*Quality Assurance and Safety of Crops & Foods, 2014; 6 (3): 291-297*
**SPECIAL ISSUE***: Food chemistry and safety*

# Some selected discrepancies observed in food chemistry proficiency tests

**M. Sykes, M. Knaggs, S. Hunter, E. Leach, C. Eaton and D. Anderson**

*The Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, United Kingdom; mark.sykes@fera.gsi.gov.uk*

Received: 2 December 2013 / Accepted: 13 April 2014


**REVIEW ARTICLE**

## Abstract

Hundreds of food chemistry proficiency tests are carried out annually across the world. The protocols for these are well established. Very occasionally, the data associated with a proficiency test are unexpected in relation to the dispersion or distribution. This may reflect differences between the test material characterisation data and the participants' data in terms of repeatability and reproducibility. The dispersion of participants' data could also reflect a difficulty with a particular analyte/matrix combination. There might also be some unanticipated chemistry occurring in a test material. In all circumstances, there are means of characterising the differences in order to obtain an appropriate outcome for the test. This paper describes some of these differences, observed from a food chemistry proficiency testing scheme.

**Keywords:** food chemistry, repeatability, reproducibility

## 1. Introduction

Experience of running professional proficiency testing schemes shows that many of the participants focus only on their own assessment. When a poor assessment is received by a laboratory, it is generally investigated, especially if the accreditation demands it. This investigation might reveal, for example, that an analysis is more challenging for the laboratory than it might have anticipated through validation. However, the analytical challenges run much deeper than that implied by one laboratory's poor performance in a proficiency test (PT). It is important for a laboratory to understand the context in which its own performance lies, and that is the basis for this paper.

Food chemistry PTs are now well established. Professional proficiency testing providers issue hundreds of tests a year, many of which are accredited to the appropriate standard (ISO, 2010). With the numbers of participants per test typically between 20 and 150, an enormous amount of data are generated. It is unusual, therefore, to come across distributions which don't fit the expected norm of dispersion. This is clearly a good thing, since laboratories rely on the consistency to monitor their own performance.

The majority of quantitative PTs, certainly in food chemistry, use the z-score as the means of assessment. This type of score is standardised via the use of an independently-derived standard deviation for proficiency assessment (SDPA). The form of the z-score equation is:

$$z = \frac{(X - X_a)}{SDPA}$$

where X is the laboratory result and $X_a$ is the assigned value for the property being measured.

The assigned value, $X_a$, and SDPA are determined by the PT provider, so it might appear that the only variable, from the point of view of the participating laboratory, is its own result, X. However, the assigned value is subject to its own analytical challenges which might not be immediately obvious to the participant. There are two major challenges which will be looked at in this paper: homogeneity testing and deriving the assigned value from the consensus of participants' results.

It is common for guides to the use of PT (Eurachem, 2011) to attempt to categorise the z-score, whereby $|z| \leq 2$ is considered 'satisfactory', $2 < |z| < 3$ is 'questionable' and $|z| \geq 3$ is 'unsatisfactory'. Although this is a useful guideline, it is

important for participants to understand the context of the z-score. If we assume that the normal distribution underlies the received data, a laboratory ought to expect that a z-score >2 will occur with a frequency of approximately 1 in 20. This would be expected to happen even if the laboratory is operating to normal parameters. On this basis, participants, PT providers and quality managers should be considering overall trends and distributions in data.

Selected examples have been taken from the food analysis performance assessment scheme (FAPAS) to illustrate some unusual deviations from the norm experienced by laboratories carrying out homogeneity testing, participant laboratories and the PT provider.

## 2. Proficiency test material characterisation

A PT material has to be sufficiently homogeneous such that any sample-to-sample variability will not significantly affect the assessment. To ensure this, the material has to be tested (preferably prior to its distribution to participants) by a sufficiently expert laboratory. An established procedure should be used (Fearn and Thompson, 2001; Thompson *et al.*, 2006) whereby 10 randomly selected samples are tested in duplicate under repeatability conditions. The homogeneity testing establishes the degree of sample-to-sample variation and also serves to give an indication of the level of determinand in the PT material. This is not a reference value generation but the indicative value of this is discussed later.

