TDL.ORG TEXAS DIGITAL LIBRARY

Journal of Regulatory Science

http:\\journalofregulatoryscience.org

Journal of Regulatory Science 10(1) (2022) 1–13

Regulatory Science

Analytical Uncertainty in Animal Feed Laboratories: A Current Evaluation of AAFCO Proficiency Testing Data for Select Analytes

Timothy J. Herrman[a,*], Kyung-Min Lee[a], Yi-Cheng Hsieh[a], Sara Williams[a]

[a] *Office of the Texas State Chemist, Texas A&M AgriLife Research, Texas A&M University System, College Station, Texas 77841, United States.*

Abstract

The Association of American Feed Control Officials (AAFCO) presents analytical variations (AVs) for each specific nutrient that can partially account for an inherent variability that occurs within and between laboratories. The AV was created using the AAFCO proficiency testing (PT) program results and represent the coefficient of variation times 2 (2 CV). This study was performed to evaluate the published AVs within the association's Official Publication (OP) using AAFCO PT data from 2014 to 2020. This study also reevaluated the current AV values by applying the concept of the measurement uncertainties including the expanded uncertainty (U), Horwitz value, and 2 CV for the assessment of the label guarantees. Although greater variations were observed at lower concentrations, the majority of data points from the PT rounds showed an acceptable reproducibility for the select analytical methods except for cobalt, fat (crude fat and acid hydrolysis fat), fiber (crude fiber, acid detergent fiber, and neutral detergent fiber), moisture, and selenium where 2 CV was higher than the current AAFCO AV. Regression models were developed based on the 95% tolerance interval of expanded uncertainty and 2 CV and were validated using the 2019-2020 AAFCO PT data. The validation data fell within the expanded uncertainty and 2 CV model bounds indicating the new models were fit-for-purpose and accurately characterized current AAFCO PT data. The present study results imply the AAFCO AV and its potential application need further study.

Keywords: analytical variability; animal feed; measurement uncertainty, Association of American Feed Control Officials (AAFCO); proficiency test

1. Introduction

The Association of American Feed Control Officials (AAFCO) was established in 1909 as an advisory body comprised of industry, state, and federal government agencies. A major function of AAFCO is to publish model laws and regulations, ingredient definitions, and enforcement polices for states to regulate the manufacture and sale of animal feeds [3]. The AAFCO Official Publication (OP) introduced Permissible Analytical Variations (PAVs) in 1973 based on the Check Sample Program data from 1956 to 1972. The PAVs were updated and changed to Analytical Variations (AVs) in the 1986 OP that included corresponding analyte concentrations [2]. No record exists of the data or statistical methods used to develop the AVs, however, it is believed AVs only consider analytical error. The AVs explain inherent variability in select analytes while they do not include the manufacturing variations [1]. The animal feed regulatory community has used the AVs as guidelines to help their decisions on acceptability of feed products. However, they are not intended to permit any deficiencies in analytical procedures and are also distinguished from tolerances or investigational allowances, the latter of which include sampling error, and which AAFCO hasn't prescribed yet. A regulatory agency may use the AVs to decide if the product is in compliance based on the proximity of the test results to the label guarantee compared to the AV. Many regulatory laboratories analyze potentially violative products in duplicate. The use of replicates will impact the AVs since increasing the number of replicates will smooth data [19]. The AVs are published within the association's Official Publication (OP) and are accompanied with the following explanation by the section editor: *AVs are a measure of between-lab variance among laboratories when analyzing the same material for the same analytes, at the same analyte concentration, using the same method* [3]. The limitations of AVs are discussed within the OP.

Proficiency testing (PT), also called external quality assessment, is a quality control technique commonly used by laboratories with a quality system including regulatory laboratories, commercial laboratories, diagnostic laboratories and others pursuing accreditation under the ISO/IEC 17025 standard [12, 24, 28]. The PT is conducted by distributing a test material (usually reference material) by the PT provider to participating

*Corresponding author: Timothy J. Herrman , Phone: 979-845-1121 , Fax: 979-845-1389, Email: tjh@otsc.tamu.edu

laboratories. The participants analyze the test material for the target analyte without being informed of its concentration. The participants return their results by a specified closing date and the provider then evaluates the reported results by laboratory, procedures used to measurement the analyte(s), and the values for analytes, which may be a consensus mean or assigned mean. The participants' performance is reported by a z-score which may be calculated using either a consensus or assigned mean and standard deviation. AAFCO utilizes a consensus mean and robust standard deviation that are estimated after excluding a small portion of outliers. The International Organization for Standardization (ISO) published a general guide in organizing and performing the PT program as well as statistical methods for data analysis and estimation [17]. The International Harmonized Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories has been published and revised in collaboration of ISO, IUPAC, and AOAC international to better unify the designs and methods of organizations and statistical features of PT [26, 28].

The Harmonized Protocol recommends an examination of z-scores given by $z = (x - X)/\sigma$, where x is the participating laboratory's result, X is the assigned value of the test material, and σ is the target value for standard deviation. The σ parameter is the maximal permissible error among the laboratories. The z-score indicates the degree of deviation of a participant result from the normal distribution of the mean and a variance of σ^2. The probability of the participant that has z-score outside the range ±2 is only about 5% while the chance of the value outside the range ±3 is below 1%. The PT provider should initiate remedial action when large and consistent discrepancies of the assigned value and the fitness-for-purpose-based standard deviation with consensus values of the participants occur. Horwitz proposed a simple equation for evaluating large data sets including those available from PT and collaborative trials: the relative standard deviation of an analytical method among laboratories,$\mathrm{RSD_R}$ (%) = $2^{(1-0.5logC)}$, where C is a concentration expressed in mass fraction of the analyte [10]. The Horwitz equation was derived based on data from many inter-laboratory studies of the Association of Official Analytical Chemists and seems to hold the general relationship between the relative standard deviation among laboratories and concentration, irrespective of matrix, analyte, and analytical method [4, 10, 13, 25]. Laboratories that yield relative standard deviations much greater than the value obtained from the Horwitz equation may be considered performing in an unacceptable manner and should investigate the source of the discrepancy. The Horwitz function has been experimentally verified and found to be valid for a variety of analytes and analytical conditions as well as under extensive statistical testing [12, 27]. Thompson and Lowthian [25] examined the consistency of Horwitz equation with the collaborative trial data used for development of Horwitz equation and found that in the range of mass factions of $10^{-6.92}$ to 0.1, called the Horwitz region, the estimated values were close to those predicted by the equation. However, the large discrepancy of the data with the predicted values was observed at a lower concentration presumably due to higher random variations at those concentrations [27].

The test procedure employed in analytical chemistry generally consists of 3 steps including sampling, sample preparation, and chemical analysis. The relative contribution of these steps to the variation in analytical results have been extensively studied by the authors [1, 7, 9, 29, 32]. The total variation: $\sigma_{total} = \sigma_{sampling} + \sigma_{sampleprep} + \sigma_{analytical}$. The variability associated with different steps makes it hard to determine the true concentration of a bulk lot and accurately classify lots for risk management and regulatory decisions [7]. To reduce misclassification of sample lots, it is important to properly design sampling plans and evaluate the variability characteristics associated with the test procedure [32]. In addition to the three steps, the degree of heterogeneity of the material being sampled, sample mass/volume, and number of increments contribute to the random error (measurement uncertainty) of analytical results [20].

Measurement uncertainty (MU) is a parameter that is used for expressing a measurement result with an associated level of uncertainty due to a lack of knowledge [6, 11, 30]. Accreditation bodies require an estimate of MU for testing laboratories to comply with international quality standards such as ISO/IEC 17025:2017 [16]. MU is one of the process components to increase confidence in the measurement result rather than an alternative validation procedure or doubt on the appropriateness of the result [6, 11, 30]. As a result of MU stemming from random errors, all decisions should be followed by gains or losses [8, 33]. MU may result from diverse sources which can be easily quantified or very difficult to deal with in the system, including measurement error, systematic error, natural variation, inherent randomness, model uncertainty, and subject judgement [22]. There are several approaches to calculate MU that are very specific to the applying discipline and available budget in order to obtain satisfactory estimates of MU [11, 31]. The most common and generally accepted approach is to carry out the inter-laboratory study as described in the International Harmonized Protocol or ISO/IEC protocols. As stated earlier, the Horwitz equation to the anticipated concentration can be a useful option to obtain rough calculations of among-laboratory reproducibility, a within-laboratory repeatability, and expanded measurement uncertainty even before starting the analysis of the reported results. In some areas, suitable estimate of MU is mean value ±2 standard deviation and 2 CV (coefficient of variation = SD/mean value x 100%) that is approximately equal to a level of 95% confidence [6]. A combined standard uncertainty $u_c(y)$ is a sum of the square root of the total variance of all the uncertainty components and also estimated by other alternative methods. An expanded uncertainty (U) that is obtained by multiplying the combined standard deviation by a coverage factor (k) is frequently used to provide an interval within which a true value of measurement likely fall. The coverage factor can be chosen according to the level of the desirable confidence (e.g $k = 2$ for 95% confidence interval) [6]. PT that is aimed to test laboratory's performance periodically can provide useful data and information for estimating and evaluating the laboratory's measurement uncertainty using test items given under fulfilled conditions and reasonable approaches because the laboratory results should not deviate significantly from the robust mean

over the number of PT rounds [6, 28].

