

ORIGINAL ARTICLE

Commutability of food microbiology proficiency testing samples

M. Abdelmassih¹, M. Polet², M.-J. Goffaux³, V. Planchon⁴, K. Dierick² and J. Mahillon¹

¹ Laboratory of Food and Environmental Microbiology, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

² Institute of Public Health - Foodborne pathogens Section, Brussels, Belgium

³ REQUASUD, Gembloux, Belgium

⁴ Agriculture and Natural Environment Department, Walloon Agricultural Research Centre (CRA-W), Gembloux, Belgium

Keywords

commutability, food microbiology, matrix effect, methods, proficiency testing, samples.

Correspondence

Marleen Abdelmassih, UCL – Laboratory of Food and Environmental Microbiology, Croix du Sud 2-L7.05.12, B-1348 Louvain-la-Neuve, Belgium.

E-mail: marleen.abdelmassih@uclouvain.be

2013/1657: received 14 August 2013, revised 7 November 2013 and accepted 16 November 2013

doi:10.1111/jam.12396

Abstract

Aim: Food microbiology proficiency testing (PT) is a useful tool to assess the analytical performances among laboratories. PT items should be close to routine samples to accurately evaluate the acceptability of the methods. However, most PT providers distribute exclusively artificial samples such as reference materials or irradiated foods. This raises the issue of the suitability of these samples because the equivalence—or ‘commutability’—between results obtained on artificial vs. authentic food samples has not been demonstrated. In the clinical field, the use of noncommutable PT samples has led to erroneous evaluation of the performances when different analytical methods were used. This study aimed to provide a first assessment of the commutability of samples distributed in food microbiology PT.

Methods and Results: REQUASUD and IPH organized 13 food microbiology PTs including 10–28 participants. Three types of PT items were used: genuine food samples, sterile food samples and reference materials. The commutability of the artificial samples (reference material or sterile samples) was assessed by plotting the distribution of the results on natural and artificial PT samples. This comparison highlighted matrix-correlated issues when nonfood matrices, such as reference materials, were used. Artificially inoculated food samples, on the other hand, raised only isolated commutability issues.

Conclusions: In the organization of a PT-scheme, authentic or artificially inoculated food samples are necessary to accurately evaluate the analytical performances. Reference materials, used as PT items because of their convenience, may present commutability issues leading to inaccurate penalizing conclusions for methods that would have provided accurate results on food samples.

Significance and Impact of the Study: For the first time, the commutability of food microbiology PT samples was investigated. The nature of the samples provided by the organizer turned out to be an important factor because matrix effects can impact on the analytical results.

Introduction

Food microbiology proficiency testing

Proficiency testing (PT), also called interlaboratory comparisons, has been conducted for half a century to guide

laboratories, to assess their performances and to harmonize the analytical procedures. These external quality controls are the best way to ensure that a sample analysed by different laboratories will yield consistent and accurate analytical results, regardless of which laboratory conducted the analyses (Vander Heyden and Smeyers-Verbeke

2007). Today, regular participation in PTs is compulsory for laboratories under ISO 17025 (Anonymous 2005a) certification.

In the field of food microbiology, many PT schemes are organized to evaluate the analytical performances of the laboratories in conditions close to routine. The ISO 22117 (Anonymous 2010b) standard for the organization of food microbiology PT specifies that the nature of the PT samples is a critical feature, as the analyses must detect one target micro-organism in the presence of an important background flora and interfering biological substances. Following this standard, many 'matrix-related effects' are likely to influence the results during the analysis of real food products, such as the presence of bacteriostatic components, the natural flora of the sample or the interaction between fats and fibres and resident micro-organisms. An important issue for PT organizers is thus to provide samples that mimic real-life samples, so that the analytical methods can be effectively evaluated for their applicability in routine analyses (Rej 1994).

Most microbiology PT providers still exclusively distribute reference materials (*e.g.* pellets or powders) or sterile matrices (*e.g.* skim milk powder or irradiated meat) that are artificially inoculated with lyophilized microbial strains: these samples are easy to produce, relatively stable and provide a precise assigned value. However, some of these artificial PT items are far from the food samples analysed routinely and the issue of their 'fitness for purpose' or 'commutability' has been raised.