The value of the homogeneity test itself is that it indicates whether any observed variance is due to sample heterogeneity or to the analytical test itself. If the test is carried out under repeatability conditions, with randomised sampling and analysis sequence, then the two sources of variance can be independently detected and accounted for. One analytical outlier can be excluded and still prove sufficient homogeneity. Multiple analytical failings can mean the analysis has to be repeated, usually with sufficient evidence that the material is actually homogeneous. Unacceptable sample homogeneity (rejection of data from its statistical analysis) will mean that the test material has to be re-prepared from first principles.

Provided that the homogeneity testing is done prior to the sample distribution, then there is no impact on the PT itself. The only exception to this would be a delay to the start of the test while sample preparation and/or homogeneity testing is repeated to a satisfactory end.

A competent laboratory would be used to carry out the preparation and homogeneity testing. The competence would need to be demonstrated by relevant expertise, experience and (usually) accreditation or other recognised qualification. In some cases, it would be possible to build up such a history of relevant experience that the homogeneity test itself could justifiably be modified to a simpler protocol (Thompson and Fearn, 2011).

## 3. Repeatability and reproducibility

Examples from PTs can be used to illustrate the difference between repeatability and reproducibility. FAPAS PT 3036 (FAPAS, 2012c) was a melamine and cyanuric acid analysis in animal feed matrix. The homogeneity plot for melamine, shown in Figure 1A, demonstrates that the repeatability conditions (all samples analysed in a single batch) used in the homogeneity test are evident in the data. In contrast, the data received from the participants are widespread and multi-modal (Figures 1B and 1C). The conclusion is that the method is not reproducible for these analytes, in this matrix, at these concentrations. The mean value from the homogeneity test (3.53 mg/kg) is in good agreement with the assigned value derived from the major mode (Lowthian and Thompson, 2002) of the participants' results (3.33 mg/kg). Although the data from the homogeneity test are not used in the generation of the consensus, this supports the validity of using the major mode as the assigned value in this case.

A second example is somewhat less straightforward but demonstrates the skill that the PT provider needs to have to assess datasets fully. FAPAS PT 03106 was the analysis of colours in soft drink (FAPAS, 2012a). The homogeneity data and distribution of participants' results for Brilliant Blue are shown in Figures 2A and 2B, respectively. The homogeneity data shows good agreement (repeatable) but the participants' results are bimodally distributed with the major mode being closest to the homogeneity mean value. In contrast, the corresponding data for sunset yellow are shown in Figures 3A and 3B. The homogeneity test highlighted a Cochran's outlier (pair 7; Fearn and Thompson 2001). Exclusion of the outlying pair reduced the analytical standard deviation component from 0.469 to 0.290. Following the outlying pair's removal, the data passed the statistical analysis, even though they look slightly heterogeneous. The participants' data, however, show a normal and symmetrical distribution. Although the mean value of the homogeneity in this example is in keeping with the major mode of the participants' data, the 'look' of the homogeneity data is a poor predictor of the outcome of the PT.

## 4. Difficult analyses – nitrate

Nitrate (and nitrite) is analysed in a number of food products. It may be incurred (in green leafy vegetables, for example) or added as a preservative (in cured meats, for example). Regulatory limits (EC, 2006) demand that analyses are carried out to ensure compliance. In some food types, the analysis is apparently difficult, which is