For the last decade, AAFCO working group and committee have been formed and reviewed the AV values using statistical approaches for comparison, but no conclusions and reports have been made. Typically, a feed product is sampled by a state feed regulatory agency and shipped to the state feed laboratory to be processed prior to the selection of a test portion. The test portions are drawn from the analytical sample and tested. During the series of processes, error propagates as the square root of the sum of errors, so the largest errors have the greatest impact. Error includes both random and systematic error (bias). There are major errors that are not evaluated in proficiency testing include those associated with field sampling (primary sampling) and laboratory sampling [6, 11]. Given the results and findings described above, a need exists to re-evaluate the application of AAFCO AVs and to explore alternative methods and models to evaluate test results in context with the regulatory limit.

2. Materials and Methods

2.1. Data collection and evaluation of AAFCO AVs

AAFCO proficiency test (PT) results from the 2014-2018 check sample program were used to evaluate the existing AV values. In total there were 86 proficiency test reports used during this period of the program. The reports included a wide range of feed types to account for different nutrient compositions in animal feed. Animal feed PT items were sent out monthly while the pet food ingredient PT items were sent on a quarterly basis. The AAFCO PT results from participating laboratories were submitted through the AAFCO online portal for the program. The number of laboratories participating in each of the 86 PT rounds offered during this period and the number of analytes per round are reported in Table 1. For evaluation of the PT results, all 86 reports were combined to form a single dataset and results exceeding a z-score of above 3 or below -3 were excluded due to their high degree of influence on mean and other statistics and thus may not appropriately define the model. Eighteen analytes including two different methods for fat analysis and three different methods for fiber analysis were evaluated for current AAFCO AV values in this study, along with their AAFCO AV applicability concentration range for which the AV is applicable and the number of PT test items whose concentrations fell outside that concentration range. Therefore, the data deviated from the main data were removed from the entire datasets before proceeding with an assessment of the suitability of the AVs to explain the normal variance in the AAFCO PT program.

2.2. Estimation of measurement uncertainty

The descriptive statistics summarizing each of these 86 reports were compiled and used to compare the proficiency test results with the AAFCO AVs calculated for 15 nutrients and 18 analytes (Table S1). The AAFCO PT program has provided the combined standard uncertainty $u_c(y)$ of the assigned value through the PT rounds:

$$\mu_c(y) = \frac{1.25 \cdot s}{p}$$

where 1.25 is the factor to adjust robust standard deviation (s) to estimate of the normal standard deviation, and p is the number of the participating laboratories in the PT round. However, for the present study, the relative standard deviation (coefficient of variation, CV), Horwitz value (RSD_R), and the expanded uncertainty (U) for each PT round and for each nutrient were estimated as follows:

$$RSD = \frac{s}{X} \cdot 100$$
$$RSD_R = 2^{(1-0.5logC)}$$
$$U = \frac{2.0 \cdot s}{p}$$

where X is the robust mean of the assigned value and C is a concentration expressed in mass fraction of the analyte. Expanded uncertainty (U) was calculated with a coverage factor ($k = 2$) for 95% confidence interval. The robust means and standard deviations were derived from ISO 13528:2005 [14]. A comparison was performed between the estimated measurement uncertainties for each PT round.

2.3. Development and validation of new AV values

The dataset for each analyte was analyzed to develop new AVs for a total of 18 analytes using SAS software (ver. 9.4, SAS Institute, Cary, NC). The AAFCO PT program reported several analytes (fat and fiber) under different analytical methods. Specifically, fat results were reported separately using the methodology to analyze proficiency test items for crude fat (Method 3) and acid hydrolysis (Method 13) fat and fiber results were reported using the methodology to analyze proficiency test items for crude fiber (Method 4), acid detergent fiber (Method 8), and neutral detergent fiber (Method 9). The separate regression model analyzed results by an analytical method. The AAFCO PT 2019-2020 datasets independent from those used to develop the models to calculate revised AVs were used to validate the revised AVs.

Microsoft SQL Server ver. 2014 database management system (Microsoft Corporation, Redmond, WA) was used to sort and retrieve data from the OTSC laboratory information management system (LIMS) that is built on Thermo Scientific Nautilus ver. 9.3 software.

3. Results

3.1. An overview of the AAFCO PT results

Table 1 shows the distribution of AAFCO proficiency test (PT) results by feed type with the number of participating laboratories and the number of tested analytes from 2014 to 2018. They were obtained for evaluation of AV values and development of a regression model. In 2014, the number of laboratories ranged from a high of 300 laboratories testing dry cat food to a low of 47 laboratories testing canned cat food. The analyte

Table 1: Description of datasets of proficiency test results from AAFCO check sample program

Year	Feed Type	# of laboratories submitting the PT results		# of analytes
		Min	Max	
2014	Beet pulp	5	155	45
	Calf starter/grower, medicated	6	209	58
	Canned cat food	5	47	40
	Chicken starter, medicated	5	203	56
	Dry cat food	6	300	56
	Equine feed	7	213	58
	Ewe developer gestation feed, medicated	10	214	35
	Fish meal - menhaden	5	154	40
	Horse mineral	6	135	55
	Marine & animal protein concentrate	6	213	54
	Medicated swine starter	5	211	62
	Non-fat dried milk	5	149	35
	Pelleted beef feed, medicated	5	216	55
	Pelleted sheep concentrate, medicated	7	212	56
	Potato flour	5	158	39
	Poultry layer feed	5	212	54
	Show pig primer, medicated	5	204	60
	Swine grower, medicated	5	155	55
2015	Barley, pet food,	5	57	64
	Beef supplement	5	208	42
	Brewers dried yeast	5	57	46
	Brown rice	5	52	60
	Calf feed, medicated	5	215	62
	Chick starter, medicated	5	224	62
	Dairy feed, medicated	5	209	52
	Distillers dried grains with solubles	5	210	62
	Dog food	5	234	63
	Equine feed	5	224	57
	Fish food	5	199	60
	Goat feed, medicated	5	212	60
	Milk replacer, medicated	5	208	41
	Pork and bone meal	5	59	46
	Rabbit feed	5	217	59
	Swine starter, medicated	5	196	61
	Wine mineral and vitamin supplement	5	189	62
2016	Cattle feed supplement	5	174	67
	Cattle mineral	5	137	59
	Cheese powder, pet food	5	48	38

Year	Feed Type	# of laboratories submitting the PT results		# of analytes
		Min	Max	
2016	Dairy beef feed	5	199	38
	Dry cat food	5	215	62
	Goat feed	5	211	63
	Lamb feed, medicated	5	198	63
	Lamb meal	5	56	44
	Llama feed	5	198	64
	Molasses product, dehydrated	5	186	56
	Oat flour	5	53	43
	Poultry feed, medicated	5	201	64
	Poultry/game bird feed	5	221	59
	Swine feed, medicated	5	207	62
	Tomato pomace, pet food	5	58	46
2017	Barley, pet food	5	61	45
	Beef feed, medicated	5	191	61
	Canola meal	5	57	42
	Dairy beef feed, medicated	5	193	64
	Dairy feed minerals	8	29	13
	Dog food mineral	10	29	14
	Dry dog food	5	188	60
	Equine feed	5	192	62
	Fish food, carnivore	5	191	55
	Fish meal	5	197	57
	Flaxseed meal	5	56	47
	Lamb feed, medicated	6	192	58
	Llama & alpaca mineral	5	129	56
	Meat and bone meal	6	56	41
	Milk replacer, medicated	5	193	59
	Pheasant & turkey feed	5	189	56
	Poultry layer feed, medicated	5	196	63
	Swine feed, medicated	5	199	66
	Swine feed, mineral	12	26	13
2018	Beef feed	5	187	64
	Blood meal, pet food	5	51	38
	Bone meal, pet food	5	60	14
	Bone meal mineral	6	25	42
	Brewers dried yeast	7	51	39

Year	Feed Type	# of laboratories submitting the PT results		# of analytes
		Min	Max	
2018	Cattle feed, medicated	5	198	57
	Chicken starter, medicated	5	192	59
	Cottonseed meal	5	197	53
	Dairy feed	7	26	14
	Dry cat food	5	224	67
	Equine crumbles	5	186	60
	Equine feed	5	177	56
	Fish feed, mineral	8	30	14
	Lamb feed	5	193	60
	Milk replacer	5	196	57
	Pig feed, medicated	5	197	62
	Poultry feed	6	204	37
	Rabbit feed	5	202	56
	Sweet potato powder, pet food	5	51	43
	Swine feed, medicated	5	207	61
	Turkey feed, mineral	9	27	14

number ranged from a high of 62 for swine starter to a low of 35 analytes for ewe developer feed or non-fat dried milk in the same period. In 2015, dog food items were analyzed by 234 participants (high) while brown rice exhibited the lowest number of participating laboratories (52). The highest level of participation was observed for dry cat food tested by 215 laboratories and the low number of participating laboratories (48) were for cheese powder in the 2016 PT activity. In 2017, the medicated swine feed test items were tested by the most laboratories (199) and mineral feed for swine involved the lowest laboratories (26). During the 2018 AFFCO PT program, dry cat feed products were the most tested items by the laboratories (224) as observed in the 2014 and 2015 PT while the lowest number of laboratories (27) participated in testing mineral feed for turkey.

