Accreditation and Quality Assurance
https://doi.org/10.1007/s00769-026-01717-6

RESEARCH

# A control-chart approach for sample homogeneity verification in food and water microbiology proficiency testing schemes

**Colin Howell[1] · Thanh Thuy Nguyen[1] · Nicolas Choquet[1] · Sandrine Nguyen[1] · Anne Tirard[1] · Romain Le Neve[1] · François Le Nestour[2] · Abdelkader Boubetra[1]**

Received: 23 April 2026 / Accepted: 9 June 2026


**Abstract**
Verifying the homogeneity of the samples used for a proficiency test is a normative requirement to ensure that the variation in the results is due to the laboratories' analyses rather than the heterogeneity of the test material. A general method based on the assessment of between-sample and within-sample (analytical) variance components of 10 samples analyzed in duplicate according to the recommendation of the ISO 13528 standard can be strict for microbiological analyses. This paper proposes an alternative approach to evaluating sufficient sample homogeneity of a new proficiency test based on standard deviation control charts plotted for successive homogeneity and proficiency tests in previous rounds. The proposed method is based on the observed stability in the variation in colony counts at different levels of contamination in food and water microbiology proficiency tests organized by Bureau Interprofessionnel d'Études Analytiques (BIPEA). However, the control limits are adapted to different test microorganisms and matrices due to their distinctive characteristics. Examples of these control charts for the enumeration of microorganisms in water, meat, and wine over the 2017 to 2024 period are presented.

**Keywords** Control chart · Sample homogeneity control · Proficiency testing schemes · Food and water microbiology

## Introduction

Proficiency testing is an important tool that enables analytical laboratories to assess and ensure the validity of their results according to the requirements of the ISO/IEC 17025:2017 standard [1]. Fundamentally, proficiency tests (PTs) involve laboratories performing the same analyses on identical samples and having their results evaluated by the organizer according to pre-established criteria. It is therefore essential that PT providers be able to produce homogeneous and stable samples; this allows certainty that different results between participants can be attributed to the laboratories' analyses rather than to the heterogeneity of the test material.

Microbiology is one of the most significant testing areas for analytical laboratories, since microorganisms are ubiquitous in the environment and some can have harmful, wide-reaching impacts on human health. These threats can additionally originate from a variety of sources, including contact with environments, such as water or surfaces that contain pathogenic bacteria, or direct consumption of contaminated food. By participating in proficiency tests, laboratories carrying out microbiological analyses integrate external controls into their quality management procedures in accordance with regulatory requirements.

While the overall framework of proficiency testing is independent of the matrices and analytes involved, there are specific challenges associated with microbiological analyses. General guidelines and obligations for PT providers are stipulated in the ISO/IEC 17043:2023 standard, but specific requirements and considerations for proficiency testing of microbiological examinations are detailed in ISO 22117:2019 [2, 3]. As living organisms, microorganisms can present increased difficulty for accurate analysis, particularly for quantitative methods. For example, microorganisms can behave differently within the same material, whether dying, reproducing, or clumping, leading to a number of microorganisms that is too low or too high, unpredictable distribution, and difficulties in the determination of a "true" value for the concentration of a test microorganism. In

✉ Colin Howell
chowell@bipea.org

[1] Bureau Interprofessionnel d'Études Analytiques (BIPEA), 189 Rue d'Aubervilliers, 75018 Paris, France

[2] Laboratoire Microsept, Z.A. La Sablonnière, 15 Rue Denis Papin, 49220 Le Lion d'Angers, France

Published online: 18 June 2026

addition, results can be both method-dependent, as analytical procedures used to quantify a specific target microorganism impact the entire environment of target and non-target microorganisms in the test material, and operator-dependent [4–6]. The reproducibility of results is thus not consistent for the same target microorganism, matrix, or contamination level over repeated tests.

Given the uncertainty in microbiological testing, sample homogeneity control in food and water microbiology proficiency testing schemes can be challenging. The general requirements and procedures for sample homogeneity testing are described in the ISO/IEC 17043, ISO 13528 and ISO 22117 standards. The conclusion of sufficient homogeneity of the test material is based on the statistical assessment of the variance between the samples, selected at random from the same batch [2, 3, 7]. Under the ISO 13528 standard, homogeneity testing is recommended to be performed on 10 samples, analyzed in duplicate and conducted under repeatability conditions. Then, one-way analysis of variance is run to estimate the between-sample and within-sample (analytical) variance of the obtained measurements, expressed in colony-forming units (CFU) on a log-transformed scale. The material is considered sufficiently homogeneous if the standard deviation between the homogeneity test (HT) samples, $s_s$, is less than 30 % of the standard deviation used for laboratory proficiency assessment, $\sigma_{pt}$. When this criterion is satisfied, $s_s$ contributes less than 10 % to the variance of the PT results, which is considered an acceptable limit [7–9]. The evaluation criterion can be calculated to include both between-sample and within-sample variance, using the F-test, when the within-sample standard deviation, $s_w$, is smaller than 50 % of $\sigma_{pt}$. The limits were recommended for use in chemistry proficiency testing and then standardized for PTs in other areas, including microbiological analysis [3, 7, 8].