Commutability of the PT samples

First coined in 1973 in the clinical field, the term 'commutability' is defined as 'a property of a PT sample whereby the sample has the same numeric relationship between measurement procedures as is observed for a panel of representative clinical patient samples' [CLSI C53-A (Anonymous 2010a)]. It thus refers to the adequacy of a PT item *vs* a 'real' analytical sample, by assessing the equivalence of the results obtained on a sample when using different analytical methods. In other words, the samples proposed in a PT scheme should behave the same way as real-life samples during the analysis, regardless of the analytical method used. In clinical PT, the sample commutability is considered as 'one of the most important concepts affecting the design and interpretation of PT schemes' (Miller *et al.* 2011). The use of commutable PT samples avoids erroneous conclusions that penalize some analytical methods, which would have provided accurate results on real patient samples (Vesper *et al.* 2007).

Commutability has become a source of great concern for clinical PT organizers: several studies have outlined

that *c.* 50% of the artificial samples used for clinical PTs are not commutable with clinical patient samples (Miller *et al.* 2011).

The notion of commutability, largely documented in the clinical and biochemical fields, has been completely disregarded in food microbiology. Yet, the commutability of the PT samples should be assessed in this field where a large number of analytical methods having different properties (*e.g.* sensitivity and selectivity) are validated for the measurement of some parameters. For instance, a basic analysis like the enumeration of total aerobic flora in foods can be performed using numerous validated methods based on colony-count, MPN (most probable number), Petrifilm™ or even oxygen consumption (AFSCA 2013). When different measurement techniques are used, some PT samples can turn out to be noncommutable and therefore not suitable for the evaluation of the performances.

Consequences of noncommutable PT samples

A noncommutable PT sample will give rise to incoherent results when different analytical methods are used by the participants: differences will be observed in the PT results that would not appear on genuine samples. The anomaly observed is called 'matrix-related bias' or 'matrix effect': an intrinsic property of the artificial sample influences the analytical results obtained with some measurement procedures [CLSI C53-A (Anonymous 2010a)].

When a matrix-related bias is present in PT samples, this will complicate the interpretation of the results: it will be hard to assess whether an erroneous PT result (an 'outlier') is the result of a measurement malfunction or is only due to an incompatibility between the measurement procedure and the artificial test sample. In the latter case, the acceptability of the method on real food samples remains undefined.

In the case of noncommutable PT samples, the agreement between the results obtained by the participants using different measurement methods will not reflect the agreement observed on real-life samples. The results can thus only be analysed in clusters of participants using the same method (or group of methods supposed to show similar matrix-related bias). If the analysis in clusters is not possible due, for instance, to a too low number of results in some groups, the only option is to analyse all the results together and to integrate the (previously quantified) matrix-related bias into the uncertainty on the assigned value. This approach, illustrated in Fig. 1, requires a preliminary quantification of the matrix-related bias by the PT organizer. The main drawback of the grouped analysis is that it can lead to very large tolerance intervals due to the poor agreement between