Table 2 shows the PT results for select 18 analytes with their corresponding AAFCO AV concentration range applicable to the AV, the number of datasets for each analyte, and the percentage of datasets of PT test results outside the applicability concentration range, based on the assigned value. If the AAFCO AV applicability concentration for select analyte is at the level of ppm (mg/kg) as selenium in Table 2, the AAFCO PT test results for the analyte are most likely fall within the AV applicability concentration range. Contrarily, the concentrations of all the PT test items for cobalt were lower and out of the concentration ranges during the 2014-2018 AAFCO PT rounds, indicating the AAFCO AV applicability concentration range is out-of-date and needs to be replaced. Fat and fiber analytes also revealed higher percentages outside the concentration range for the respective AV. The different testing methods for fat and fiber showed the range of 16.3 and 30.8% outside the AV concentration range through the PT rounds.

3.2. Measurement uncertainty quantification in the AAFCO PT program

Since the publication of the AAFCO AVs, laboratories participating in the AAFCO PT program that are ISO accredited to the 17025 standard [15, 16] need to report analytical uncertainty. The AAFCO PT program achieved ISO/IEC 17043:2010 accreditation in 2017, which necessitated a change in the PT management systems and statistical approach to calculating and reporting PT results. In addition to the quality management practices, the instrumentation and analytical methods that are reported to in the AAFCO PT program differ from those practiced when the AV concept was originally introduced and updated. The organizations for chemists working in analytical chemistry have provided general information and guides about how to quantify measurement uncertainty using the PT data [6, 18]. AAFCO also reports measurement uncertainties for the assigned values for each analyte, which may be considered to obtain one of the uncertainty components that are used for comparing with the label guarantee. The laboratory's participation in the PT program can largely help the laboratory fully and correctly evaluate its measurement uncertainty. However, the AAFCO PT data should be used to estimate the measurement uncertainty only in the conditions in which the concentrations

Table 2: Proficiency test (PT) results out of the current AAFCO AV applicability concentration range of selected analytes

Analyte[a]	AAFCO Concentration range (%)[b]	Total No. of datasets[c]	Out of AAFCO AV applicability concentration range[d]	
			n	%
Ash	2-88	86	6	7.0
Calcium	0.5-25	86	4	4.7
Cobalt	0.01-0.16	78	78	100.0
Fat (Method 3)	3-20	86	14	16.3
Fat (Method 13)		86	28	32.6
Fiber (Method 4)	2-30	86	25	29.1
Fiber (Method 8)		82	11	13.4
Fiber (Method 9)		80	18	22.5
Iron	0.01-5.00	86	14	16.3
Lysine	0.5-4.0	83	14	16.9
Magnesium	0.01-15.0	86	0	0.0
Manganese	0.01-15.0	86	42	48.8
Moisture	3-40	83	1	1.2
Phosphorus	0.5-20.0	86	11	12.8
Potassium	0.04-8.00	86	0	0.0
Protein	10-85	86	10	11.6
Selenium	ppm	86	0	0.0
Zinc	0.002-6.00	86	2	2.3

[a] Methods 3 and 13 for fat analysis are for determination of crude fat and acid hydrolysis fat, respectively. In the meantime, Methods 4, 8, and 9 for fiber analysis are for determination of crude fiber, acid detergent fiber, and neutral detergent fiber, respectively.
[b] The current AAFCO concentration range of analyte that the AV can apply for.
[c] The total number of datasets of proficiency test results.
[d] The number of AAFCO datasets with mean values out of the concentration range of analytes defined in the AAFCO official publication.

of PT test items cover the routine analytical range of actual regulatory samples and the variation in the PT test results among the participating laboratories is minimal which should provide the low uncertainty of the assigned value [5, 6]. Therefore, the quality and suitability of the PT test items such as sufficient homogeneity and stability are critical and should be assessed prior to the distribution of the test items to the participating laboratories [14].

3.3. Comparison and analysis of expanded uncertainty estimates

The relative expanded uncertainties and relative standard deviations estimated based on the AAFCO PT 2014-2018 data are presented in Table 3, along with the current AAFCO AV. Figures 1 and S1 also show the relationship between relative expanded uncertainty and analyte concentration and comparison of the uncertainty with 2 CVs and the relative standard deviations (Horwitz values) predicted by Horwitz equation. Table 3 shows that the relative expanded uncertainties of most analytes were below 5%. The range for the uncertainty values through the PT rounds was wide due to higher imprecision at lower concentrations and for those rounds outside the AAFCO AV applicability concentration ranges. The predicted Horwitz values were generally in good agreement with the observed relative standard deviations (RSDs) for most analytes including ash, calcium, cobalt, iron, lysine, magnesium, manganese, phosphorus, potassium, protein, and zinc within the AAFCO AV applicability concentration range of the analytes (Figures 1 and S1). The majority of dataset points from the PT rounds that falls within the Horwitz band and fulfills 0.5 – 2.0 times of the Horwitz values indicate that the reproducibility precision of the analytical methods employed in the PT rounds are

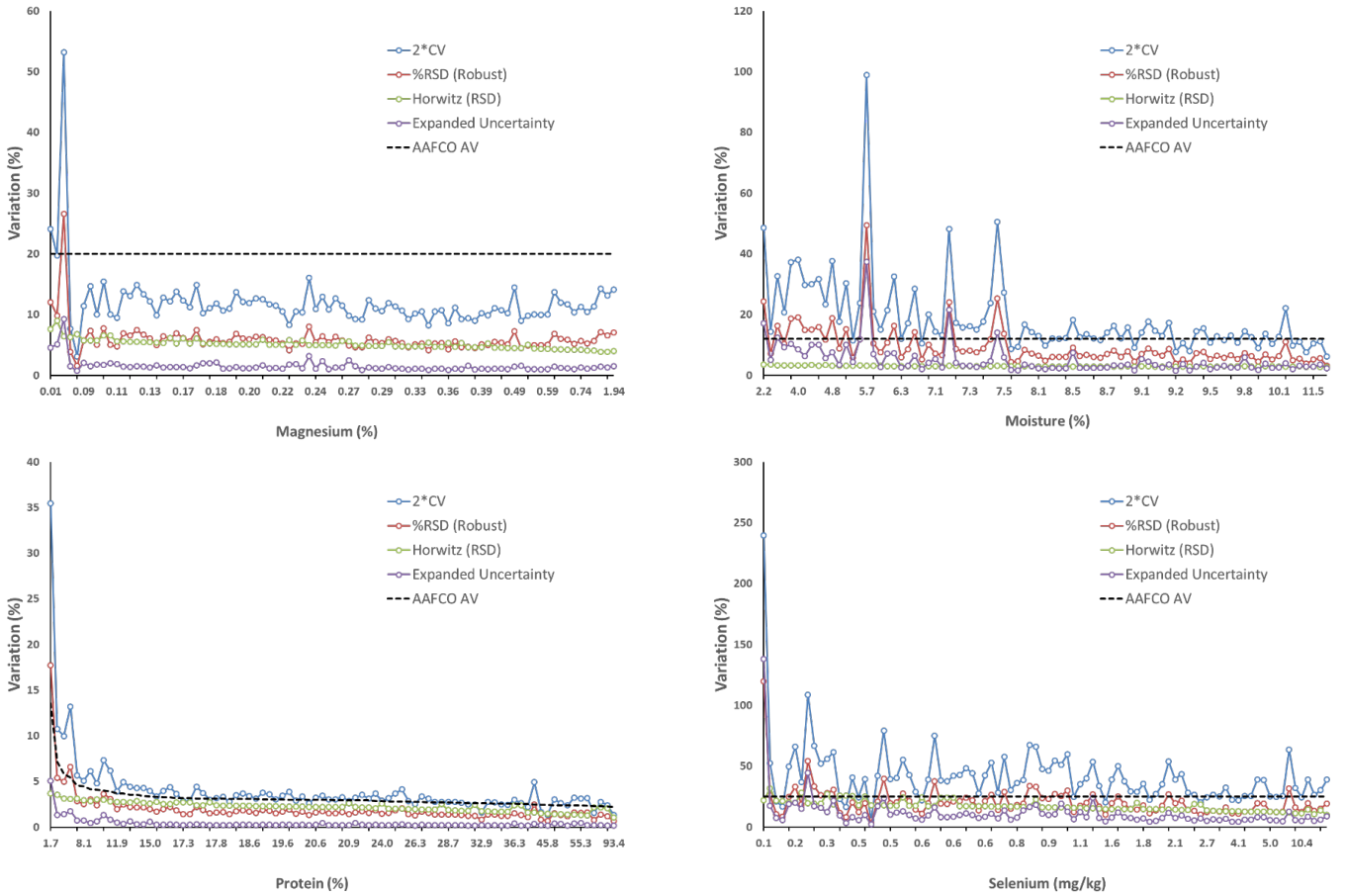

Figure 1: A scatter plot of assigned values of select analytes (magnesium, moisture, protein, and selenium) vs relative expanded uncertainties and relative standard deviations estimated or predicted using the 2014-2018 AAFCO Proficiency Test data: AV, analytical variation (%); RSD, relative standard deviation (%); Horwitz, relative standard deviation (%) predicted by Horwitz equation; and 2*CV, two times coefficient of variation (%).

considered acceptable and successful [23, 27]. However, the larger discrepancy at the lower concentrations and outside the AAFCO AV applicability concentration ranges were also observed. For those analytes, the relative expanded uncertainties of the assigned values were smaller than the Horwitz values within the AAFCO AV applicability concentration range. The expanded uncertainty for fat, fiber, and moisture was close to the Horwitz values in the AAFCO AV applicability concentration range. The inconsistency in the magnitude of measurement uncertainty among analytes through the PT rounds indicates that the contribution of analytical imprecision varies among analytes and methods. The previous study also implies that more reliable measurement uncertainties may be obtained after the laboratory acquired the minimum number of test results and participated in the minimum consecutive trials [21].