However, this technique has some limitations. Firstly, unless $\sigma_{pt}$ is determined prior to the analysis of the PT results, it remains unknown at the time the HT is conducted. Secondly, $s_s$ is calculated by the square root of a difference which tends to be negative when $s_s$ is smaller than $s_w$. In this case, the between-sample variance is expected to be small and is considered equal to 0 in the ISO 13528 standard [9, 10]. This statistical approach is likely to result in a low risk of failing to account for the heterogeneity of the samples in the calculation of the standard deviation for proficiency assessment. Thirdly, the probability of falsely rejecting homogeneous HT samples is high when the $\sigma_{pt}$ estimated from the participating laboratories' results is small. In such a scenario, $s_s$ may be greater than 30 % of $\sigma_{pt}$ although the PT results show little variation and the samples are considered sufficiently homogeneous. Fourthly, due to the inherent biological variability of microorganisms, the heterogeneous (random or contagious) distribution of microorganisms, and the variation in test material composition, test results are sensitive to the analytical methods used for each target microorganism, as well as to the type of matrices. Fifthly, replicate analysis is not always applicable because of the destructive nature of microbiological testing. For example, after water is passed through a membrane filter, it is not possible to recover the filtered volume [4]. Furthermore, microorganisms can multiply or die between duplicate analyses, leading to large differences in results even though the samples are homogeneous. Finally, the number of HT samples is usually small as homogeneity testing involves high cost, lowering the power of statistical tests for homogeneity [11].

Alternative methods of sample homogeneity verification should therefore be considered to avoid the rejection of good samples when the standard evaluation criteria are not met. As homogeneity testing is performed in the same laboratory, applying the same procedure on the same number of samples of similar test materials, it is feasible to apply a statistical process control. Standard deviation (SD) control charts plotted for successive homogeneity and proficiency testing rounds can be a fit-for-purpose approach to homogeneity testing in PT schemes. The variability of the results is comparable as they are obtained by the same testing procedure on similar test materials. The trend analysis of standard deviations and mean colony counts of HT and PT results over time allows for evaluation of sample homogeneity of a new round for the same target microorganism, matrix, or contamination level.

A control-chart approach for monitoring sample homogeneity verification is based on the observed stability in the variation in colony counts for different levels of contamination in food and water microbiology proficiency tests organized by BIPEA, an accredited PT provider. This paper aims to plot the control charts for standard deviations of HT and PT results round by round and against the mean colony counts of individual rounds. Examples of control charts for PTs studying the enumeration of microorganisms in water, meat, and wine over the 2017 to 2024 period are presented to illustrate the standard deviations and control limits for different microorganisms, matrices, and contamination levels.

## Methodology

The data used to build the control charts were BIPEA's HT and PT results of three schemes for the enumeration of aerobic bacteria revivable at 22 °C and 36 °C in water, *Escherichia coli* and *Staphylococcus aureus* in meat (minced beef), and yeasts and mold in wine, from 2017 to 2024. These were chosen to demonstrate the wide applicability of the method, with matrices ranging from homogeneous (water) to heterogeneous (minced beef) and microorganisms that include both Gram-negative and Gram-positive bacteria with varied

morphologies. For each target microorganism and matrix, different methods of sample production and homogeneity control were applied.

## Samples and homogeneity control

To produce the samples of water, a volume of the matrix was sterilized and contaminated with a stock suspension of *Bacillus cereus,* before being homogenized by stirring and distributed into sterile diluent-filled vials for a final volume of 250 ml per sample and target concentrations ranging from $5\times10^1$ to $10^3$ CFU/ml. For the samples of meat, a batch of minced beef was first divided into individual Stomacher bags and sterilized to remove endogenous flora. A calibrated pooled stock solution containing all target bacteria, including *Escherichia coli* and *Staphylococcus aureus*, was then produced and used to inoculate each sample individually for a final mass of 25 g per sample and target initial concentrations of $10^2$ to $10^4$ CFU/g of the two bacterial species of interest. Finally for the wine samples, separate spiking solutions of *Saccharomyces cerevisiae* and *Aspergillus brasiliensis* were used to contaminate a batch of wine, before homogenization by stirring and division into 100 ml samples with target initial concentrations of $10^2$ to $10^5$ CFU/ml for both microorganisms. Depending on the number of participants in each test, the approximate batch sizes ranged from 90 to 110 water samples produced per test, from 75 to 125 meat samples produced per test, and from 40 to 60 wine samples produced per test.

All series of samples were then stored at $5 \pm 3$ °C and transported for both homogeneity testing and proficiency testing by refrigerated shipment. For each round and each scheme, 10 samples were taken randomly across the batch to be analyzed in duplicate for the target microorganisms by a subcontracting laboratory: the water samples were analyzed for aerobic bacteria revivable at 22 °C and 36 °C by a laboratory operating the reference method ISO 6222 [12] under ISO 17025 accreditation, the meat samples were analyzed by a laboratory using the reference methods ISO 16649-2 and ISO 6888-2 for the enumeration of *Escherichia coli* and *Staphylococcus aureus*, respectively [13, 14], under ISO 17025 accreditation, and the wine samples were analyzed for the parameters yeasts and mold by Petri dish colony count and culture on selective media, respectively.

## Standard deviation control charts

Shewhart charts have been applied in statistical process control of PT schemes to detect drifts in sample production and laboratory performance [7, 15]. The idea of using them as a tool in homogeneity testing has been introduced in chemistry [16]. However, control limits need to take into account the specific requirements for food microbiology described in ISO 22117, the relationship between standard deviations and microbial counts, and the uncertainty of the quantification of living microorganisms.

Most control charts are based on the assumption of the normal distribution of the data. In microbiological analysis, the results of colony counts (of more than 35 CFU per plate) frequently conform to the lognormal distribution; therefore, they are transformed into $\log_{10}$ CFU to have an approximately normal distribution before statistical analysis [3, 6].

Based on the observed variability of the HT and PT results of batches of samples which were concluded to be sufficiently homogeneous, the conclusion on the homogeneity of the new round batch can be drawn by comparing its standard deviations with those of the previous rounds for a similar level of mean colony counts of the same microorganism and matrix. Two charts were used to achieve this objective, by plotting HT and PT standard deviations against mean colony counts to assess the effect of the matrix complexity on the variability of the measurements, and by charting HT and PT standard deviations on a round-by-round basis to monitor the variability of the colony counts over time.