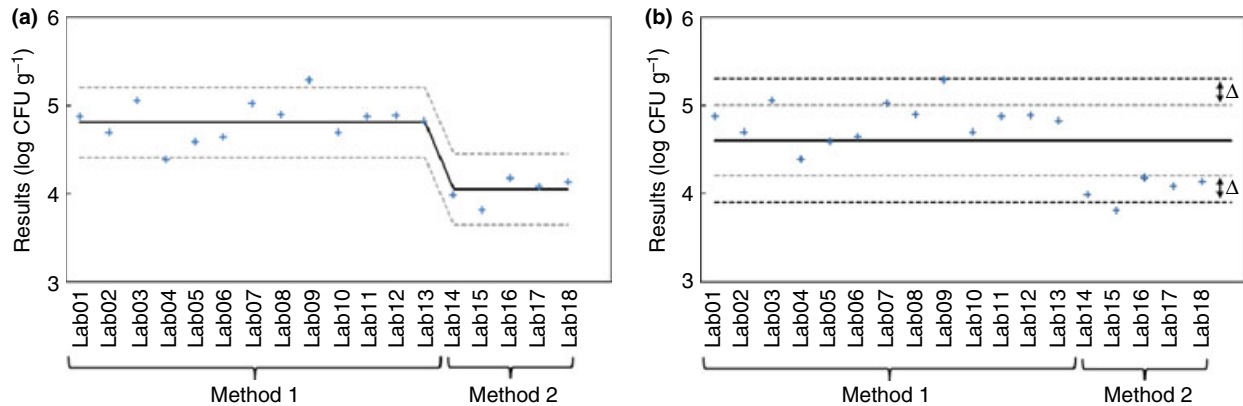


Figure 1 Analysis of PT results when a noncommutable sample is used: assigned value (—); participants results (+); tolerance limits (---); matrix-related bias (Δ); extended tolerance limits (---). The laboratories analysed the sample using two different methods (1 and 2), giving different results on the noncommutable PT sample. The analysis in clusters (a) is recommended, but not statistically robust as the second group contains less than six participants. When no clustering is done (b), the tolerance limits are supplemented with the matrix-related bias to yield the extended tolerance limits. The grouped analysis leads to erroneous conclusions as doubtful results (Lab09) are now within the acceptance zone.

results: outliers might then erroneously be considered satisfactory.

To avoid such issues, PT providers should make sure to distribute only ‘universal’ commutable samples that truly inform on the analytical results yielded by different measurement methods on real-life samples (Miller *et al.* 2011).

How to produce commutable PT samples?

The best way to ensure that the food microbiology PT samples are commutable is to propose only authentic naturally contaminated food matrices, close to the laboratory real-life conditions. However, the introduction of real food samples into a PT scheme requires special precautions and raises technical challenges, as the natural contamination of foodstuffs is generally heterogeneous, unstable and highly variable. Due to these technical constraints, most food microbiology PT providers made the choice to propose only sterile spiked matrices or reference materials as test samples, without knowing precisely if these samples behave as routine samples or if they exhibit a matrix-related bias for some measurement methods.

In food microbiology, such matrix effects can be due to the composition and structure of the sample (*e.g.* fats, enzymes, fibres, salts or preservatives), the presence of a microbiological flora and the preparation process. It has been demonstrated in clinical chemistry that the processing of a sample (*e.g.* lyophilization, freezing, sterilization or addition of stabilizing components) can significantly modify the matrix properties and compromise the commutability of clinical samples (Vesper *et al.* 2007).

When authentic PT samples are used, an artificial inoculation (or ‘spiking’) of the samples is often needed to reach adequate concentrations of the analytes. It is generally accepted that the supplementation of a sample with small amounts of purified analytes will not alter the matrix composition and will not jeopardize the sample commutability. This assumption was, for instance, successfully demonstrated on clinical serum samples supplemented with creatinine (NKDEP 2012).

Statistical analysis to assess commutability

Several methodologies have been described in the clinical literature to assess the commutability of reference materials or PT samples. The clinical laboratories have approved two guidelines to validate the commutability of samples: CLSI EP14-A2 (Anonymous 2005b) for the evaluation of matrix effects and CLSI C53-A (Anonymous 2010a) for the validation of clinical reference materials. These documents postulate that the evaluation of commutability requires the measurement of the artificial sample to test, in parallel with a ‘real-life’ sample (*e.g.* patient sample). The general approach is to analyse both types of samples with a reference and with an alternative method and to evaluate whether the artificial samples follow the same distribution as the authentic samples or if a bias (or ‘matrix effect’) is detected (Eckfeldt and Copeland 1993).

The most documented methodology to assess commutability includes two steps. First, a regression analysis is performed to plot the mathematical distribution of results obtained on genuine samples with two different measurement methods. The second step of the analysis is to add, on the same graph, the results obtained on the

artificial sample using the two analytical methods. If the results of the artificial sample are outside the 95% prediction interval of the authentic samples' distribution, there is a matrix effect and the sample is considered noncommutable with natural samples.

In a more developed way, multivariate analysis has been used in clinical chemistry to compare simultaneously the results obtained with more than two analytical methods: here, the cluster of the 'real' clinical samples results is delimited by an ellipse on the multivariate graph. The points obtained by the analysis of artificial samples are then placed on the graph and one can observe if they are found within the cluster of 'real' samples. This multivariate approach has two major drawbacks: it requires a large number of analytical results (for each method to assess) and the conclusions are based on a visual interpretation: no objective numerical criteria are provided to conclude on the commutability of the sample (Vesper *et al.* 2007).

A third way to assess commutability is based on the plotting of the residuals from the regression analysis. This method is of little use because it is less intuitive, and it is only applicable to results with a constant relative variance over the whole measurement range (Eckfeldt and Cope-land 1993).