3.4. Comparison and analysis of 2 CV estimates

Analytical laboratories often utilize 2 CVs to assign upper and lower warning limits and three CVs are designated as control chart limits using working controls (often reference material) to monitor intra-laboratory variation. For PT programs, a laboratory that is continuing to report analytical results with greater or less than 2 CVs implies a systematic error and needs to investigate the entire analytical process for corrective action(s) to remove or minimize the systematic error. The summary performance statistics for a PT round report a CV that is approximately double the test/reference material CV. In the 2014-2018 AAFCO PT data, the AV in the defined concentration range appeared to be larger than other relative standard deviations for most analytes, except for cobalt, fat, fiber, moisture, and selenium where the calculated variability (2 CV) exceeded the AAFCO AV (Figures 1 and S1). The higher 2 CV over the AAFCO AV implies the published values in the OP for select analytes poorly explain the inter-laboratory variability by the AV and need to reflect the current practice of the laboratories. For those analytes, the high percentages of PT test items out of the AAFCO AV applicability concentration ranges coincided with significantly higher 2 CVs over the AAFCO AV. The PT data also showed the expanded uncertainties for most of these analytes do not meet the ISO criteria against the standard deviation for PT ($U \leq \sigma_{PT}$) (Table 3). These observations indicate a need to reevaluate/model new AVs for those analytes. On the other hand, if PT participating laboratories employ diverse analytical methods to report the results for a single

Table 3: Comparison of relative interlaboratory standard deviations and uncertainty calculated using the 2014-2018 AAFCO Proficiency Test (PT) data with those predicted by Horwitz equation

Analyte[a]	AAFCO AV (%)[b]	Horwitz (%)[c]		2 CV (%)[d]		Relative expanded uncertainty (%)		≥ $0.3\sigma_{PT}$[e]	
		Ave	Range	Ave	Range	Ave	Range	*n*	%
Ash	45/x+3	2.8	1.1-4.2	5.7	1.3-31.0	0.6	0.2-5.0	17	19.8
Calcium	14/x+6 /10/12	4.1	2.4-8.0	13.4	6.7-49.7	1.6	0.7-9.5	21	24.4
Cobalt	25	16.0	9.0-36.3	51.0	15.9-172.8	13.2	3.5-53.0	78	100.0
Fat (Method 3)	10	3.4	1.7-6.0	31.5	3.9-258.5	4.7	0.3-66.7	24	27.9
Fat (Method 13)		3.1	1.6-4.7	27.6	4.8-111.4	4.5	0.8-27.9	27	31.4
Fiber (Method 4)	30/x+6	3.4	1.7-6.0	43.7	9.3-239.6	7.2	0.9-56.5	24	27.9
Fiber (Method 8)		3.1	1.5-4.7	40.1	8.2-220.0	9.4	1.2-89.8	32	39.0
Fiber (Method 9)		2.5	1.4-4.4	34.7	6.7-207.7	8.4	0.9-73.4	42	52.5
Iron	25	7.3	4.0-17.7	24.6	9.9-209.2	3.7	1.4-54.0	21	24.4
Lysine	20	4.1	2.9-6.4	12.2	4.2-67.7	3.3	1.0-25.6	83	100.0
Magnesium	20	5.2	3.9-9.0	12.0	3.1-53.2	1.5	0.7-9.3	21	24.4
Manganese	30	9.2	5.2-25.3	20.7	8.2-242.7	3.7	1.0-80.9	21	24.4
Moisture	12	3.0	2.8-3.6	18.7	6.2-98.8	5.1	1.4-37.4	83	100.0
Phosphorus	3/x+8	4.2	3.1-7.0	10.0	6.5-23.0	1.1	0.5-3.9	86	100.0
Potassium	15	4.1	3.3-5.8	12.0	8.1-47.4	1.5	0.8-8.3	21	24.4
Protein	20/x+2	2.2	1.0-3.7	4.0	1.3-35.5	0.4	0.2-5.1	1	1.2
Selenium	25	17.2	10.0-31.7	41.5	5.7-238.4	11.5	2.3-138.2	86	100.0
Zinc	20	7.7	4.6-12.6	17.7	8.5-132.9	2.5	1.1-32.2	21	24.4

[a] Methods 3 and 13 for fat analysis are for determination of crude fat and acid hydrolysis fat, respectively. In the meantime, Methods 4, 8, and 9 for fiber analysis are for determination of crude fiber, acid detergent fiber, and neutral detergent fiber, respectively.
[b] AV(%) represents current AAFCO analytical variation (AV, %). x = % guarantee. Calcium: (14x+6), 10, and 12 % at the ranges of 0.5-25%, 10-25%, and <<10%, respectively.
[c] Relative standard deviation (%) predicted by Horwitz equation
[d] 2 CV, two times of coefficient of variations (relative standard deviation).
[e] The number of dataset with standard deviation for proficiency testing greater than ISO 13528 (2005) criteria (0.3 σ_{PT})

analyte and thus multiple methods are associated with a single AV, inter-laboratory variation and measurement uncertainty could be larger. Robust mean and standard deviations could be also biased by using defective methodology. Under these conditions, the calculated expanded uncertainty of the assigned value is more likely to not fully meet the ISO criteria.

3.5. Development of statistical models for evaluation of the PT data

The AAFCO PT data were used to develop and validate the possible alternative models to reevaluate the current AV values and to better understand the relationship between measurement uncertainty and analyte concentration. For the statistical modeling of the AAFCO PT data, a 95% tolerance interval was applied. This means that we can claim with 95% confidence that the true measurement uncertainty would fall within that tolerance interval.

A one-sided upper tolerance interval was used to construct the model that was applied for predicting and determining the external validity of the AAFCO values. During the development of the models for measurement uncertainty, a greater variance at lower concentrations were observed indicating that the AV values in some of the analytes were more dependent on concentrations (Figures 1 and S1). At lower concentrations, the expanded uncertainty and other relative standard deviations were more spread out than at the higher concentrations. In the statistical modeling procedure, a simple linear regression line was fit to the data points and then the tolerance interval was found. The tolerance interval was not a single value, but a curved line was fitted to the interval. Another issue in developing the model for reevaluation of the AAFCO AV was that the concentrations of some analytes did have neither a high correlation with estimated measurement uncertainty values nor normal distributions. That part of the problem was solved by removing outliers or using a log-normal transformation.

Table 4: A new regression model developed on the 2014-2018 AAFCO PT data for determination of new AV values of select analytes[a]

Analyte[b]	2 CV			Expanded uncertainty (U)		
	A	B	r^2	A	B	r^{2c}
Ash	9.9	-0.007	> 0.99	1.1	-0.006	> 0.99
Calcium	18.9	-0.01	> 0.99	2.6	-0.005	> 0.90
Cobalt	89.1	-0.006	> 0.99	21.9	-0.01	> 0.99
Fat (Method 3)	64.0	-0.038	> 0.95	10.6	-0.01	> 0.95
Fat (Method 13)	56.3	-0.033	> 0.95	9.2	-0.029	> 0.95
Fiber (Method 4)	79.2	-0.006	> 0.95	15.2	-0.005	> 0.95
Fiber (Method 8)	80.7	-0.026	> 0.95	22.8	-0.005	> 0.95
Fiber (Method 9)	87.1	-0.006	> 0.99	21.4	-0.005	> 0.99
Iron	33.3	-0.006	> 0.95	4.2	-0.000006	> 0.90
Lysine	19.5	-0.039	> 0.99	6.0	-0.02	> 0.95
Magnesium	15.5	0.040	> 0.90	2.2	-0.107	> 0.95
Manganese	24.3	-0.006	> 0.90	6.1	-0.0004	> 0.99
Moisture	78.3	-0.006	> 0.99	17.9	-0.104	> 0.95
Phosphorus	13.0	-0.006	> 0.95 7	1.7	-0.005	> 0.99
Potassium	24.0	-0.105	> 0.99	2.8	-0.005	> 0.99
Protein	8.5	-0.009	> 0.99	0.6	-0.005	> 0.99
Selenium	96,8	-0.011	> 0.99	19.7	-0.005	> 0.99
Zinc	26.0	-0.006	> 0.99	4.6	-0.00005	> 0.99

[a] Regression model, 2CV (or U) = $A \cdot e^{Bx}$

[b] Methods 3 and 13 for fat analysis are for determination of crude fat and acid hydrolysis fat, respectively. In the meantime, Methods 4, 8, and 9 for fiber analysis are for determination of crude fiber, acid detergent fiber, and neutral detergent fiber, respectively.