To build these two types of charts, the following parameters were estimated:

–Mean colony count of a homogeneity test, $x_{\mathrm{ht},i}$, was the average of 20 measurements of the homogeneity test $i$ ($i=1, 2, \ldots, n$).

–Standard deviation of a homogeneity test, $s_{\mathrm{ht},i}$, was the standard deviation of 20 measurements from the mean colony count of the homogeneity test $i$. It was assumed to include both between-sample and within-sample (analytical) standard deviations.

–Pooled SD of homogeneity tests, $\bar{s}_{\mathrm{ht}}$. As the 20 measurements had the same expected standard deviation for all homogeneity tests as shown in the charts in the next section, $\bar{s}_{\mathrm{ht}}$ was calculated by averaging the standard deviations of all homogeneity tests during the study period.

–95 % and 99 % confidence interval (CI) upper limits for the SD of a HT round. To detect the homogeneity tests displaying microbial counts with large deviations, only upper limits were useful and plotted. Under the assumption of normal distribution of the test results, the variance $s^2_{\mathrm{ht},i}$ of 20 HT measurements would be expected to be a random variable, and $19s^2_{\mathrm{ht},i}/\bar{s}^2_{\mathrm{ht}}$ would be expected to have a Chi-square distribution with 19 degrees of freedom [17]. The standard deviation of the results of a HT round would be expected to fall with a probability of 95 % or 99 % under the CI upper limits of $\bar{s}_{\mathrm{ht}}\sqrt{\chi^2_{1-\alpha/2,19}/19}$ with $\alpha=0.05$ or $\alpha=0.01$, respectively. If a HT standard deviation was found above the CI upper limits, the conclusion of insufficient homogeneity was based on the additional examination of production and analytical procedures, trend analysis of test results as a function of production or measurement order, homogeneity

evaluation according to ISO 13528, the variation in the laboratories' results, and confirmation by a complementary test if necessary, in order to decide if the large standard deviation was due to the heterogeneity of the test material or the analytical performance of the laboratory conducting the HT. In the case where test samples were concluded to be insufficiently homogeneous, the proficiency test could be canceled and a replacement test would be organized.

–Advancing Standards Transforming Markets (ASTM)'s upper control limit (UCL) of a $s_{ht}$ chart. As homogeneity testing was conducted on the same number of samples and similar test materials for successive rounds, the UCL of a $s_{ht}$ chart could be calculated by multiplying $\bar{s}_{ht}$ with a constant $B_4$ corresponding to the number of test results, that is $UCL = B_4\,\bar{s}_{ht}$. $B_4 = 1.490$ if 20 measurements were obtained for each test. The method was presented in the ASTM E2587-16 and is fit for the purpose of homogeneity testing [18, 19].

–Mean colony count of a proficiency test, $x_{pt,i}$, and its associated standard deviation, $\sigma_{pt,i}$. For each test microorganism, a robust mean and its associated standard deviation, expressed in $\log_{10}$ CFU, were estimated from the laboratories' results, applying Algorithm A in accordance with Annex C3 of ISO 13528:2022 [7]. They were chosen as the assigned value and the standard deviation for proficiency assessment in BIPEA's PT schemes.

–Pooled SD of proficiency tests, $\overline{\sigma}_{pt}$. A pooled SD of all the PT rounds organized during this study, weighted by the number of participating laboratories in each PT round, was calculated, using the formula (1) below:

$$\overline{\sigma}_{pt} = \sqrt{\frac{\sum_{i=1}^{n}\left(p\left(x_{pt,i}\right)-1\right)\sigma_{pt,i}^{2}}{\sum_{i=1}^{n}p\left(x_{pt,i}\right)-n}} \quad (1)$$

where

$i$ is the proficiency test $i$ ($i = 1, 2, \ldots, n$);

$p(x_{pt,i})$ is the number of laboratories participating in the estimation of the assigned value of the proficiency test $i$;

$\sigma_{pt,i}$ is the standard deviation for proficiency assessment of the proficiency test $i$;

$\overline{\sigma}_{pt}$ is the pooled SD of all proficiency tests.

–95 % and 99 % confidence interval (CI) upper limits for the SD of a PT round. Consistent with the CIs of a HT round, the variance $\sigma_{pt,i}^2$ of the results of the proficiency test $i$, where $i = 1, 2, \ldots, n$, would be expected to be a random variable, and $\left(p\left(x_{pt,i}\right)-1\right)\sigma_{pt,i}^2/\overline{\sigma}_{pt}^2$ would be expected to have a Chi-square distribution with $p(x_{pt,i}) - 1$ degrees of freedom (DF), where $p(x_{pt,i})$ was the number of laboratories participating in the estimation of the assigned value of the proficiency test $i$ [17]. The standard deviation of the results of the PT round $i$ would be expected to fall with 95 % or 99 % probability under the CI upper limits of $\overline{\sigma}_{pt}\sqrt{\chi^2_{1-\alpha/2,p(x_{pt,i})-1}/\left(p(x_{pt,i})-1\right)}$ with $\alpha = 0.05$ or $\alpha = 0.01$, respectively. The upper limits of the confidence intervals were plotted for the minimum and maximum numbers of participating laboratories observed during the study period, with $DF = p(x_{pt,min}) - 1$ and $DF = p(x_{pt,max}) - 1$, to take into account the variations in the round-by-round participation.

In addition, target SDs of 0.25 and 0.35 $\log_{10}$ CFU proposed in the ISO 22117 standard for laboratory proficiency assessment in food microbiology schemes were also plotted on the control charts as reference control limits for meat and wine matrices [3].