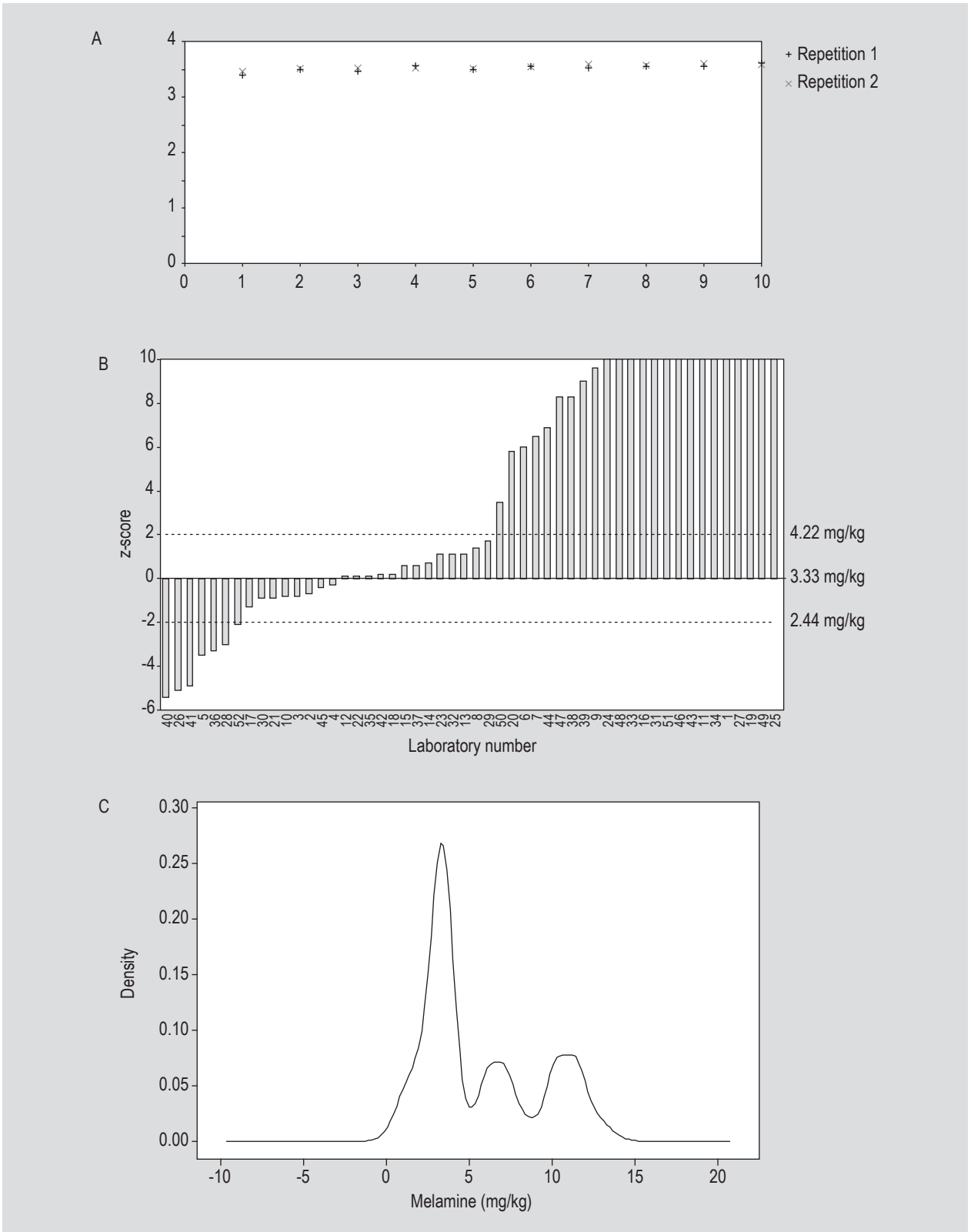


**Figure 1. PT 3036 melamine (A) homogeneity plot, (B) z-score histogram and (C) kernel density plot (FAPAS, 2012c).**

reflected in the associated PT data. In the FAPAS meat nitrate PTs, the percentage of z-scores within ±2 is typically between 60 to 70%, even though a relatively generous SDPA is applied (12.17% relative compared to the Horwitz SD of about 7.9% relative (FAPAS reports and unpublished data)). This is low compared with, for example, nitrate in green

