Regular Article

ANALYTICAL
AND BIOANALYTICAL
CHEMISTRY
RESEARCH

Published by the
Iranian Chemical Society

***Anal. Bioanal. Chem. Res.**, Vol. 2, No. 2, 113-118, December 2015.*

Application of Statistics to Evaluate Iranian Analytical Laboratories Proficiency: Case of Aflatoxins in Pistachio

L. Fotouhi*, A. Meshkani and M. Hooshmand

Marjaan Khatam, Training, Research & Q.C. Lab. Services, Tehran 1416853545, Iran

(Received 13 July 2015, Received 9 October 2015)

The aim of this study was to evaluate the utility of a proficiency testing program among limited number of local laboratories as an alternative to the IUPAC/CITAC guide on proficiency testing with a limited number of participants, specially where international schemes are not accessible. As a sample scheme we planned to determine aflatoxins (B1, G1, B2, G2, total) in Iranian pistachio matrix. A part of naturally contaminated pistachio sample was tested for sufficient homogeneity by a competent laboratory and then homogenized sub-samples were distributed among participants all across the country. The median of participants' results was selected as assigned value. Student t-test was applied to show there is no significant difference between assigned and mean values of homogeneity test results obtained by the competent laboratory. Calculated z-scores showed that 6 out of 8 results in aflatoxin B1, 7 out of 8 results in aflatoxin B2, 5 out of 8 results in aflatoxin G1, 7 out of 8 results in aflatoxin G2 and 6 out of 9 results in aflatoxin total were in satisfactory range. Together our studies indicate that the approach described here is highly cost efficient and applicable for quality assurance of test results when there is no access to international proficiency testing providers.

Keywords: Proficiency testing, aflatoxins, z-Score, Pistachio, Quality assurance, Assigned value

INTRODUCTION

Aflatoxins are highly toxic and carcinogenic, and are detected in various food commodities including pistachio nuts [1]. Institute of Standard and Industrial Research of Iran controls the quality of pistachio distributed in international markets using legal maximum levels defined in ISIRI 5925 [2]. The European Union has also set special conditions for pistachios imported from Iran, due to the risks of aflatoxin contamination, in commission regulation (EC) 1152/2009 [3]. These two authorities follow the same objective individually: quality control of pistachios for aflatoxin contamination.

Achieving quality of measurement in the framework of the concept "tested once, accepted everywhere" requires comparability and compatibility of the test results [4]. Measurement results have to mainly meet specific criteria of accuracy, precision, and reliability. In the case of mycotoxins, the analytical process is more influenced due to various sources of errors arising from different steps of determination [5]. Therefore, there must be a tool for checking the laboratory performance.

Participation in external quality control programs is a part of the establishment of ISO/IEC 17025 standard [6] in a laboratory, which proves proficiency of laboratory and confirms the accuracy of results. The process is common everywhere, homogenized test samples are distributed by a provider and the participants test the samples. The results are then submitted to the provider for evaluation. The provider processes the results and returns a confidential report for each laboratory in which their status is presented. Participation in such a scheme provides comparisons with external references for laboratories.

There are currently international guidelines and protocols providing procedures for evaluating the performance of participants in PT (proficiency testing) schemes [7-9]. These references highlighted the possible limitations or problems arising from situations where the

*Corresponding author. E-mail: l.fotouhi@ut.ac.ir

number of participants is limited. Belli *et al*. [10] studied the metrological background for the selection and use of PT for a limited number of participants. Their study showed that consensus mean values and observed standard deviation of measured results of the participants in PT are insufficiently reliable for the assessment of a laboratory's performance with a limited number of participants. A new IUPAC/CITAC Guide on selection and use of PT for a limited number of participants also reached a similar conclusion on this issue [11] and suggested that traceable assigned values of test items like a certified reference material (CRM) or an in-house reference material (HRM) or a spike should be used whenever possible. However, the reality is that appropriate CRMs might not be available or it might not be practical using in-house reference materials or spiked samples as testing materials for PT testing, especially in the area of food testing. A recent research conducted by Siu-kay Wong applying Mont Carlo Simulation [12] provided statistical evidence that consensus value approach should be regarded acceptable given that 10-20% of possible inconsistency is tolerable.

The aim of the present study was to show that it is practically possible to count consensus values as assigned value when statistical evaluations indicate no significant difference between consensus values and mean of homogeneity test of AF (aflatoxins) in pistachio samples. The homogeneity test data has been obtained from a competent laboratory which has three successive acceptable *z*-scores in international schemes. As there have been many debates around the application of PT for limited number of laboratories, this new data treatment was introduced.

