L., Vickery, M.C., Bowen, M.D., Meyer, R.F., DePaola, A., 2003. Detection of pathogenic Vibrio parahaemolyticus in oyster enrichments by real time PCR. J. Microbiol. Methods 53, 149–155.

Bland, J.M., Altman, D.G., 2007. Agreement between methods of measurement with multiple observations per individual. J. Biopharm. Stat. 17, 571–582.

Center for Food Safety and Nutrition, U.S Food and Drug Administration, 2005. Quantitative Risk Assessment on the Public Health Impact of Pathogenic Vibrio Parahaemolyticus in Raw Oysters. Online. https://www.fda.gov/downloads/Food/FoodScienceResearch/UCM196914.pdf, Accessed date: 13 June 2018.

Centers for Disease Control and Prevention, 2017. Vibrio species causing Vibriosis, selected outbreaks. online. https://www.cdc.gov/vibrio/index.html, Accessed date: 29 June 2018.

Cruz, C.D., Hedderley, D., Fletcher, G.C., 2015. Long-term study of vibrio parahaemolyticus prevalence and distribution in New Zealand shellfish. Appl. Environ. Microbiol. 81 (7), 2320–2327.

DePaola, A., Kaysner, C.A., 2004. Vibrio., Bacteriological Analytical Manual. U.S. Food and Drug Administration, Washington DC online. https://www.fda.gov/food/foodscienceresearch/laboratorymethods/ucm070830.htm, Accessed date: 26 June 2018.

Duan, J., Su, Y.C., 2005. Occurrence of Vibrio parahaemolyticus in two Oregon oyster--growing bays. J. Food Sci. 70, M58–M63.

Feinberg, M., 2007. Validation of analytical methods based on accuracy profiles. J. Chromatogr. A 1158, 174–183.

Fung, D.Y., 1994. Rapid methods and automation in food microbiology: a review. Food Rev. Int. 10, 357–375.

Glynn, B., Lahiff, S., Wernecke, M., Barry, T., Smith, T.J., Maher, M., 2006. Current and emerging molecular diagnostic technologies applicable to bacterial food safety. Int. J. Dair. Technol. 59, 126–139.

Hanneman, S.K., 2008. Design, analysis, and interpretation of method-comparison studies. AACN Adv. Crit. Care 19, 223–234.

Horwitz, W., 1994. Nomenclature of interlaboratory analytical studies (IUPAC Recommendations 1994). Pure Appl. Chem. 66, 1903–1911.

ISO, D.I.N., 1994. 5725-1: Accuracy (Trueness and Precision) of Measurement Methods and Results– Part 1: General Principles and Definitions. Beuth, Berlin.

ISO, D.I.N., 2002. 5725-2: Accuracy (Trueness and Precision) of Measurement Methods and Results Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method (12.02). Beuth, Berlin.

ISO, D.I.N., 2015. 13528: Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparison. Beuth, Berlin.

ISO, D.I.N., 2016. 16140-2: Microbiology of the Food Chain-Method Validation Part 2: Protocol for the Validation of Alternative (Proprietary) Methods against a Reference Method. Beuth, Berlin.

Jasson, V., Jacxsens, L., Luning, P., Rajkovic, A., Uyttendaele, M., 2010. Alternative microbial methods: an overview and selection criteria. Food Microbiol. 27, 710–730.

Kaysner, C.A., DePaola, A., 2001. Vibrio, Compendium of Methods for the Microbiological Examination of Foods, fourth ed. American Public Health Association, Washington, DC, pp. 405–420.

Lee, R., Murray, L., 2010. Components of Microbiological Monitoring Programmes, Safe Management of Shellfish and Harvest Waters. World Health Organization, London, UK, pp. 91–108.

Letchumanan, V., Chan, K.G., Lee, L.H., 2014. Vibrio parahaemolyticus: a review on the pathogenesis, prevalence, and advance molecular identification techniques. Front. Microbiol. 5, 705.

Lindemann, S., Kmet, M., Reddy, R., Uhlig, S., 2016. Matrix-specific method validation of an automated most-probable-number system for use in measuring bacteriological quality of grade "A" milk products. J. Food Protect. 79, 1911–1918.

Nordstrom, J.L., Vickery, M.C., Blackstone, G.M., Murray, S.L., DePaola, A., 2007. Development of a multiplex real-time PCR assay with an internal amplification control for the detection of total and pathogenic Vibrio parahaemolyticus bacteria in oysters. Appl. Environ. Microbiol. 73, 5840–5847.

PROLab Plus software package. online. https://www.quodata.de/en Accessed June 13, 2018..

Taverniers, I., De Loose, M., Van Bockstaele, E., 2004. Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance. Trac. Trends Anal. Chem. 23, 535–552.

U.S Food and Drug Administration, 2015. National shellfish sanitation program guide for the control of Molluscan shellfish. online. https://www.fda.gov/downloads/food/guidanceregulation/federalstatefoodprograms/ucm505093.pdf, Accessed date: 13 June 2018.

World Health Organization & Food and Agriculture Organization of the United Nations, 2016. Selection and Application of Methods for the Detection and Enumeration of Human-Pathogenic Halophilic vibrio Spp. In Seafood: Guidance. World Health Organization online. , 2019.

Yeung, P.S., Boor, K.J., 2004. Epidemiology, pathogenesis, and prevention of foodborne Vibrio parahaemolyticus infections. Foodb. Pathog. Dis. 1, 74–88.