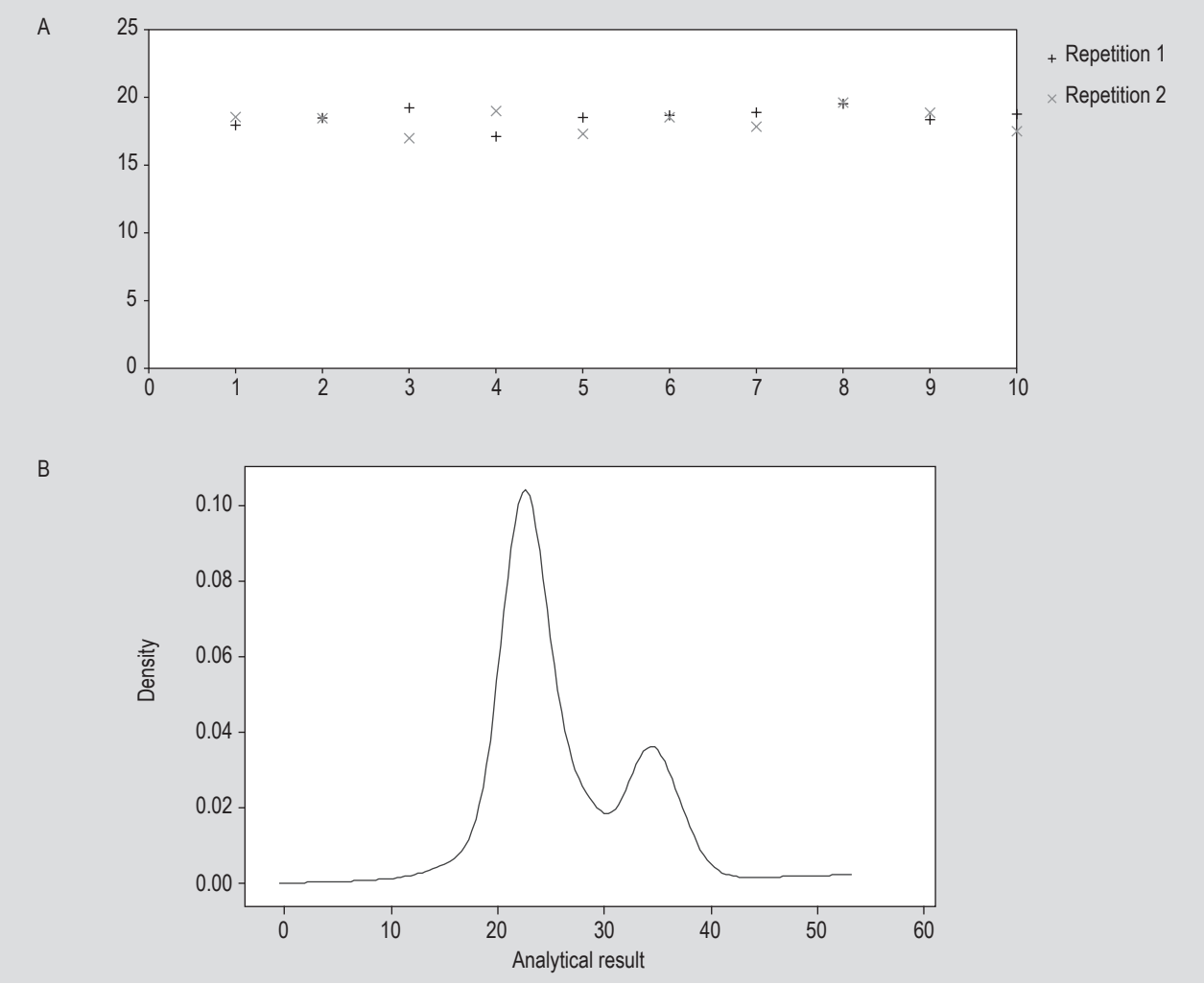


**Figure 2. PT 03106 brilliant blue (A) homogeneity and (B) kernel density plot (FAPAS, 2012a).**

leafy vegetables which typically achieves 80-85% |z| ≤2. Participants' data are generally symmetrically distributed, it is just more widespread than expected. A study carried out on FAPAS PT data (unpublished data) found no root cause for the high variance.

One of the observations made in the course of the study was that participants were normally asked to submit their results as the ionic form, i.e. the nitrate anion. Some of the standard methods, however, reported the result as the salt. In order to investigate whether the form of the determinand was affecting the variance in the results, in the course of one PT 1568 (FAPAS, 2010b), FAPAS asked participants to report their nitrate result as both the ion and the sodium salt. The histograms of z-scores can be seen in Figures 4A and 4B. The distributions are effectively identical, just that the sodium salt version is transposed by a factor equivalent to the molecular weight ratio of the nitrate salt to the anion. There is, therefore, something fundamentally more at stake with the analysis than a simple calculation.