EXPERIMENTAL

The procedure followed was recommended by the IUPAC and FAPAS [7,13].

Test Material

The test material used in this study was slurry pistachio sample naturally contaminated by aflatoxin B1 (AFB1) at 1.33 µg kg^{-1} and spiked by aflatoxin B2 (AFB2), aflatoxin G1 (AFG1), aflatoxin G2 (AFG2) at levels of 0.72, 1.18, 0.33 µg kg^{-1} providing total amount of aflatoxin total (AFT) 3.56 µg kg^{-1}.

Homogeneity Evaluation

The sample was mixed with appropriate amount of water using a slurry machine for 1 h to achieve sufficient homogeneity. The homogeneity of the sample was checked randomly, taking 12 samples (110 g each) and analyzing in duplicates by a competent laboratory holding ISO/IEC 17025 certificate of accreditation for the test method. Cochran's C test carried out prior to ANOVA to detect outliers. The results showed no outlier existed. In the next step the ANOVA was carried out to check for significant inhomogeneity. The outcome presented sufficient homogeneity in the bulk sample. (See Supplementary Material for homogeneity statistics)

After making sure that the sample is homogenous (the data were kept secret), the samples were distributed to the participating laboratories and they tested the sample and returned their independent result to us.

Estimation of Assigned Value

The most critical step is to estimate the assigned value. Various approaches are possible for determining the assigned values and this value is subjected to the method used and choice of statistics. In the present project, the consensus of participants was chosen as assigned value. Median has been selected as the robust estimator of the consensus of the participants, which excludes the outliers. To confirm trueness of the assigned value In fact trueness refers to " the closeness of agreement between the average value obtained from a large series of test results and an accepted value" here we assumed the median as obtained value from participants results and compared it to the mean of preliminary homogeneity data previously obtained by a competent laboratory. (Data is provided as Supplementary Material) applying an appropriate form of student t-test.

$$t_{ex} = \left| X - \hat{X} \right| \frac{\sqrt{n}}{\hat{\sigma}}$$

Where t_{ex} = the experimental t value, X = the assigned value, $\hat{X}$ = the mean value of homogeneity data, $\hat{\sigma}$ = the robust standard deviation calculated from participants' results and n = the number of data points used to calculate the assigned value. The final decision is made by comparison of experimental t value (t_{ex}) and critical t value (t_c) extracted

from statistical table at 95% confidence level.

The uncertainty (u) of the median was estimated with the following equation:

$$u = \frac{\hat{\sigma}}{\sqrt{n}} \times 1.25$$

where: $\hat{\sigma}$ = the robust standard deviation, n = the number of data points (all participant submitted results) used to calculate the median.

z-Score

A z-score compares an estimate of the error of a result with a target value for the standard deviation (σ). The participants' z-scores were calculated using this equation:

$$z = \frac{(x - X)}{\sigma}$$

where x = the participant's reported result (concentration in μg kg^{-1}) corrected for recovery, X = the assigned value, and σ = either the target value for standard deviation calculated from Horwitz equation (σ_p) or the robust standard deviation ($\hat{\sigma}$). The σ_p is the numerical value of the standard deviation of a measurement result, which has been designated as a goal for measurement quality [14].

We used Horwitz equation to predict σ_p [13]. The appropriate form of Horwitz equation in this study is:

$$\sigma_p = \frac{0.22c}{w}$$

where c = concentration, i.e., the assigned value, X , and w = dimensionless mass ratio (μg kg^{-1}).

Horwitz equation presents dispersion of results in extreme conditions in which there are many variables, and therefore, standard deviation is expected to be close to the maximum possible limit. Thus, when the number of test laboratories is limited, and so as the variables, it is not appropriate to take Horwitz equation as criterion. If we use Horwitz equation instead it reduces the confidence level of the test. Whenever σ_p is greater than $\hat{\sigma}$, as a strict tool for calculation of z-score, the robust standard deviation ($\hat{\sigma}$) is better to be applied. Although resulting z-scores are satisfactory, some results may exist which are not satisfactory. In this project due to the limited number of laboratories the robust standard deviation ($\hat{\sigma}$) was smaller than what was expected from Horwitz equation. Therefore, the robust standard deviation ($\hat{\sigma}$) was used for calculation of z-scores.