[c] Coefficient of determination (r^2) values were estimated base on the upper bound of tolerance interval for new AV values.

Tolerance intervals are similar to confidence intervals, but the tolerance interval is a percent bound for the range of data, and confidence intervals are for one particular parameter of the data (e.g. mean). For the present study, a 95% tolerance interval was used. Therefore, at least 95% of the estimated measurement uncertainty values fall within that tolerance interval. The tolerance interval was also chosen to avoid overfitting the data. A new optimal AV value for each analyte could be determined based on the results of the 95% tolerance interval.

Although lower concentrations showed higher variability than higher concentrations in some analytes, the calibration models developed on the AAFCO 2014-2018 data showed a good linearity between analyte concentration and the estimated measurement uncertainties: expanded uncertainty and 2 CV regardless of the type of analyte (Table 4). Separate models were created for fat (crude fat and acid hydrolysis) and fiber (crude fiber, acid detergent fiber, and neutral detergent fiber) because different methods do not test for the same analyte. In fact, the AVs in the AAFCO OP were established only for specific analytical methods. Therefore, different methods for fat and fiber should not be grouped together. Many analytes in the AAFCO PT programs do not still have an AV calculation including some minerals (e.g. arsenic, chromium, and lead), vitamin Bs, vitamin D, vitamin E, and mycotoxins.

3.6. Validation and evaluation of the models

For the validation of the models, the AAFCO 2019-2020 PT dataset were used to confirm the reliability and validity of estimated measurement uncertainties and their regression models from the previous years' PT data (Figures 2 and 2S). The assigned values and robust standard deviation of the validation PT data fell between the expanded uncertainty and 2 CV models in most analytes that showed lower expanded uncertainty and similar RSD and Horwitz values (Figures 1 and 1S). However, the standard deviations accompanied by the assigned values were somewhat larger and fell within the wider predictive

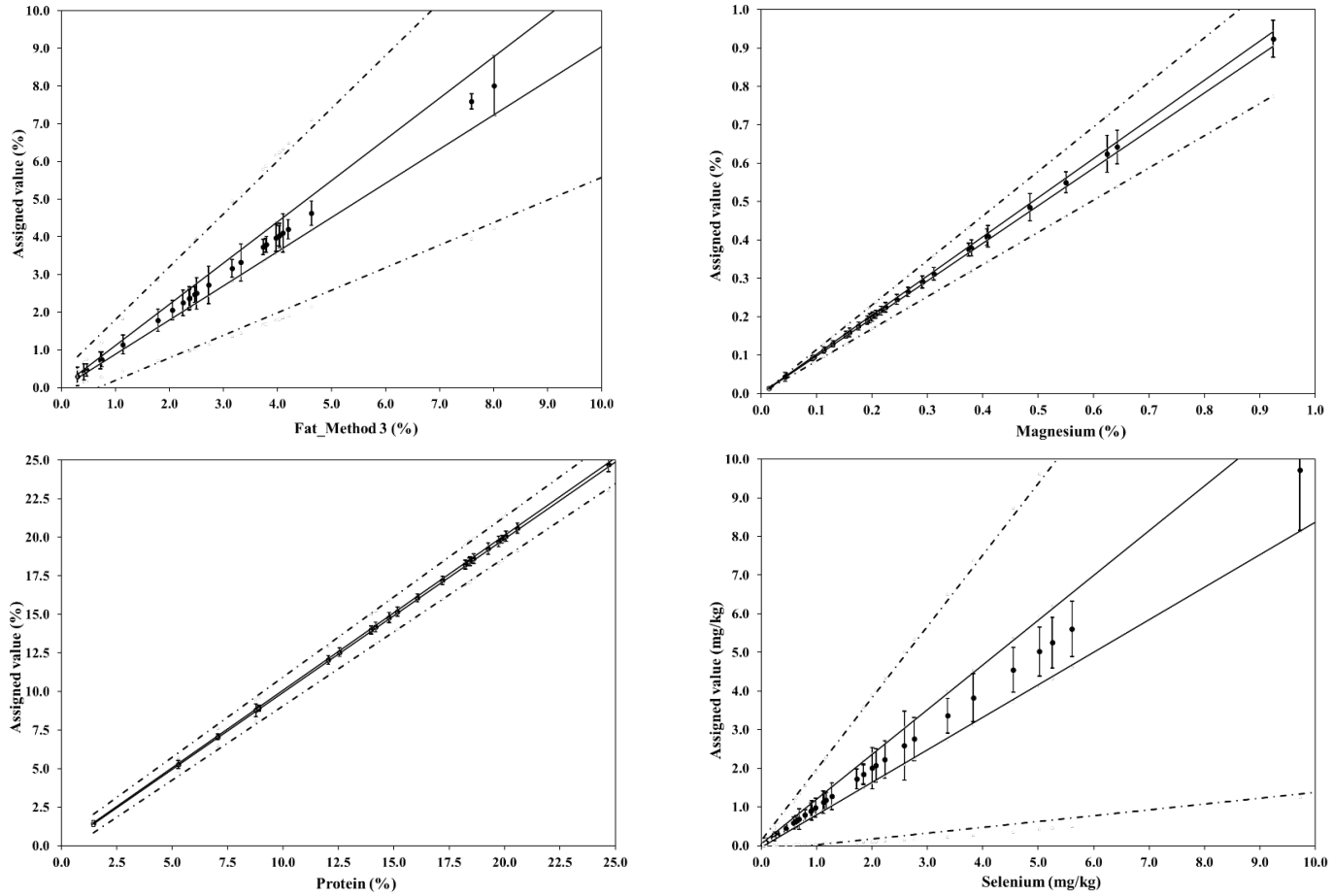

Figure 2: A scatter plot of the assigned value and its standard deviation obtained from the validation data and a regression line of the expanded uncertainty (*solid line*) and 2 CV (*dashed line*) models for fat (method 3), magnesium, protein, and selenium. Error bars represent standard deviations of the assigned value.

intervals of the expanded uncertainty models over the concentration ranges for some analytes (e.g. cobalt, fiber, moisture, and selenium) whose 2 CV values are greater than the current AAFCO AV. This can be attributed to a significant variation in laboratory measurements through the PT rounds and thus the estimated uncertainties used for developing the regression models, as particularly seen in selenium. Meanwhile, with the developed regression models for fat PT data, the standard deviation range of the assigned value fell within the predictive interval of expanded uncertainty model at higher concentrations while it likely exceeded the predictive uncertainty interval at lower concentrations (Figure 2). These analytes appeared to have had greater difficulties and inconsistency in analyzing PT test items at the lower concentrations.

3.7. Limits and challenges of the AAFCO AV

The observations and findings from the present study imply most of the current AV values may be appropriate while the AV values of some analytes needs to be re-examined. As reported in the previous studies [12, 25] and most experts agree, the present study also showed the measurement uncertainty decreased as analyte concentration increased and a greater variance was observed at lower analyte concentrations following the Horwitz function. Besides, certain analytes clearly demonstrated a better suitability of fitness to the concept of establishing the AAFCO AV than others through the PT data. This can be explained by the fact that the AVs are unique to analyte of interest, analyte concentration, sample matrix, and analytical method. There should be also other influential factors on the AVs including unique analytical variations associated with each individual method and laboratory procedure. However, such information is difficult to be extracted from the given PT data.

4. Conclusion

The objective of this study was to evaluate the current AV value which has not been performed since 1986 and the data and statistical methodology used to create the AV value are no longer available. While the current conversation within the AAFCO is that a need exists to consider a global estimation error including a total variability associated with primary sampling, laboratory sampling, and testing – this paper serves to benchmark the current functionality of AAFCO AVs. This study revealed that many of the AVs are currently applicable for their intended purpose while the analytical methods for cobalt, fat,

fiber, moisture, and selenium presented a challenge. The estimated new AVs created using regression modeling and validated using measurement uncertainties did not fit the AVs for the these analytes. Despite improvements with new AVs, some analytes still remain a challenge to measure within the expanded AVs. For these and existing AVs – pursuit of investigational allowances that consider primary sampling and laboratory sampling error may not offer the solution some within the regulatory and regulated community expect. The revised AVs based on recent AAFCO PT results appeared to better capture and represent analytical variability for select analytes. However, the hypothesis drawn from the current study results and findings needs to be tested as well as the suitability of the revised AVs proposed in this paper to actual regulatory data.

The present study also provided useful information about the measurement uncertainties of the select analytes relevant to the current AAFCO AVs, which would allow the PT provider and participating laboratories to identify and recognize analytes required for improvement of participating laboratory's performance or analytical methodologies. The observations and findings from the present study also indicated that the current AV values could be considered acceptable while the analytical methods employed in the AAFCO PT rounds are in general precise and reproducible. The AAFCO PT program is still one of the best ways to trace and estimate the uncertainty of the analytical results of feed regulatory agencies' laboratory.