Materials and methods

REQUASUD is a Belgian organization that has been providing, for more than 20 years, PTs displaying conditions close to those encountered in routine food microbiological laboratories. The Institute of Public Health (IPH), Belgian national reference laboratory for food microbiology, is a food microbiology PT provider since 2010. The PT items proposed by both organizations include naturally contaminated food samples, artificially contaminated samples and reference materials. The results obtained with these three types of samples have been compared in this study.

Proficiency testing design

The PT data were collected from the REQUASUD and IPH proficiency testing from 2009 to 2013, including participants from Belgium, Luxemburg, France and Spain. Ten international collaborative trials, involving 10–18 laboratories, were conducted by REQUASUD between 2009 and 2013. In the same way, 3 PT schemes, involving 19–28 laboratories, were conducted by the IPH between 2010 and 2012.

The food-based PT samples were produced using food matrices collected from the market. The matrices were first spiked (if necessary) at concentration ranges similar

to those observed in routine testing, then homogenized, divided into subsamples and distributed the same day at $4 \pm 2^\circ\text{C}$.

In terms of composition, the naturally contaminated food items (smoked salmon, ham sausage and green beans) distributed as PT samples presented the same potentially interfering components as routine samples (*e.g.* fats, proteins, sugars, fibres and preservatives). The possible artificial inoculations of these samples consisted of low quantities (between 5 and 500 μl) of pure microbial cultures. The naturally contaminated food samples, used as a reference in this study, are thus expected to be representative of 'authentic' food samples.

The second category of samples used as PT items consisted of sterilized food matrices (milk, soy milk and meatloaf) that were artificially inoculated.

The third category of samples was the reference material RM-Bc validated by REQUASUD (Abdelmassih *et al.* 2011): this PT item consists of bacterial spores (*Bacillus cereus* ATCC 13061) adsorbed on a calcium carbonate support.

The participants performed all the analyses of the PT samples 1 day after their distribution. The analytical results of the laboratories were compiled by the organizer and a \log_{10} -transformation was performed to standardize the variance. The results were then separated in peer groups, according to the analytical method used: the first cluster (called 'ISO') contained the results obtained using the reference method while the second group gathered the results obtained with an alternative method.

Commutability studies based on PT results

The commutability of the artificial PT samples (sterile spiked food and reference materials) towards authentic food samples was assessed following the protocol described below, derived from the 'comparative method' described in guideline EP14-A2.

Distribution of the results on naturally contaminated PT samples

The first step was to plot the results obtained on naturally contaminated PT samples (ham, salmon, beans), when two different analytical methods, based on a different principle, were used. The results were grouped in 'ISO' vs 'Alternative method', and the mean value obtained by both groups was calculated. The mean results of the alternative method were set as the y axis and the mean results of the reference (ISO) method as the x axis.

A linear regression analysis was carried out on these results using the JMP statistical software. This regression accurately represents the distribution of the results if there is no curvature, if the scattering is constant on the

whole concentration range and if $R^2 \geq 0.90$. After these verifications, the two-tailed 95% confidence limits (in the y direction) around the regression line were calculated by the JMP software (SAS Institute Inc., Cary, NC, USA).

This methodology was used to plot the distribution of the PT results on authentic food samples for each analytical parameter analysed in REQUASUD and IPH proficiency testing.

Assessment of commutability of artificial PT samples

The second step of the analysis was to compare the distribution of the results obtained on artificial PT samples (sterile spiked food samples or reference materials) to the distribution of authentic samples. Therefore, the PT results obtained on these artificial samples were added to the regression graphs obtained in (Distribution of the results on naturally contaminated PT samples): the artificial samples are considered commutable if their analytical results are within the confidence limits; if a significant bias is observed, a matrix effect must be suspected.

Results

Commutability of sterile spiked food samples

The commutability of sterile and artificially inoculated food samples was assessed towards naturally contaminated food samples, to evaluate the presence of matrix effects.

Figure 2 displays the distribution of the analytical results obtained on naturally contaminated PT samples

(salmon, ham and beans) when two different methods were used by the participants. The data obtained on artificially contaminated PT samples (milk, soy milk and meatloaf) were subsequently added to these graphs. All results are expressed in logarithm ($\log \text{CFU g}^{-1}$), and each point of these graphs is the mean result of one PT sample analysed by 10–28 laboratories between 2009 and 2013.