RESULTS AND DISCUSSION

Results obtained from homogeneity tests and statistical evaluations are reported in Table 1. There was no outlier in homogeneity data. Table 1 shows the assigned values (X), calculated as median, together with the mean of homogeneity data ($\hat{X}$), standard uncertainty (u), and also reports the values of $\hat{\sigma}$ used to calculate z-score compared with the standard deviation predicted from Horwitz equation (σ_p). Data depict that robust standard deviation of participants is smaller than values predicted by Horwitz equation. Hence, the values of the robust standard deviation were used to obtain z-scores.

The results of student t-test are tabulated in Table 2. The evaluation of obtained data at confidence level of 95% showed that the experimental t value in all cases was smaller than critical t value. Thus, there was no significant difference between assumed assigned value and the mean value of homogeneity data obtained by a competent laboratory and supposed to be true. Table 3 shows the results of participants, recovery percentages and z-scores in different analytes. The number and percentage of z-scores in satisfactory range ($|z| \leq 2$) for each analyte are shown in Table 4.

The Analytical Methods Used by Participants

Each participating laboratory was unambiguously coded and was requested to report their result in μg kg^{-1} corrected for recovery together with the recovery percentage. The reported results are provided in Table 3 for all analytes and participants.

Approximately 78% of the nine laboratories involved in this proficiency testing were certified by accreditation bodies. The method applied by laboratories coded 10, 11, 12, 13, 14 and 15 includes extraction of analytes using methanol and water, clean-up using immunoaffinity column, and HPLC quantification were used. Laboratories coded 16 and 18 used the same method but were not

Table 1. Statistical Evaluation of Data Obtained from Homogeneity Tests and Participants

Analyte	No. of duplicates in homogeneity data	Mean of homogeneity data (µg kg^{-1}),	Target standard deviation (µg kg^{-1}), σp	Assigned value (median) (µg kg^{-1}), X	Robust standard deviation (µg kg^{-1}), $\hat{\sigma}$	Uncertainty of median (µg kg^{-1}), u
AFB1	12	1.36	0.30	1.52	0.19	0.07
AFB2	12	0.82	0.18	0.80	0.14	0.05
AFG1	12	1.24	0.27	1.41	0.13	0.05
AFG2	12	0.36	0.08	0.38	0.08	0.03
AFT	12	3.79	0.83	4.12	0.54	0.18

Table 2. Results of Student t-Test at 95% Confidence Level

Analyte	Assigned value (median) (µg kg^{-1}), X	Mean of homogeneity data (µg kg^{-1}),	The number of data points, n	Robust standard deviation (µg kg^{-1}), $\hat{\sigma}$	Critical t value, t_c	Experimental t value, t_{ex}
AFB1	1.52	1.36	8	0.19	2.36	0.30
AFB2	0.80	0.82	8	0.14	2.36	0.07
AFG1	1.41	1.24	8	0.13	2.36	0.48
AFG2	0.38	0.36	8	0.08	2.36	0.08
AFT	4.12	3.73	9	0.54	2.31	0.20

Table 3. Results and z-Scores of (a) AFB1 (b) AFB2 (c) AFG1 (d) AFG2 (e) AFT

Laboratory code	Median of results (µg kg^{-1})	Recovery (%)	z-Score
AFB1 (1.52 µg kg^{-1})			
10	1.45	91.5	-0.37
11	1.62	98.8	0.53
12	1.3	93.5	-1.17
13	1.41	95.7	-0.59
14	2.17	84.5	3.47
15	1.59	83.2	0.37
16	1.66	95	0.75
18	0.64	47	-4.69
AFB2 (0.80 µg kg^{-1})			
10	0.81	70.6	0.11
11	0.89	96	0.67
12	0.72	91.2	-0.53
13	0.7	97.9	-0.67

Table 3. Continued

14	1.11	107.6	2.21
15	1	79.7	1.44
16	0.78	94.7	-0.11
18	0.54	34.7	-1.79
AFG1 (1.41 µg kg^{-1})			
10	1.49	87.3	0.63
11	1.5	104.50	0.71
12	1.37	88.2	-0.31
13	1.35	95.5	-0.47
14	2.27	107	6.75
15	0.92	87.8	-3.84
16	1.45	97	0.31
18	0.26	60.2	-9.02
AFG2 (0.38 µg kg^{-1})			
10	0.37	81.8	-0.12
11	0.41	103.3	0.36
12	0.28	91.6	-1.21
13	0.31	97.7	-0.85
14	0.58	100.1	2.42
15	0.24	103.2	-1.70
16	0.42	98.3	0.48
18	0.39	50.2	0.12
AFT (4.12 µg kg^{-1})			
10	4.12	-	0.00
11	4.42		0.56
12	3.67		-0.83
13	3.76		-0.67
14	6.13		3.72
15	3.76		-0.67
16	4.31		0.35
17	17.1		24.04
18	1.83		-4.24