## Results and discussion

A summary of HT and PT results over the 2017 to 2024 period is given in Tables 1 and 2 for 6 test microorganisms in three matrices (water, meat, and wine). The microbial concentrations fell in a range of 2.20 to 4.70 $\log_{10}$ CFU across 13 to 28 rounds, depending on the test microorganism and matrix. On average, the standard deviations of the HT results were 28 to 50 % less than those of PT results. This difference could be explained by the fact that participating laboratories applied different analytical methods in comparison with the analysis of HT samples, performed by an accredited laboratory using a single standard method. For example, *Escherichia coli* was quantified by the application of the reference method ISO 16649-2 for HT samples. Although the majority of participants applied the reference method, at least five other alternative methods were used by some laboratories. The larger variations in the PT results were also attributed to different analysts performing the test and the use of different equipment, reagents, diluents, or culture media for enumeration of microorganisms.

The 95 % and 99 % CI upper limits for homogeneity tests were lower than 0.12 $\log_{10}$ CFU for all microorganisms. The 95 % and 99 % CI upper limits for proficiency tests varied according to the test microorganisms and matrices. Among the test microorganisms, the pooled SDs for yeasts and mold in wine were comparable to the pooled SDs for *Escherichia coli* and *Staphylococcus aureus* in meat (less than 0.20 $\log_{10}$ CFU), yet their CI upper limits were higher due to the lower participation for the wine PT scheme.

Standard deviations of 6 test microorganism–matrix combinations were plotted against the mean colony counts (Figs. 1a, 2a, 3a, 4a, 5a, 6a) and on a round-by-round basis (Figs. 1b, 2b, 3b, 4b, 5b, 6b), allowing chronological analysis. Each PT round was assigned a unique number in sequential order, corresponding to the 97–125 rounds for aerobic bacteria, the 31–52 rounds for *Escherichia coli*, the 31–52 rounds for *Staphylococcus aureus*, and the 14–26 rounds for yeasts and mold organized during the 2017 to 2024 period

**Table 1** Statistical summary of homogeneity tests on the enumeration of microorganisms in water, meat, and wine from 2017 to 2024

| Test microorganism | Unit | Homogeneity test (HT) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | No. of rounds | No. of measurements | Range of mean colony counts | Pooled SD | Upper 95% CI | Upper 99% CI | ASTM upper control limit |
| | | $n$ | $m$ | $x_{ht}$ | $\bar{s}_{ht}$ | $\bar{s}_{ht}\sqrt{\chi^2_{0.975,19}/19}$ | $\bar{s}_{ht}\sqrt{\chi^2_{0.995,19}/19}$ | $1.490\bar{s}_{ht}$ |
| Aerobic bacteria revivable at 22 °C | log(CFU/100 ml) | 28 | 20 | (2.335 to 3.098) | 0.055 | 0.072 | 0.076 | 0.082 |
| Aerobic bacteria revivable at 36 °C | log(CFU/100 ml) | 28 | 20 | (2.298 to 2.976) | 0.052 | 0.069 | 0.072 | 0.078 |
| *Escherichia coli* | log(CFU/g) | 22 | 20 | (2.288 to 3.677) | 0.085 | 0.112 | 0.117 | 0.127 |
| *Staphylococcus aureus* | log(CFU/g) | 23 | 20 | (2.426 to 4.113) | 0.081 | 0.107 | 0.112 | 0.121 |
| Yeasts | log(CFU/ml) | 13 | 20 | (2.865 to 4.698) | 0.057 | 0.075 | 0.079 | 0.085 |
| Mold | log(CFU/ml) | 13 | 20 | (2.386 to 3.728) | 0.081 | 0.106 | 0.112 | 0.121 |

of this study. Each plus-marked data point represents the standard deviation of 20 results from a homogeneity test, and each circle denotes a standard deviation for proficiency assessment. The 95 % and 99 % CI upper limits for the SDs of homogeneity tests and proficiency tests, ASTM's UCL for homogeneity tests, and the ISO 22117 target SD for proficiency assessment were also charted where relevant.

The SD control charts within the framework of this study demonstrate the following key results:

- Constant standard deviations were observed across different levels of mean colony counts. The homoscedasticity of the data justified the assumption of constant variance underlying the use of SD control charts, and the consistent variation in the laboratories' results over time confirmed the homogeneity of the test materials.
- Outlying data points were easy to detect. Charting both HT and PT standard deviations on the same plot enabled verification if the SDs were outliers for both HT and PT of the same round. The simultaneous heterogeneity of the HT and PT results could signal the insufficient homogeneity of the samples. For example, the HT No. 24 on the enumeration of mold in wine lay above the 95 % and 99 % CI upper limits, and $\sigma_{pt}$ of this round was slightly higher than the SDs of other rounds (Fig. 6). However, this point was under the target SD, and the observed variability had been considered in the application of a tolerance value equal to 2 $\sigma_{pt}$ from the mean colony count in the assessment of laboratory performance. The variability may also have been due to the particular morphology of mold in test samples.
- The results were not always simultaneously heterogeneous for both HT and PT. An example is *Staphylococcus aureus* in meat (Fig. 4). The dispersion of the laboratories' results in Round No. 33 was significantly higher than all other rounds, with its robust SD reaching 0.315 $\log_{10}$ CFU/g. However, the SD of the HT lay slightly above the 95 % and 99 % CI upper limits but remained under the ASTM's UCL and therefore did not provide evidence of insufficient homogeneity of the samples (Fig. 4b). Another example was Round No. 44. The outlying SD of HT results ($s_{ht}$= 0.199 $\log_{10}$ CFU/g) was attributed to a trend of decreasing concentration observed across samples in the order of sample production. However, the SD of the proficiency test (0.212 $\log_{10}$ CFU/g) was slightly higher than the pooled SD (0.193 $\log_{10}$ CFU/g) but remained below the target SD for proficiency assessment (Fig. 4b). Although all HT results in this round fell inside the tolerance interval of ($x_{pt}$± 2 $\sigma_{pt}$), the observed heterogeneity of the samples related to the production should be integrated in the interpretation of the laboratories' results. According to ISO 13528, the between-sample standard deviation should be incorporated in the calculation of the standard deviation for proficiency assessment, that is, $\sigma_{pt}' = \sqrt{\sigma_{pt}^2 + s_s^2}$. When $\sigma_{pt}$ is calculated from the participants' results, using a robust algorithm, the heterogeneity of the test samples is included in $\sigma_{pt}$. Therefore, the homogeneity acceptance criteria may be relaxed,