## 5. Difficult analyses – aflatoxins in ginger

One of the very first PTs that FAPAS carried out when it was set up in 1990 was for aflatoxin analysis. Aflatoxins are found in a wide variety of food products but typically these include nuts, dried fruits and spices. The analysis is well-characterised and very routine for many food testing laboratories. This is reflected in the PTs with the percentage of |z|-scores ≤2 typically 85-90%. The PTs are conducted for a wide range of matrices and participants can report for aflatoxin $B_1$, $B_2$, $G_1$, $G_2$ and/or total aflatoxin. The SDPA is derived from the (modified) Horwitz equation (Thompson, 2000).

FAPAS PT 04185 used ginger as the matrix (FAPAS, 2012b). The results of the PT are summarised in Table 1. The percentage of |z|-scores ≤2 is abnormally low, especially for $G_1$ and $G_2$ aflatoxins. The distributions, however, are symmetrical. This PT generated a number of comments from participants. The critical attribute of this PT is the matrix; ginger is acknowledged as being a difficult matrix

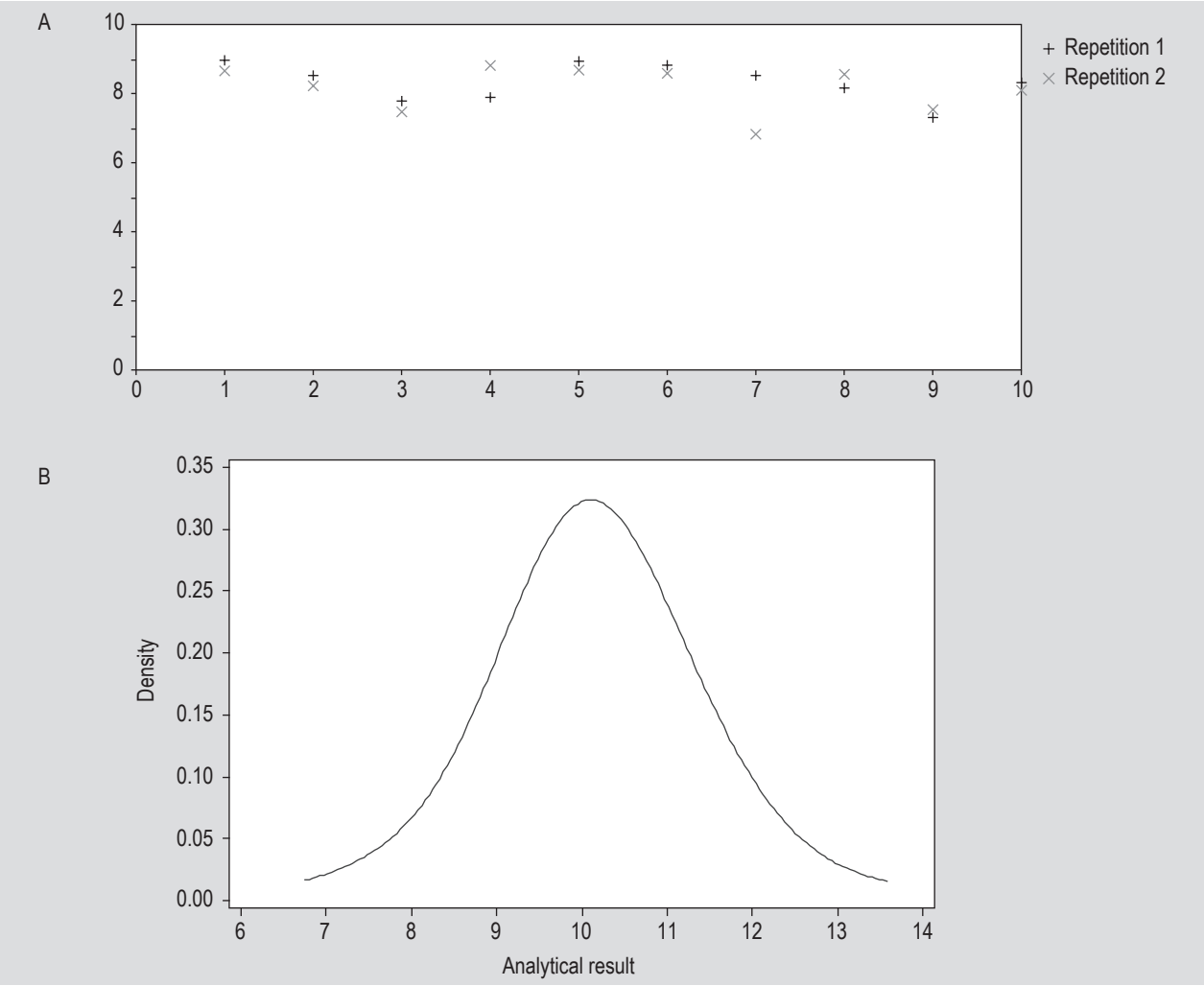


**Figure 3. PT 03106 sunset yellow (A) homogeneity and (B) kernel density plot (FAPAS, 2012a).**

**Table 1. PT 04185 summary: aflatoxin (AF) in ginger (FAPAS, 2012b).**

| Analyte | Assigned value (µg/kg) | \|z\| ≤2 | Total z | % \|z\| ≤2 |
|---|---|---|---|---|
| $AFB_1$ | 6.49 | 31 | 43 | 72 |
| $AFB_2$ | 2.15 | 33 | 43 | 77 |
| $AFG_1$ | 1.23 | 27 | 42 | 64 |