5. Declaration of Conflicting Interest

The authors declare no conflicts of interest.

6. Acknowledgement

The authors are thankful to AAFCO to allow the use of the proficiency test results from the 2014-2020 proficiency testing program for the present study. We also thank Karen Fischer who performed the preliminary analysis of AAFCO PT data.

7. References

[1] Association of American Feed Control Officials (AAFCO). (2018). GOOD Test Portions: Guidance on Obtaining Defensible Test Portions. AAFCO, Champaign, IL. http://www.aafco.org/Publications/GOODTestPortions.

[2] AAFCO. (1986). Official publication. Association of American Feed Control Officials.

[3] AAFCO. (2000). Official publication. Association of American Feed Control Officials.

[4] Boyer, K.W., Horwitz, W., & Albert, R. (1985). Interlaboratory variability in trace element analysis. *Analytical Chemistry, 57(2)*, 454-459.

[5] Earnshaw, A, Smith, R. A., & Owen, L. (2009). How proficiency testing can improve the quality of analytical data using vitamin analysis as an example. *Food Chemistry, 113(3)*, 781-783.

[6] Ellison, S. L. R., & Williams, A. (2012). Quantifying uncertainty in analytical measurement. 3rd ed. Eurachem/CITAC guide. http://www.eurachem.org.

[7] Farkas, Z., Slate, A., Whitaker, T. B., Suszter, G., & Ambrus, Á. (2015). Use of combined uncertainty of pesticide residue results for testing compliance with maximum residue limits (MRLs). *Journal of Agricultural and Food Chemistry, 63(18)*, 4418-4428.

[8] Fenton, N., & Neil, M. (2018). Risk assessment and decision analysis with Bayesian networks. Boca Raton, FL: CRC Press.

[9] Fussell, R. J., Hetmanski, M. T., Macarthur, R., Findlay, D., Smith, F., Ambrus, Á., & Brodesser, P. J. (2007). Measurement uncertainty associated with sample processing of oranges and tomatoes for pesticide residue analysis. *Journal of Agricultural and Food Chemistry, 55(4)*, 1062-1070.

[10] Horwitz, W. (1982). Evaluation of analytical methods used for regulation of foods and drugs. *Analytical chemistry, 54(1)*, 67-76.

[11] Horwitz, W. (2003). The certainty of uncertainty. *Journal of AOAC International, 86(1)*, 109-111.

[12] Horwitz, W., Britton, P., & Chirtel, S. J. (1998). A simple method for evaluating data from an interlaboratory study. *Journal of AOAC International, 81(6)*, 1257-1266.

[13] Horwitz, W., Kamps, L. V. R., & Boyer, K. W. (1980). Quality assurance in the analysis of foods for trace constituents. *Journal of the Association of Official Analytical Chemists, 63(6)*, 1344-1354.

[14] International Organization for Standardization. (2005). ISO 13528:2005 Statistical method for use in proficiency testing by interlaboratory comparison. ISO, Geneva.

[15] International Organization for Standardization. (2010). ISO 17043:2010 Conformity assessment-general requirements for proficiency testing. ISO, Geneva.

[16] International Organization for Standardization. (2017). ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories. ISO, Geneva.

[17] International Organization for Standardization. (1997). ISO/IEC Guide 43–1 Proficiency testing by interlaboratory comparisons—Part 1: Development and operation of proficiency testing schemes. 2nd ed. ISO, Geneva.

[18] JCGM. (2012). Evaluation of measurement data–the role of measurement uncertainty in conformity assessment. JCGM 106.

[19] Johnson, R. L., Latimer Jr., G. W., & Spiegelman, C. (1994). Use of trimmed duplicates derived from laboratory data to estimate standard deviation. *Journal of AOAC International, 77(6)*, 1660-1663.

[20] Omeroglu, P. Y., Ámbrus, Á., Boyacioglu, D., & Majzik, E. S. (2013). Uncertainty of the sample size reduction step in pesticide residue analysis of large-sized crops. *Food Additives & Contaminants: Part A, 30(1)*, 116-126.

[21] Patriarca, M., Chiodo, F., Castelli, M., & Menditto, A. (2006). Estimates of uncertainty of measurement from proficiency testing data: a case study. *Accreditation and Quality Assurance: 11(8)*, 474-480.

[22] Regan, H. M., Colyvan, M., & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications, 12(2)*, 618-628.

[23] Taverniers, I., De Loose, M., & Bockstaele, Van E. (2004). Trends in quality in the analytical laboratory. I. Traceability and measurement uncertainty of analytical results. *TrAC Trends in Analytical Chemistry, 23(7)*, 480-490.

[24] Thompson, M. & Lowthian, P. J. (1996). Statistical aspects of proficiency testing in analytical laboratories. Part 1. Ranking of participants on scores is misleading. *Analyst, 121(11)*, 1589-1592.

[25] Thompson, M. & Lowthian, P. J. (1997). The Horwitz function revisited. *Journal of AOAC International, 80(3)*, 676-680.

[26] Thompson, M., & Wood, R. (1993). The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories (Technical Report). *Pure Appl Chem, 65(9)*, 2123-2144.

[27] Thompson, M. (2012). The characteristic function, a method-specific alternative to the Horwitz function. *Journal of AOAC International, 95(6)*, 1803-1806.

[28] Thompson, M., Ellison, S. L., & Wood, R. (2006). The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). *Pure and Applied Chemistry, 78(1)*, 145-196.

[29] Trucksess, M. W., Whitaker, T. B., Weaver, C. M., Slate, A., Giesbrecht, F. G., Rader, J. I., & Betz, J. M. (2009). Sampling and analytical variability associated with the determination of total aflatoxins and ochratoxin A in powdered ginger sold as a dietary supplement in capsules. *Journal of Agricultural and Food Chemistry, 57(2)*, 321-325.

[30] Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software, 63*, 24-31.

[31] Wallace, J. (2010). Ten methods for calculating the uncertainty of measurement. *Science & Justice, 50(4)*, 182-186.
[32] Whitaker, T.B., Hagler, W. M., Giesbrecht, F. G., & Johansson, A. S. (2000). Sampling, sample preparation, and analytical variability associated with testing wheat for deoxynivalenol. *Journal of AOAC International, 83(5)*, 1285-1292.
[33] Williams, A. (2008). Principles of the EURACHEM/CITAC guide "Use of uncertainty information in compliance assessment". *Accreditation and Quality Assurance, 13(11)*, 633-638.

8. Supplemental Data

Table S1: Descriptive statistics of the proficiency test results of each nutrient

	Mean	Standard Error	Median	Standard Deviation	Kurtosis	Skewness	Range	Count
Ash, %	12.78	1.36	8.02	12.64	11.35	3.06	1.25-76.61	86
Calcium, %	2.79	0.38	1.39	3.49	5.71	2.30	0.01-17.89	86
Cobalt, mg/kg	6.00	1.15	1.81	10.11	9.20	2.93	0.06-52.59	77
Fat (Method 3), %	5.28	0.52	3.68	4.85	15.46	3.08	0.18-34.83	86
Fat (Method 13), %	7.32	0.67	5.22	6.17	8.60	2.50	0.60-39.18	86
Fiber (Method 4), %	6.42	0.68	4.43	6.29	5.35	1.90	0.19-36.67	86
Fiber (Method 8), %	9.11	0.86	6.58	7.76	3.33	1.60	0.32-42.64	82
Fiber (Method 9), %	18.13	1.22	17.52	10.92	-0.28	0.52	0.57-50.02	80
Iron, mg/kg	711.33	156.91	344.95	1455.11	28.99	5.08	3.51-10593.0	86
Lysine, %	1.45	0.14	1.06	1.23	11.67	2.84	0.129-8.28	83
Magnesium, %	0.33	0.03	0.25	0.30	10.90	2.81	0.01-1.94	86
Manganese, mg/kg	211.88	41.54	104.54	385.19	15.16	3.69	0.34-2345.47	86
Moisture, %	7.77	0.23	8.27	2.13	-0.37	-0.41	2.19-12.16	83
Phosphorus, %	1.21	0.14	0.764	1.29	13.56	3.26	0.08-8.61	86
Potassium, %	1.13	0.06	1.05	0.58	3.08	1.25	0.22-3.60	86
Protein, %	26.36	1.82	20.77	16.91	3.57	1.71	1.73-93.44	86
Selenium, mg/kg	1.70	0.21	0.83	1.97	6.38	2.33	0.08-10.36	86
Zinc, mg/kg	567.77	101.37	176.95	940.09	5.03	2.42	4.78-3821.6	86