**Table 2** Statistical summary of proficiency tests on the enumeration of microorganisms in water, meat, and wine from 2017 to 2024

| Test microorganism | Unit | No. of rounds | Proficiency test (PT) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | No. of laboratories | | Range of mean colony counts | Pooled SD | Upper 95% CI | | Upper 99% CI | |
| | | | $p(x_{pt,min})$ | $p(x_{pt,max})$ | $x_{pt}$ | $\overline{\sigma}_{pt}$ | DF = $p(x_{pt,min})-1$ | DF = $p(x_{pt,max})-1$ | DF = $p(x_{pt,min})-1$ | DF = $p(x_{pt,max})-1$ |
| Aerobic bacteria revivable at 22 °C | log(CFU/100 ml) | 28 | 33 | 63 | (2.320 to 2.995) | 0.110 | 0.136 | 0.129 | 0.142 | 0.133 |
| Aerobic bacteria revivable at 36 °C | log(CFU/100 ml) | 28 | 34 | 65 | (2.295 to 2.924) | 0.116 | 0.144 | 0.136 | 0.150 | 0.140 |
| *Escherichia coli* | log(CFU/g) | 22 | 34 | 70 | (2.313 to 3.692) | 0.175 | 0.217 | 0.205 | 0.226 | 0.210 |
| *Staphylococcus aureus* | log(CFU/g) | 23 | 32 | 70 | (2.120 to 4.285) | 0.193 | 0.241 | 0.225 | 0.250 | 0.231 |
| Yeasts | log(CFU/ml) | 13 | 12 | 27 | (2.936 to 4.764) | 0.201 | 0.284 | 0.256 | 0.302 | 0.267 |
| Mold | log(CFU/ml) | 13 | 10 | 21 | (2.416 to 3.908) | 0.172 | 0.250 | 0.225 | 0.267 | 0.236 |

depending on the effect. Additionally, all participants are informed about this sample heterogeneity information in the interlaboratory comparison report [7].

- All target SD, ASTM's UCL, and the 95 % and 99 % CI upper limits can be fit for the purpose of homogeneity verification and should be taken into account by the PT provider. There was a general agreement among the control limits determined by different methods for meat and wine (Figs. 4, 5, 6). However, the laboratories' results of aerobic bacteria revivable at 22 °C and 36 °C in water showed low variability across 28 successive rounds (Figs. 1, 2), which is unsurprising given the homogeneous nature of the matrix itself. The 95 % and 99 % CI upper limits, varying from 0.129 to 0.150 $\log_{10}$ CFU/ml, were significantly lower than the target SD in food PT schemes (0.250 $\log_{10}$ CFU/ml). In the absence of a normative target SD for proficiency assessment in water microbiology, the 95 % and 99 % CI upper bounds could be used as their reference control limits.

Adopting a statistical process control approach to verifying sample homogeneity highlights other advantages.

- It can assist in homogeneity evaluation when standard procedures result in a "Type 1" error, falsely rejecting homogeneous samples. An example is Round No. 120 of aerobic bacteria revivable at 22 °C in water (Fig. 1). The small variance of both HT and PT results did not meet the ISO 13528 evaluation criteria for homogeneity testing. However, the charts demonstrated the homogeneity of the test material in both HT and PT tests.
- The requirement for replicate analysis of HT samples can be relaxed. When test protocols are in control over a long period of time, single analysis of the samples is sufficient to evaluate homogeneity, which implies a reduction in HT costs [16].
- SD control charts contribute to characterizing reproducibility in the quantification of a microorganism in a matrix under the stable testing procedure over a long period of time, allowing a $\sigma_{pt}$ to be established for each microorganism–matrix combination. $\sigma_{pt}$ can be lower or higher than the target SDs recommended in ISO 22117, taking into account the inherent heterogeneity in microbiological distribution and matrix composition. This approach also allows minimum and maximum SD limits to be defined to ensure consistent evaluation of laboratory performance in a PT scheme, which is a long-term aim of this study. Even in the absence of a homogeneity test, $\sigma_{pt}$ "in control" is an indicator of homogeneous samples used for a PT.
- Building SD control charts for the sample strain of microorganisms in different matrices can be effective in the evaluation of the matrix effect on the same test

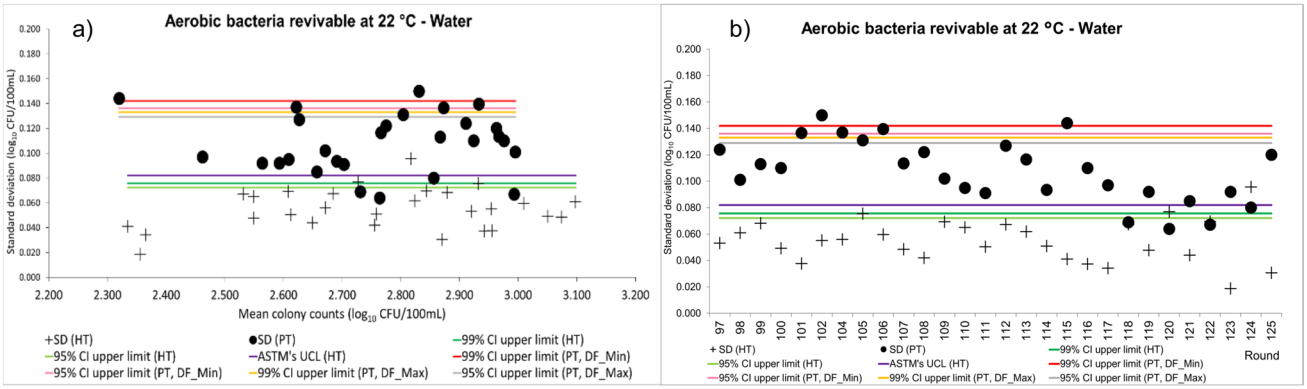


**Fig. 1** Standard deviations of 28* successive homogeneity tests and proficiency tests on the enumeration of aerobic bacteria revivable at 22 °C in water from 2017 to 2024 (**a**) against mean colony counts and (**b**) for individual rounds. * No HT was scheduled for the round No. 103