| $AFG_2$ | 0.76 | 22 | 41 | 54 |
| Total AF | 10.4 | 34 | 43 | 79 |

for aflatoxin analysis (S. MacDonald and B. Hirst, personal communications). The results of the PT simply reflect this difficulty. The generation of the assigned values is consistent for FAPAS aflatoxin PTs and, importantly, the SDPA is still derived from the same formula. We should expect that, all things being equal, participants would find this PT more difficult than most other matrices. If the SDPA was made more generous for ginger, this would defeat the purpose of demonstrating a more challenging interlaboratory comparison.

## 6. Unexpected chemistry

Very occasionally, something unexpected occurs to the analytes in a PT. In the case of pesticide levels decreasing in the PT material, this is not so unexpected and has been described elsewhere (Sykes *et al.*, 2013). In the case of FAPAS PT 02155 (FAPAS, 2010a), veterinary drug residues in bovine liver, participants were required to identify and quantify avermectins and benzimidazoles in the test material. The test material was spiked with oxfendazole (and also contained incurred eprinomectin). The EU maximum residue limits definition (EC, 2010) is the sum of all extractable residues that can be oxidised to fenbendazole sulfone, i.e. fenbendazole + oxfendazole + fenbendazole sulfone. Participants could report the individual residue components as well as the total fenbendazole sulfone. The results are summarised in Table 2.