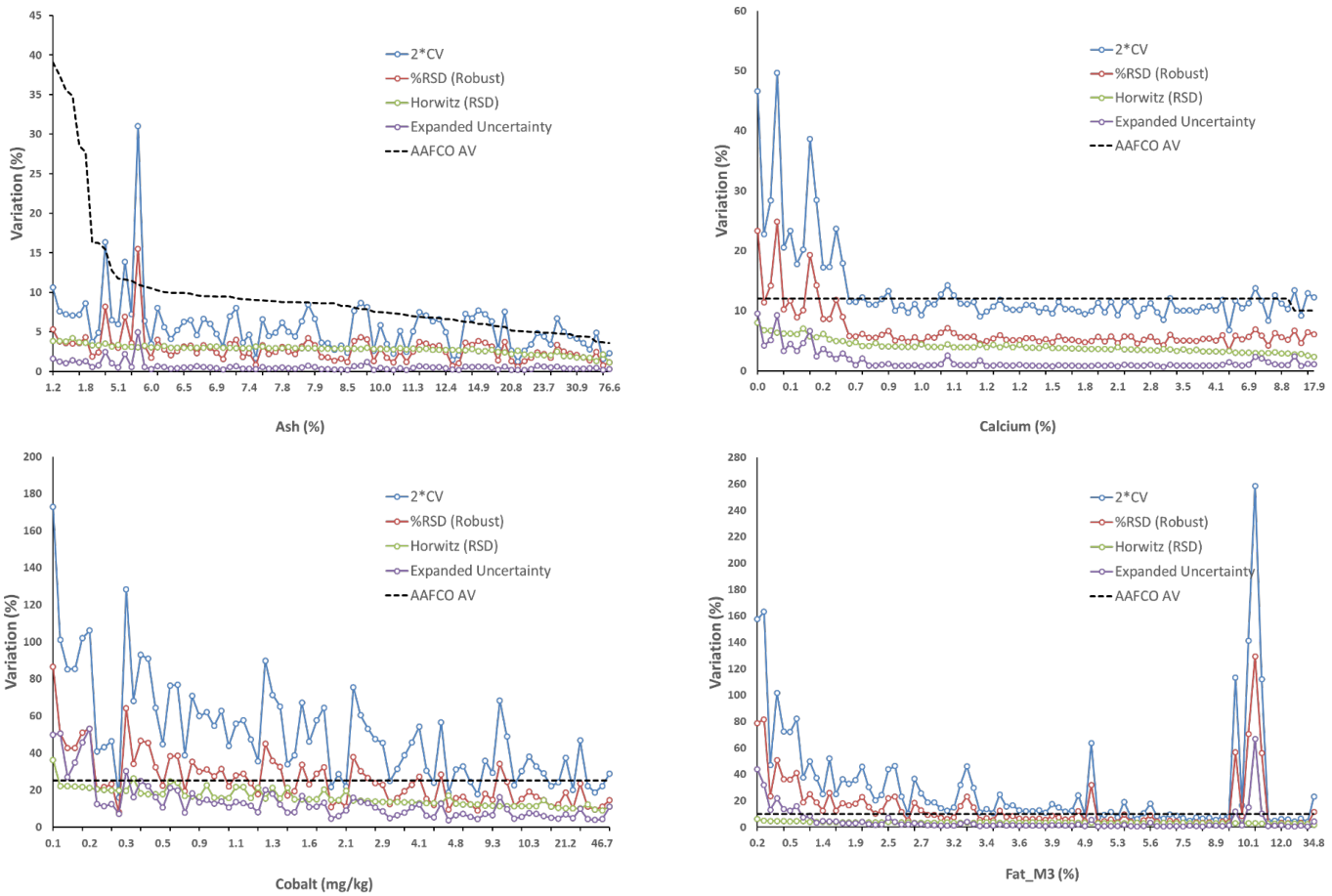
Variation (%)
2*CV
%RSD (Robust)
Horwitz (RSD)
Expanded Uncertainty
AAFCO AV
Ash (%)
Calcium (%)
Cobalt (mg/kg)
Fat_M3 (%)

Figure S1(A)

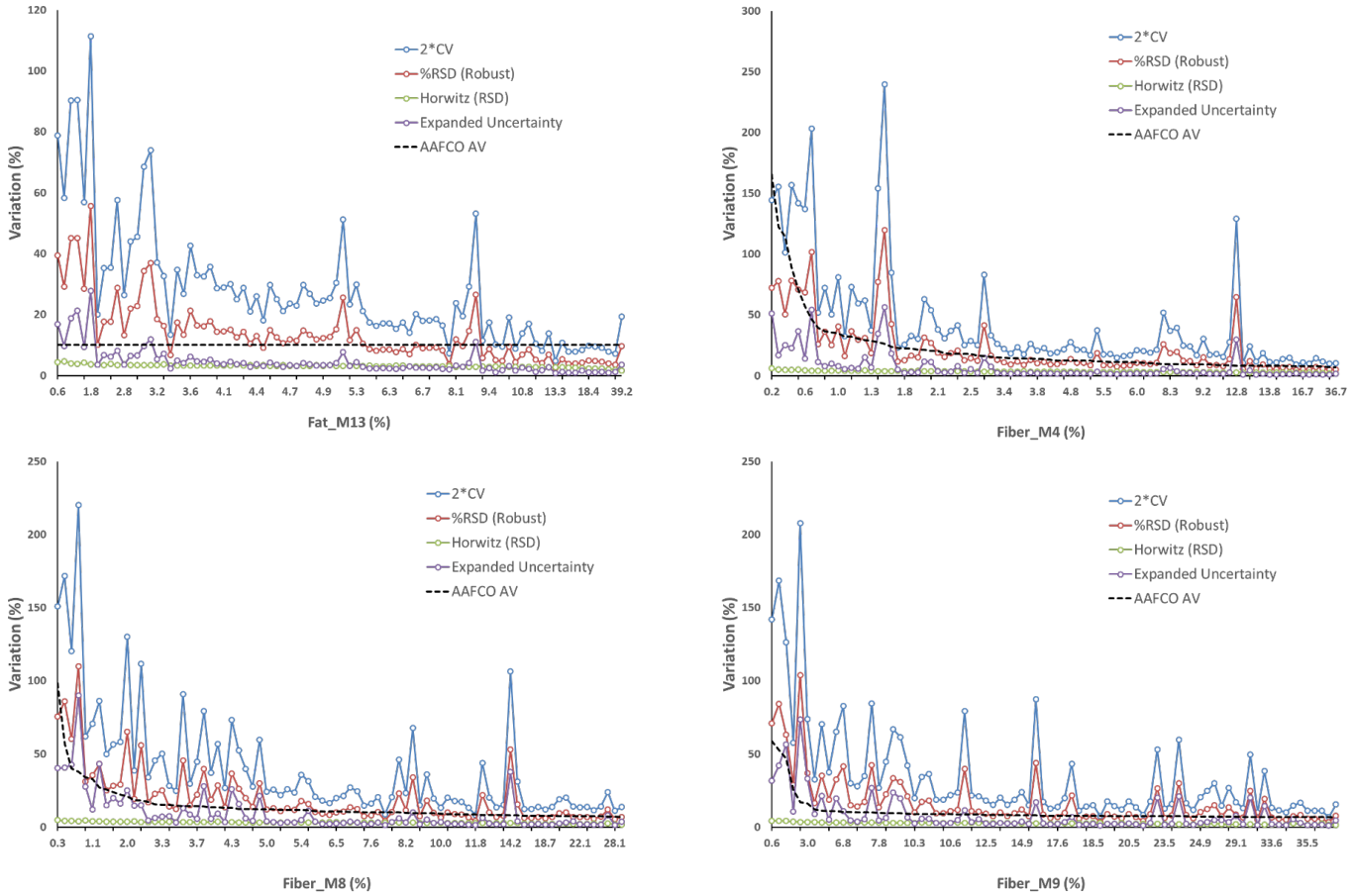
Variation (%)
2*CV
%RSD (Robust)
Horwitz (RSD)
Expanded Uncertainty
AAFCO AV
Fat_M13 (%)
Fiber_M4 (%)
Fiber_M8 (%)
Fiber_M9 (%)

Figure S1(B)