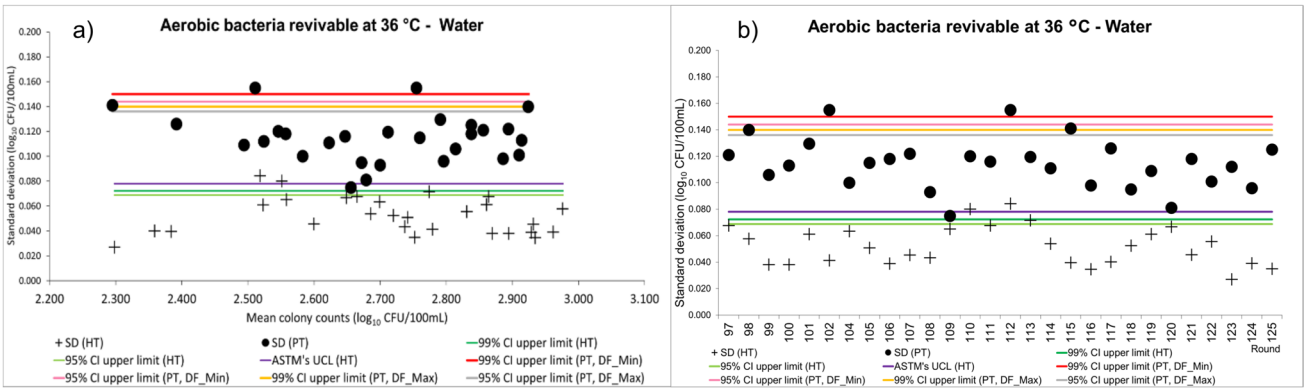


**Fig. 2** Standard deviations of 28* successive homogeneity tests and proficiency tests on the enumeration of aerobic bacteria revivable at 36 °C in water from 2017 to 2024 (**a**) against mean colony counts and (**b**) for individual rounds. * No HT was scheduled for the round No. 103

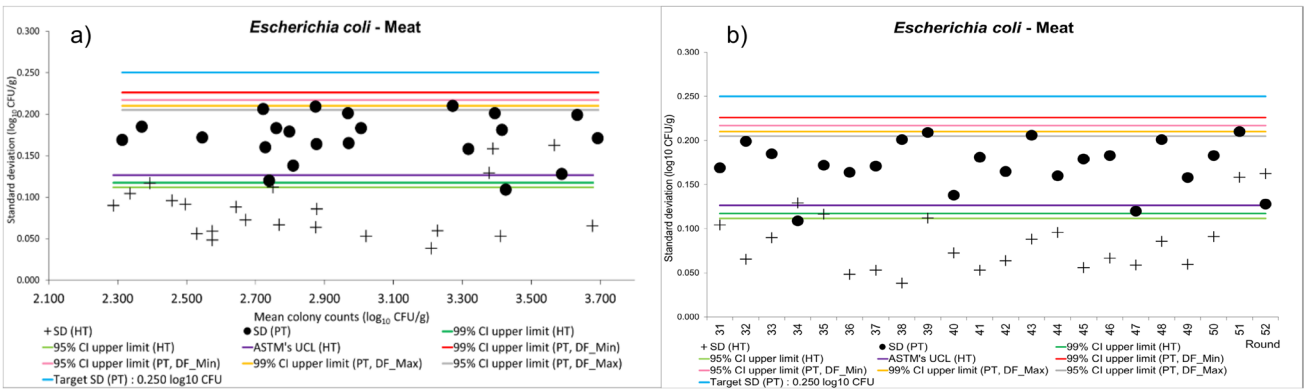


**Fig. 3** Standard deviations of 22 successive homogeneity tests and proficiency tests on the enumeration of *Escherichia coli* in meat from 2017 to 2024 (**a**) against mean colony counts and (**b**) for individual rounds

microorganism. This is essential in the construction of a characteristic function for each microorganism–matrix combination or a group of microorganisms and matrices showing similar behaviors.

- It is also feasible to use SD control charts to compare the performance of different analytical methods applied to the enumeration of the same microorganism, which needs to be explored in future research.