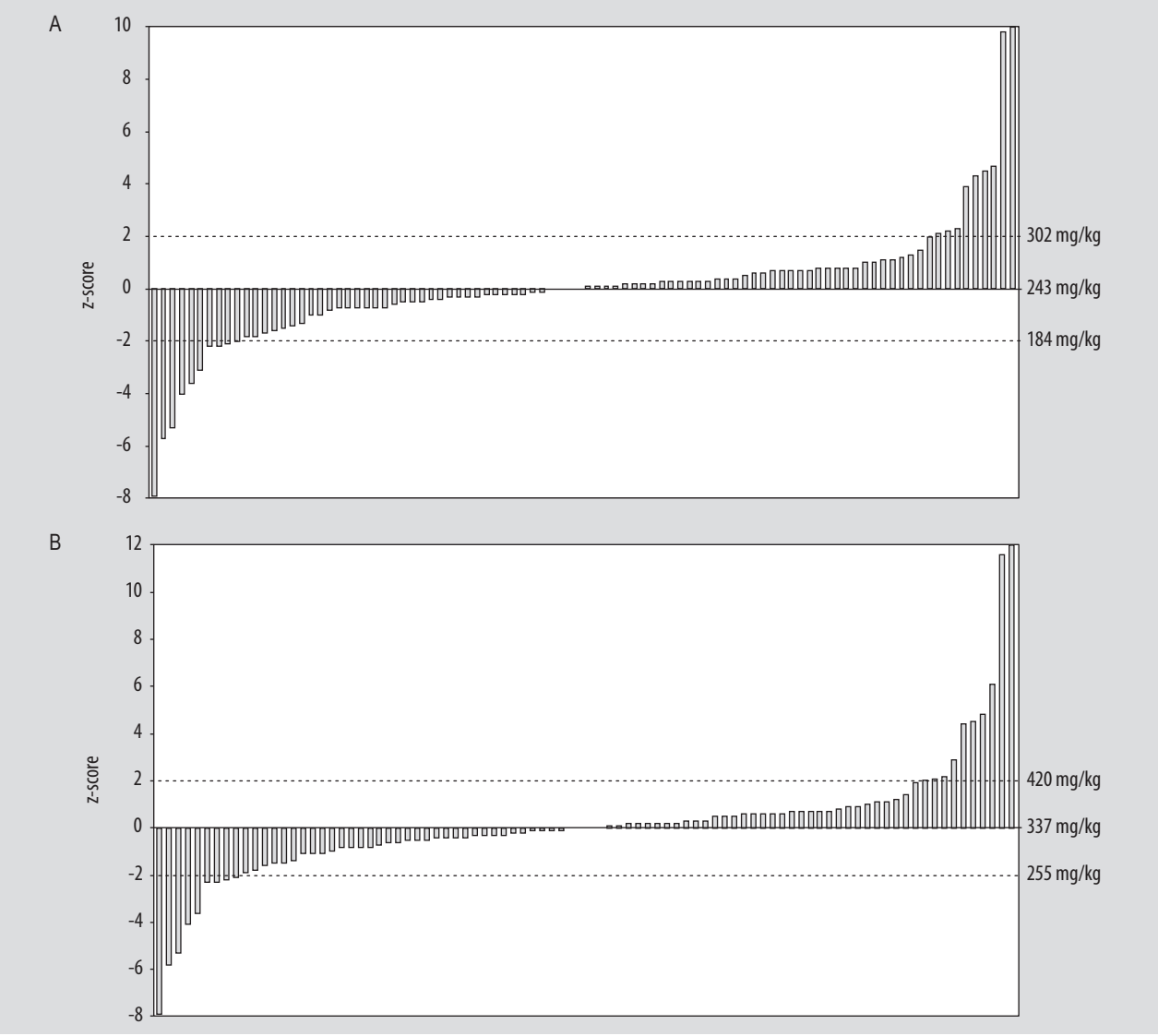


**Figure 4. PT 1568 nitrate data distribution reported as (A) ionic form and as (B) the sodium salt (FAPAS, 2010b).**

**Table 2. PT 02155 summary: veterinary drug residues in bovine liver (FAPAS, 2010a).**

| Analyte | Spike value (µg/kg) | Homogeneity mean value (µg/kg) | Assigned value (µg/kg) |
|---|---|---|---|
| Fenbendazole | NS[1] | 203 | 201 |
| Oxfendazole | 500 | 323 | 207 |
| Fenbendazole sulfone | NS | 22.4 | 12.4 IO[3] |
| Total fenbendazole sulfone | NS | NR[2] | 463 |

[1] NS = not spiked.
[2] NR = not recorded.
[3] IO = issued for information only.

During the course of the test material preparation, which involved blending a number of liver samples to obtain the appropriate level of eprinomectin, the level of oxfendazole decreased. However, instead of further oxidising to fenbendazole sulfone, it reduced chemically to the parent compound, fenbendazole. The initial spike level

of oxfendazole was 500 µg/kg. This had decreased to 65% of this level by the time the homogeneity test was conducted, and further decreased to 41% by the time the PT itself was conducted. The striking observation here is the production of fenbendazole in the matrix, and only a negligible amount of fenbendazole sulfone produced. This occurred between test material preparation and the homogeneity testing but then reached equilibration during the course of the PT. This is evident from the excellent agreement between the homogeneity means and assigned value. The fenbendazole sulfone results had sufficiently high uncertainty that assessments were issued for information only (not fully evaluative). Both fenbendazole and oxfendazole results could, however, be fully assessed.