Table 4. Number and Percentage of Satisfactory z-Scores

Analyte	Number of satisfactory scores $\|z\| \leq 2$	Total number of scores	Satisfactory (%)
AFB1	6	8	75%
AFB2	7	8	88%
AFG1	5	8	62%
AFG2	7	8	88%
AFT	6	9	67%

accredited by an accreditation body. Laboratory coded 17 used an ELISA method which was just capable of quantification of AFT (total aflatoxin).

Due to the characteristics of the proficiency testing itself, it is difficult to make a reliable correlation between the obtained *z*-scores and the efficiency of the adopted methods because of the unknown competency of the performing laboratory. Nevertheless, from the *z*-score values it is possible to observe that most laboratories with the best performance used accredited methods. However, this does not mean that if the laboratory is accredited no monitoring is needed. Alternatively, it is possible that an unaccredited laboratory produces accurate result, please see Table 4- AFB1 (laboratories number 14 and 16, respectively).

CONCLUSIONS

This inter-laboratory study supplies an overview of the most used procedures for AF determination in pistachio samples, and gives an estimate of the performance of some Iranian laboratories involved in AF determination by a new data treatment. Our results showed it is practically possible to use consensus values as assigned values after comparison with the mean of homogeneity tests. The method defined here is useful for limited number of laboratories. It is actually a reasonable alternative when there is no access to international scheme, or where there is no scheme available at all. In this study, more than 75% of results lay in the satisfactory range. This issue sounds valuable in the way of achieving the framework of the concept "tested once, accepted everywhere". Iran is one of the most Pistachio exporting countries, which has been criticized for contamination of its exporting pistachio. There are many reliable laboratories within Iran which can definitely give accurate assessments of AFT levels and reduce the cost of returning consignments.

ACKNOWLEDGEMENTS

The authors would like to thank Marjaan Khatam, Training, Research and QC Lab Services headquarters for funding this research. We would like also express our gratitude to Sepideh Farshbaf, and Dr Nader Sheibani for their contributions.

REFERENCES

[1] G. Hudler, Magical Mushrooms, Mischievous Molds, Princeton University Press, Princeton, NJ, 1998.

[2] ISIRI 5925, Food and Feed-Mycotoxins-Maximum Tolerated Level, 1380.

[3] Commission Regulation (EC) 1152/2009, Imposing Special Conditions Governing the Import of Certain Foodstuffs from Certain Third Countries Due to Contamination Risk by Aflatoxins and Repealing Decisions 2006/504/EC.

[4] I. Kuselman, M. Belli, S. Elison, A. Fajgelj, U. Sansone, W. Wegscheider, Accred. Qual. Assur. 12 (2007) 563.

[5] D. Park, H. Njapau, R. Coker, in: M. Miraglia, H. Van Egmond, C. Brera, J. Gilbert (Eds.), Mycotoxins and Phycotoxins Developments in Chemistry, Toxicology and Food Safety, Alaken Inc., Colorado, United States, 1998.

[6] ISO/IEC 17025, General Requirements for the Competence of Testing and Calibration Laboratories, 2005.

[7] M. Thompson, S. Ellison, R. Wood, Pure Appl. Chem. 78 (2006) 145.

[8] ISO 13528, Statistical Methods for Use in Proficiency for Testing by Interlaboratory Comparison. International Organization Standardization, Geneva, 2005.

[9] ILAC G13, Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes. International Laboratory Accreditation Cooperation, Sydney, 2007.

[10] M. Belli, S. Ellision, A. Fajgelj, I. Kuselman, U. Sansone, W. Wegscheider, Accred. Qual. Assur. 12 (2007) 391.

[11] I. Kuselman, A. Fajgelj, Pure Appl. Chem. 82 (2010) 1099.

[12] S. Wong, Accred. Qual. Assur. 16 (2011) 539.

[13] FAPAS Proficiency Testing Protocol for Organisation and Analysis of Data. 6th ed., 2002.

[14] M. Thompson, Analyst 125 (2000) 385.