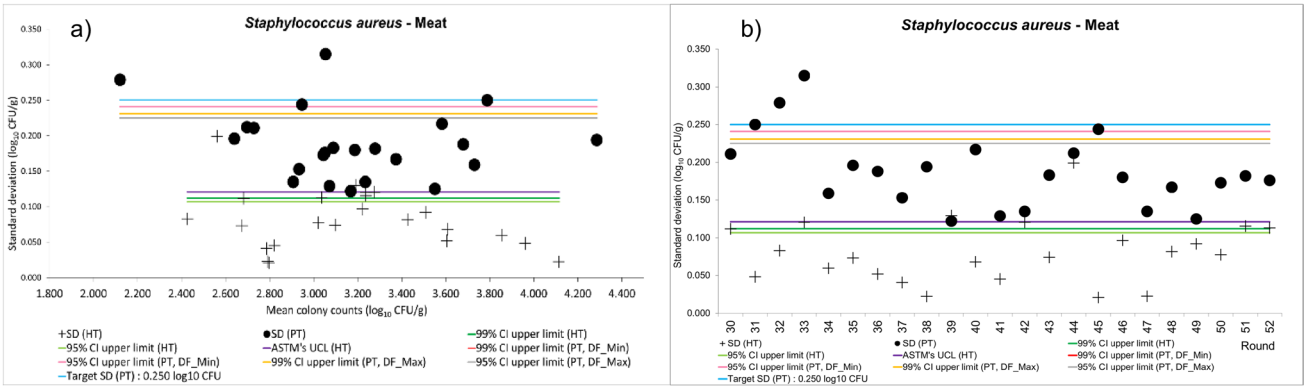


**Fig. 4** Standard deviations of 23 successive homogeneity tests and proficiency tests on the enumeration of *Staphylococcus aureus* in meat from 2017 to 2024 (**a**) against mean colony counts and (**b**) for individual rounds. The upper limit of the 99% confidence interval for the minimum number of participants overlaps with the target SD line.

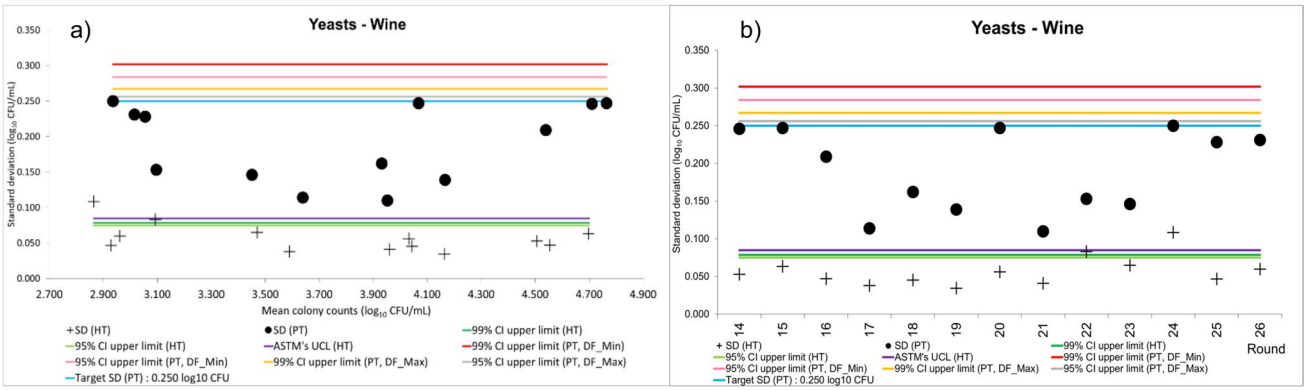


**Fig. 5** Standard deviations of 13 successive homogeneity tests and proficiency tests on the enumeration of yeasts in wine from 2017 to 2024 (**a**) against mean colony counts and (**b**) for individual rounds

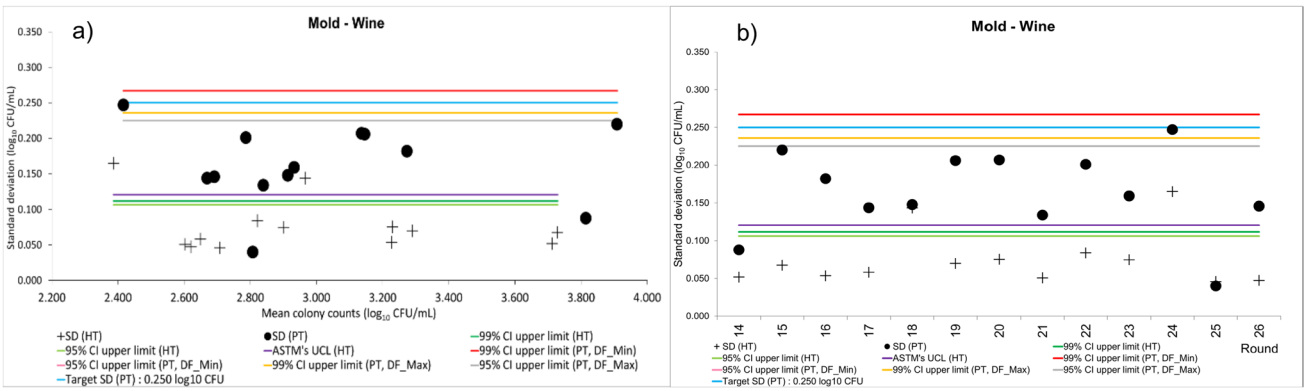


**Fig. 6** Standard deviations of 13 successive homogeneity tests and proficiency tests on the enumeration of mold in wine from 2017 to 2024 (**a**) against mean colony counts and (**b**) for individual rounds. The upper limit of the 95% confidence interval for the minimum number of participants overlaps with the target SD line

The application of standard deviation control charts presents certain constraints.

- When the conclusion of standard homogeneity testing procedures differs from the control limits indicated on the SD control charts, further verifications should be undertaken. Plotting microbial counts by sample production and measurement order enables the identification of production- and measurement-related effects on the insufficient homogeneity of the test samples. It is also important to verify the transport conditions and the dates of analysis. A temperature indicator was sent with all shipped batches to check their conformity at reception, and laboratories were requested to perform the analyses as soon as they received the samples. Prolonged exposure to elevated temperatures (more than 8 °C during 48 h) and delayed sample analysis are major sources of result variability that is not attributable to sample heterogeneity. Furthermore, supplementary homogeneity tests may be performed to confirm the results of the first homogeneity test.
- If the assumptions of homoscedasticity and data normality are not satisfied, the determination of characteristic functions that describe the relationship between microbial counts and standard deviations enables the calculation of control limits fitted to the data distribution. A goodness-of-fit test or the Shapiro–Wilk test for data normality should be performed in future works. If the data are not normally distributed, a non-parametric standard deviation control chart may be considered [20]. However, when the standard deviations are proportional to microbial counts, relative SD control charts can be constructed in a similar way to SD control charts [16]. Furthermore, measurement variability in microbial counts is observed to be stable over a limited range, suggesting that SD control charts remain effective within a specific domain of microbiological analysis.