Only the results for the enumeration of *Enterobacteriaceae* and *E. coli* with the ISO, Tempo and Petrifilm methods are presented in Fig. 2. The analysis led to the same observations when other methods were compared (*e.g.* Compass agar, Rapid'Staph (Bio-Rad, Hercules, CA, USA) or ColiID (Biomérieux, Marcy L'Etoile, France)) and for the enumeration of other parameters such as total aerobic mesophilic flora, coliforms, *Staphylococcus aureus* or *B. cereus* (data not shown).

The graphic comparison highlighted no systematic bias when sterile spiked food samples were used as PT items: most of the points follow the same distribution as naturally contaminated food samples.

However, during two PT schemes including sterile spiked milk and naturally contaminated smoked salmon, isolated cases of commutability issues were raised by the artificial samples. The laboratories using the Tempo LAB method for the enumeration of lactic acid bacteria were unable to detect the strain *Lactobacillus rhamnosus* (reference strain LMG 6400) introduced into the sterile milk samples. In the smoked salmon samples, however, the Tempo LAB method enabled a correct enumeration of

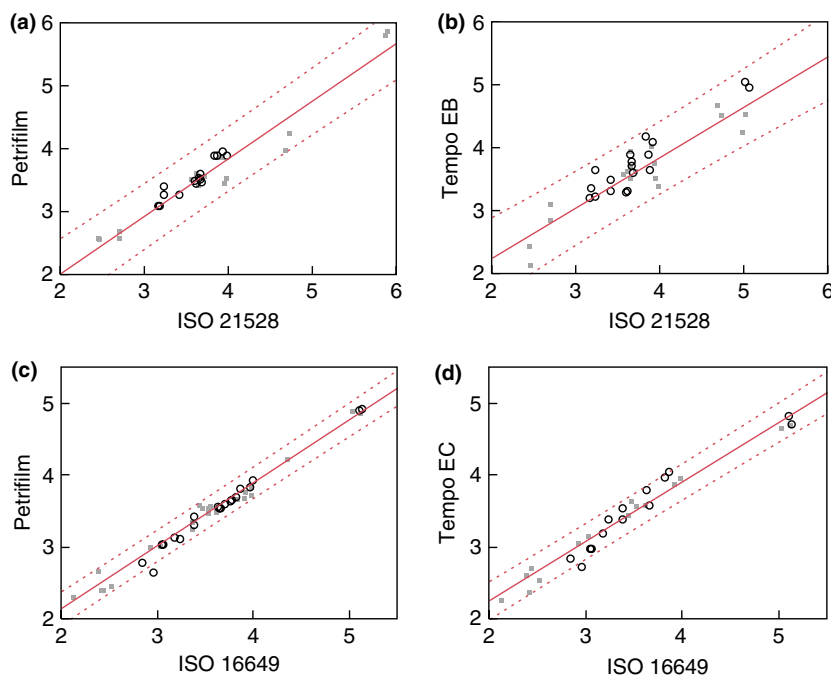


Figure 2 Distribution of the PT results on naturally contaminated PT samples (■) and sterile spiked samples (○), using two different analytical methods, for the enumeration of *Enterobacteriaceae* (up) or *E. coli* (down). The results obtained by the reference method (ISO) are plotted against the alternative method Petrifilm (a and c) or Tempo (b and d). The linear regression (—) for the results on naturally contaminated samples is indicated, with its 95% confidence limits (---).

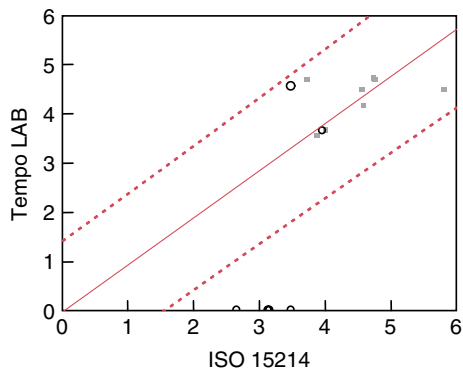


Figure 3 Results obtained on naturally contaminated smoked salmon (■) and on sterile spiked milk (○), for the enumeration of lactic acid bacteria following ISO 15214 and Tempo LAB method.

the natural lactic flora. As shown in Figs 3 and 4, PT results were located on the *x* axis: while 500–3000 CFU g⁻¹ were enumerated by the ISO method, the Tempo LAB method yielded '<10 CFU g⁻¹'.