## 7. Concluding remarks

The vast majority of food chemistry PTs run smoothly, with no undue cause for concern. Some participants will inevitably receive an assessment that is unsatisfactory in some way. However, these occasional unsatisfactory scores should be treated in the context appropriate for that analysis. Usually, this should entail the laboratory carrying out its own investigation to compliance with on-going quality control measures and long-term trends.

Occasionally, something unexpected happens with the PT data. It should be within the capability of the PT provider to investigate anomalous data and react appropriately. It should be stressed that, in the experience of FAPAS PT, this is a very rare occurrence and the examples presented here have been collated over a number of years and many hundreds of PTs.

The process of a PT starts with the test material preparation and characterisation. If an anomaly is discovered at this point, the PT should go no further. However, a difference between homogeneity data and PT results data does not necessarily mean that the PT has failed in some way (melamine and food colours examples). This might, instead, reflect the difference between the repeatability conditions of the homogeneity test and the (perhaps not) reproducibility conditions of the PT. The PT provider should have procedures in place to characterise the difference, which accreditation to the appropriate standard (ISO, 2010) would be expected to encompass.

Sometimes, the PT demonstrates that a particular analysis is more difficult than might be expected. Nitrate analysis in meat and aflatoxin analysis in ginger are good examples, in which these analytes in other matrices are relatively straightforward.

Finally, it is well to remember that chemical changes (pesticide and veterinary drug residues) can still occur outside the control of either the PT provider or the participants. These changes can be managed by the PT provider, such that assessments can still be issued under most circumstances.

## References

Eurachem, 2011. Selection, use and interpretation of proficiency testing (PT) schemes (2nd Ed.). Eurachem, Uppsala, Sweden. Available at: http://www.eurachem.org/images/stories/Guides/pdf/Eurachem_PT_Guide_2011.pdf.

European Commission (EC), 2006. Commission Regulation (EC) no. 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. Official Journal of the European Union L364: 5-24.

European Commission (EC), 2010. Commission Regulation (EC) no. 37/2010 of 22 December 2009 on pharmacologically active substances and their classification regarding maximum residue limits in foodstuffs of animal origin. Official Journal of the European Union L15: 1-76.

FAPAS, 2010a. FAPAS® proficiency test report 02155. Avermectins and benzimidazoles in bovine liver. FAPAS, York, UK.

FAPAS, 2010b. FAPAS® proficiency test report 1568. Nitrate and nitrite in meat. FAPAS, York, UK.

FAPAS, 2012a. FAPAS® proficiency test report 03106. Artificial colours in soft drink. FAPAS, York, UK.

FAPAS, 2012b. FAPAS® proficiency test report 04185. Aflatoxins B&G in ginger spice. FAPAS, York, UK.

FAPAS, 2012c. FAPAS® proficiency test report 3036. Melamine & cyanuric acid in animal feed. FAPAS, York, UK.

Fearn, T. and Thompson, M., 2001. A new test for sufficient homogeneity. Analyst 126: 1414-1417.

International Standardisation Organisation (ISO), 2010. Conformity assessment – general requirements for proficiency testing – ISO/IEC 17043. ISO, Geneva, Switzerland.

Lowthian, P.J. and Thompson, M., 2002. Bump-hunting for the proficiency tester-searching for multimodality. Analyst 127: 1359-1364.

Sykes, M., Thompson, M. and Reynolds, S., 2013. Pesticide residues in food-based proficiency test materials, spiking values versus consensus assigned values. Journal of Agricultural and Food Chemistry 61: 4205-4209.

Thompson, M., 2000. Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing. Analyst 125: 385-386.

Thompson, M., Ellison, S.L.R. and Wood, R., 2006. the international harmonised protocol for the proficiency testing of analytical chemistry laboratories. Pure Applied Chemistry 78: 145-196.

Thompson, M. and Fearn, T., 2011. A long-term look at homogeneity testing: prospects for a cheaper 'quality control' based test. Analytical Methods 3: 2529-2533.