## Conclusion

Given the limitations of the normative methods, standard deviation control charts can be effective for the verification of sample homogeneity in microbiological proficiency tests when the variance of the test results is expected to be constant across a wide range of microbial concentrations, as observed in BIPEA's PT schemes for the enumeration of microorganisms in a variety of food and water matrices. The variability of homogeneity and proficiency testing results "in control" over a long period of time would be evidence of the stability of sample production and test protocols. The pooled standard deviations and control limits obtained from a large number of successive rounds of the tests can give a reliable assessment of the homogeneity of the samples. Any drifts from the control limits on the charts at the moment of homogeneity testing should be investigated to find the causes to ensure that heterogeneous test materials are not used to evaluate the analytical performance of laboratories. If samples with "out-of-control" variances are chosen to be used for proficiency testing, the homogeneity of the samples should be reassessed a posteriori based on the examination of the laboratories' results on the same batch of samples. Plotting the standard deviations of both homogeneity and proficiency testing results on the same control chart can assist in making a quick decision on the acceptable variability of the measurements. Furthermore, the charts can be progressively improved by incorporating the results of new rounds, with the control limits reassessed annually, and can complement the standard homogeneity testing procedure, particularly when it is impossible to calculate the standard evaluation criteria.

Standard deviation control charts become a cost-saving alternative to sample homogeneity testing in the long run when costly tests can be an impediment for laboratories from participating regularly in external quality control through interlaboratory proficiency tests. PT provider experience is a key factor in the development of reliable control charts for each microorganism–matrix combination.

**Acknowledgements** BIPEA acknowledges all laboratories participating in these proficiency tests.

**Author contributions** Conceptualization was performed by Colin Howell, Thanh Thuy Nguyen, Abdelkader Boubetra; Methodology by Thanh Thuy Nguyen, Anne Tirard, Abdelkader Boubetra; Formal analysis and investigation by Thanh Thuy Nguyen, Sandrine Nguyen; Writing—original draft by Colin Howell, Thanh Thuy Nguyen, Nicolas Choquet; Writing—review and editing by Sandrine Nguyen, Abdelkader Boubetra, François Le Nestour; Project administration by Colin Howell; Supervision by Romain Le Neve, Abdelkader Boubetra

**Funding** No funding was received for conducting this study or to assist with the preparation of this manuscript.

**Data availability** Data will be made available upon request.

## Declarations

**Competing interests** All authors certify that they have no competing interests to declare that are relevant to the content of the article.

## References

1. ISO/IEC 17025 (2017) General requirements for the competence of testing and calibration laboratories
2. ISO/IEC 17043 (2023) Conformity assessment—General requirements for the competence of proficiency testing providers
3. ISO 22117 (2019) Microbiology of the food chain—Specific requirements and guidance for proficiency testing by interlaboratory comparison

4. Lightfoot NF, Maier EA (eds) (1998) Microbiological analysis of food and water: guidelines for quality assurance. Elsevier, Amsterdam
5. Noblett T (2012) Establishing the standard deviation for proficiency assessment ($\sigma$) in microbiology PT/EQA schemes. Accred Qual Assur 17:383–388. https://doi.org/10.1007/s00769-012-0909-z
6. Jarvis B (2016) Statistical aspects of the microbiological examination of foods. Academic Press, Amsterdam
7. ISO 13528 (2022) Statistical methods for use in proficiency testing by interlaboratory comparison
8. Fearn T, Thompson M (2001) A new test for 'sufficient homogeneity.' Analyst 126:1414–1417. https://doi.org/10.1039/B103812P
9. Thompson M, Ellison SLR, Wood R (2006) The international harmonized protocol for the proficiency testing of analytical chemistry laboratories. Pure Appl Chem 78(1):145–196. https://doi.org/10.1351/pac200678010145
10. Coucke W, Tanaskovic JV et al (2019) Alternative sample-homogeneity test for quantitative and qualitative proficiency testing schemes. Anal Chem 91(3):1847–1854. https://doi.org/10.1021/acs.analchem.8b03313
11. Thompson M (2015) Is your 'homogeneity test' really useful? Anal Methods 7:1627–1629. https://doi.org/10.1039/C4AY02762K
12. ISO 6222 (1999) Water quality—enumeration of culturable micro-organisms—colony count by inoculation in a nutrient agar culture medium
13. ISO 16649-2 (2001) Microbiology of food and animal feeding stuffs—horizontal method for the enumeration of β-glucuronidase-positive *Escherichia coli*—colony-count technique at 44 °C using 5-bromo-4-chloro-3-indolyl $\beta$-D-glucuronide
14. ISO 6888-2 (2021) Microbiology of the food chain—horizontal method for the enumeration of coagulase-positive staphylococci (*Staphylococcus aureus* and other species)—method using rabbit plasma fibrinogen agar medium
15. Shewhart WA (1931) Economic control of quality of manufactured product. Van Nostrand Co., Inc., New York
16. Thompson M, Fearn T (2011) A long-term look at homogeneity testing: prospects for a cheaper 'quality control' based test. Anal Methods 3:2529–2533. https://doi.org/10.1039/C1AY05425B
17. Casella G, Berger RL (2002) Statistical inference. Duxbury
18. ASTM E2587-16 (2016) Standard practice for use of control charts in statistical process control
19. Neubauer DV (ed) (2010) Manual on presentation of data and control chart analysis, 8th edition. ASTM International, West Conshohocken, PA
20. Das N (2008) A non-parametric control chart for controlling variability based on squared rank test. Econ Qual Control 2(2):114–125