Complementary tests (data not shown) confirmed that the problem came from the *Lact. rhamnosus* strain, which was not recognized by Tempo LAB, a method based on an automated MPN analysis using a dehydrated culture medium containing a fluorochrome (AFNOR 2009). The main issue was therefore associated with a monostrain inoculation of the artificial PT samples, a situation not representative of most natural food contaminations. In this case, the outliers were not only the result of a measurement malfunction but were partly attributable to an incompatibility between the measurement procedure and the artificial PT samples provided.

Commutability of reference materials

Using the same methodology as above, the interlaboratory results obtained on REQUASUD's reference material RM-Bc (calcium carbonate pellets carrying *B. cereus* spores) were compared with the results yielded on authentic food samples. The distribution of the results for the enumera-

tion of total aerobic flora and *B. cereus* are presented in Fig. 4a,b, respectively.

For the enumeration of total flora using ISO 4833 vs Tempo TVC methods (Fig. 4a), the results obtained for the RM-Bc globally follow the same distribution as the naturally contaminated food samples. A slight difference between both distributions is observed, but 14 of the 16 RM-Bc results are within the 95% confidence limits. These data suggest the presence of a nonsignificant bias when RM-Bc is analysed using the Tempo TVC method. This nonsignificant but constant bias may not be of big concern for the RM producer, but it will contribute to the total error of the analyses.

Surprisingly, for the enumeration of *B. cereus* in the reference material RM-Bc (Fig. 4b), the results of the ISO 7932 vs. Compass methods did not follow the distribution observed on naturally contaminated PT samples. A constant bias was observed, the alternative method yielding results that were 0.5–1 log inferior to those obtained with the reference method.

This bias can be explained by the composition of the Compass agar used in the alternative method: this culture medium is less nutritive and much more selective than the MYP agar used in the reference method. Supplementary 'selective agents' (a mix of antibiotics) are present in the Compass agar to avoid the growth of interfering bacteria and moulds, which are able to grow on MYP and to complicate the reading of the results (AFNOR 2010). In authentic food samples containing protecting components (*e.g.* fats, proteins and sugars) and vegetative bacteria, this increased selectivity did not significantly reduce the results of the *B. cereus* enumeration. However, while analysing an artificial sample made of spores and calcium carbonate (Abdelmassih *et al.* 2011), the growth of the sporulated bacteria seemed to be difficult on this highly selective medium. We can assume that some antibiotics present in the Compass agar might have inhibited the germination of the bacterial spores.

This noncommutability of the RM-Bc, provided as PT sample, represents a problematic situation, as this reference material behaves differently than real-life food sam-

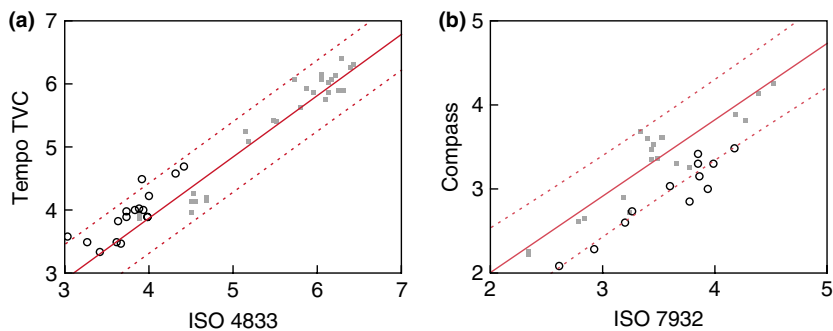


Figure 4 Results obtained on naturally contaminated PT samples (■) and RM-Bc (○), using two different analytical methods for the enumeration of total flora (a) or for the enumeration of *Bacillus cereus* (b).

ples under the Compass method and therefore penalized the laboratories using this technique.

Discussion

This study, based on the results of two Belgian proficiency testing organizers, provided a first evaluation of the commutability of food microbiology PT samples. Natural and artificial samples were introduced in the PT schemes, which enabled a direct comparison of the results obtained on both types of test items. Naturally contaminated food samples provide authentic, unbiased test materials for PT that enable an evaluation of the laboratories' analytical performances in real conditions. However, the technical hitches linked to those problematic matrices led most PT providers to distribute only sterile spiked food samples or artificial reference materials.

In terms of distribution of the results, the sterile spiked food samples tested in this study (full milk, soy milk and meatloaf) globally followed the same distribution as naturally contaminated food samples. This corroborates the theory, already verified in the clinical chemistry field, that the supplementation of a sample with a small amount of analytes will not alter the matrix composition and the commutability of the sample (Miller *et al.* 2011).

Artificial food samples can thus be used as PT items without raising systematic commutability concerns. Yet, PT providers must keep in mind that isolated commutability issues, such as the nondetection of some lactic bacteria presented in this study, may occur with artificially inoculated samples. Whenever artificial food samples are used, the PT results should be analysed in 'clusters of methods', to identify potential differences between methods.

Regarding 'nonfood' PT samples, a significant matrix effect has been highlighted in the reference material RM-Bc for the enumeration of *B. cereus* with the Compass method. This alternative method yielded underestimated results when *B. cereus* spores had to be enumerated in a nonfood matrix. In food matrices however, the Compass method provided accurate results for the enumeration of *B. cereus*. When RM-Bc was distributed as a PT item, erroneous PT results were thus generated by an actually acceptable procedure and the laboratories using the Compass method were inaccurately penalized.

The systematic use of reference materials as food microbiology PT samples is thus a questionable practice, as it can lead to the unjustified exclusion of some participants' results.

Whenever artificial samples are distributed for the evaluation of the food microbiology laboratories, it is for the PT organizers to assess these samples' possible matrix effects and their impact on the conclusions of acceptability. As long

as a doubt remains about the commutability of a PT sample, a clustering is necessary to analyse the PT results in function of the method used and to avoid inaccurate conclusions due to sample–method interactions.

Acknowledgements

This study was supported by the Walloon Region (D GARNE) by the nonprofit association REQUASUD (www.requasud.be) and by the Belgian Agency for the Safety of the Food Chain (www.afsca.be). The authors wish to thank the REQUASUD laboratories and the NRL food microbiology of the IPH for their contribution to the technical part of this study. The financial support of the Université catholique de Louvain (UCL) is also acknowledged.

Conflict of interests

No conflict of interest declared.

References

- Abdelmassih, M., Planchon, V., Anceau, C. and Mahillon, J. (2011) Development and validation of stable reference materials for food microbiology using *Bacillus cereus* and *Clostridium perfringens* spores. *J Appl Microbiol* **110**, 1524–1530.
- AFNOR. (2009) Renewal of ISO 16140 validation Tempo TVC method for the enumeration of total aerobic flora.
- AFNOR. (2010) ISO 16140 validation of COMPASS® *Bacillus cereus* Agar method for the enumeration of presumptive *Bacillus cereus*.
- AFSCA. (2013) Liste des méthodes microbiologiques reconnues par l'AFSCA, v15.
- Anonymous. (2005a) *ISO/IEC 17025: General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva, Switzerland: International Standard Organization.
- Anonymous. (2005b) CLSI document EP14-A2: evaluation of matrix effects; approved guideline. *CLSI* **25**, 4.
- Anonymous. (2010a) CLSI document C53-A: characterization and qualification of commutable reference materials for laboratory medicine; approved guideline. *CLSI* **30**, 12.
- Anonymous. (2010b) *ISO/TS 22117: Microbiology of Food and Animal Feeding Stuffs – Specific Requirements and Guidance for Proficiency Testing by Interlaboratory Comparison*. Geneva, Switzerland: International Standard Organization.
- Eckfeldt, J.H. and Copeland, K.R. (1993) Accuracy verification and identification of matrix effects. *Arch Pathol Lab Med* **117**, 381–386.
- Miller, W.G., Jones, G.R., Horowitz, G.L. and Weykamp, C. (2011) Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem* **57**, 1670–1680.

- NKDEP. (2012) Commutability study of creatinine reference materials <http://www.nkdep.nih.gov/lab-evaluation/gfr/creatinine-standardization/commutability-study.shtml>
- Rej, R. (1994) Proficiency testing, matrix effects, and method evaluation. *Clin Chem* **40**, 345–346.
- Vander Heyden, Y. and Smeyers-Verbeke, J. (2007) Set-up and evaluation of interlaboratory studies (Review). *J Chromatogr* **1158**, 158–167.
- Vesper, H.W., Miller, W.G. and Myers, G.L. (2007) Reference materials and commutability. *Clin Biochem Rev* **28**, 139–147.