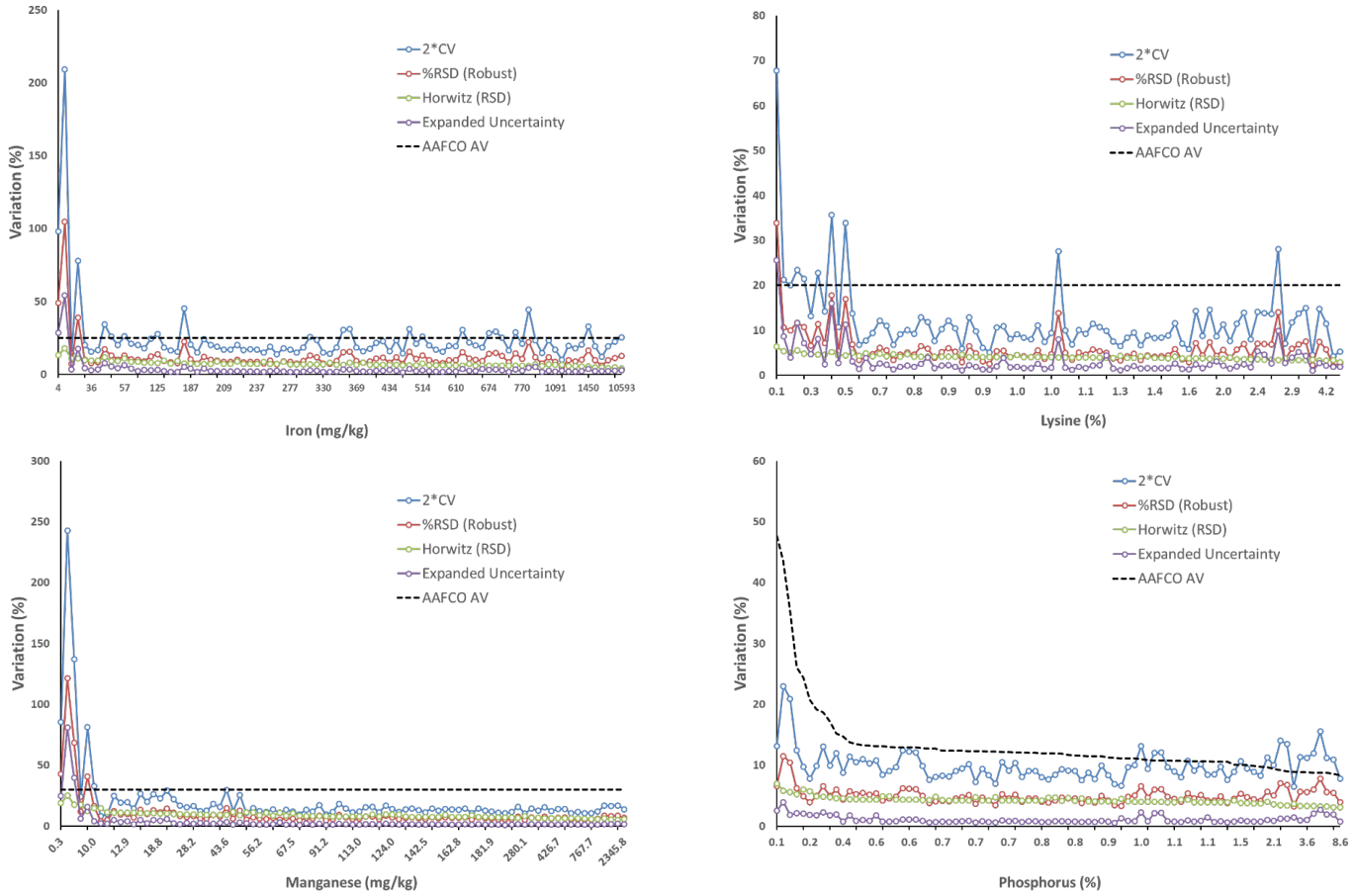

Figure S1(C)

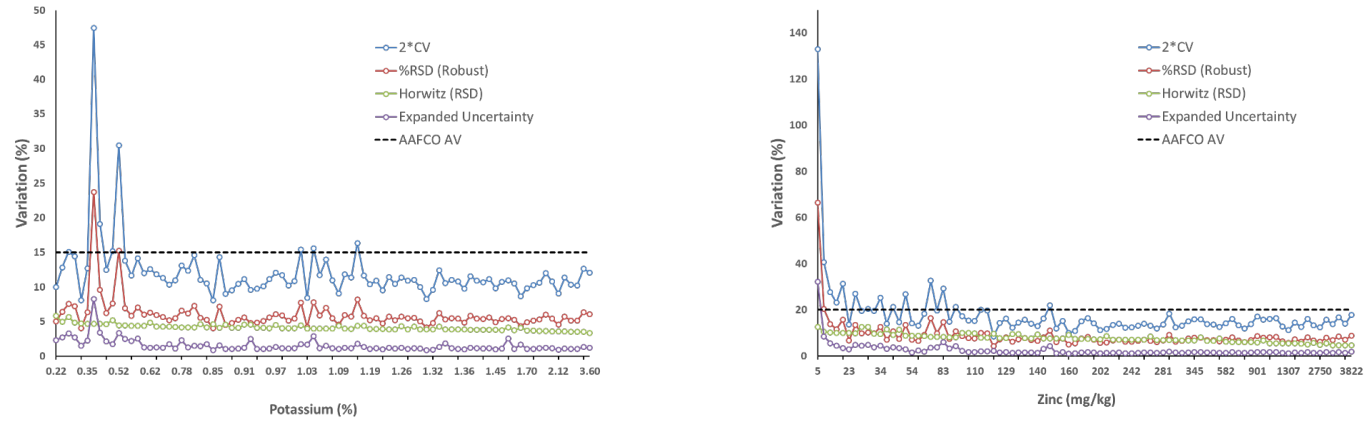

Figure S1 A-D : A scatter plot of assigned values of select analytes (ash, cobalt, phosphorus, calcium, fat (methods 3 and 13), fiber (methods 4, 8, and 9), iron, lysine, manganese, potassium, and zinc) vs relative expanded uncertainties and standard deviations estimated or predicted using the 2014-2018 AAFCO Proficiency Test data: AV, analytical variation (%); RSD, relative standard deviation (%); Horwitz, relative standard deviation (%) predicted by Horwitz equation; and 2*CV, two times coefficient of variation (%).

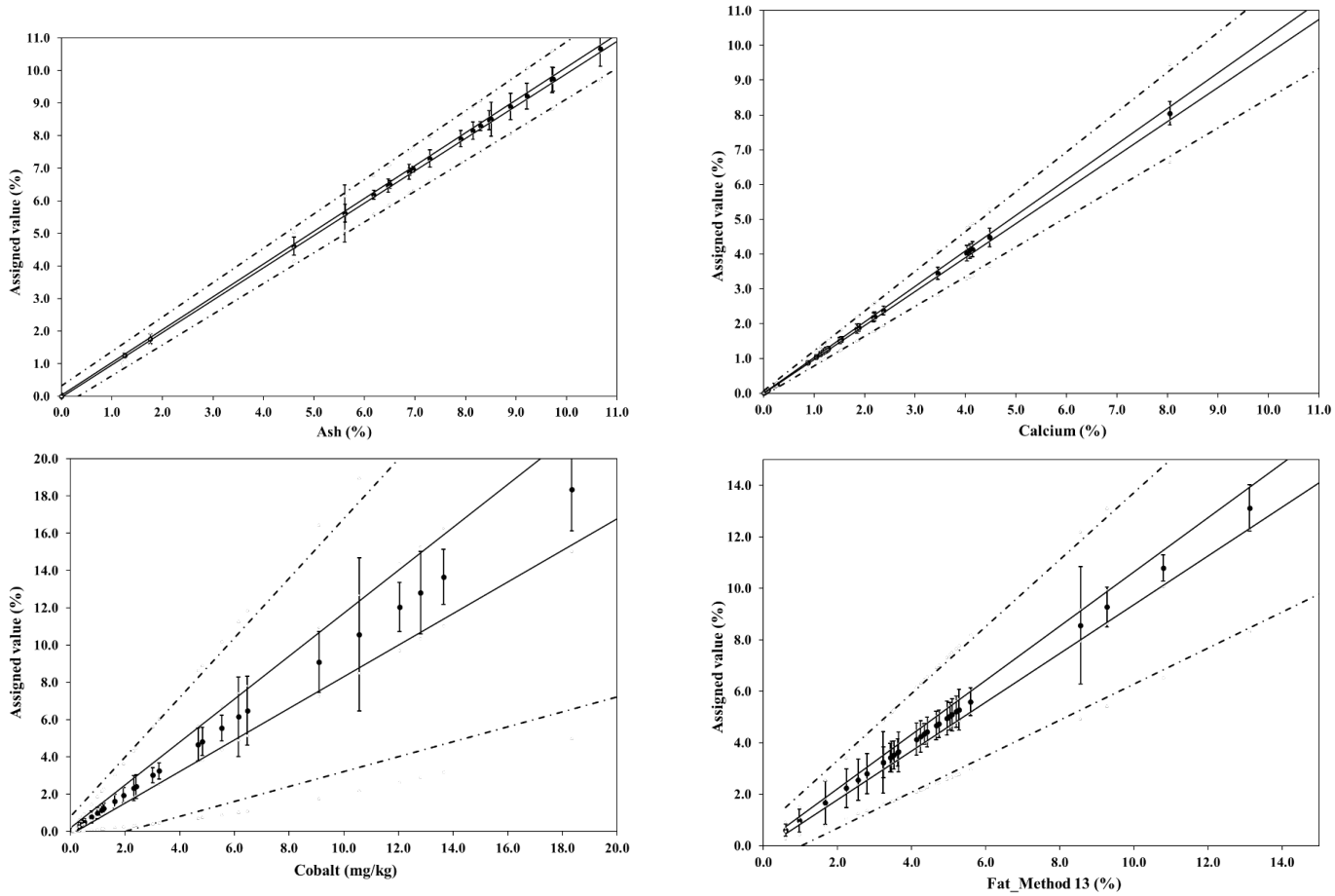
Assigned value (%)
Ash (%)
Assigned value (%)
Calcium (%)
Assigned value (%)
Cobalt (mg/kg)
Assigned value (%)
Fat_Method 13 (%)

Figure S2(A)

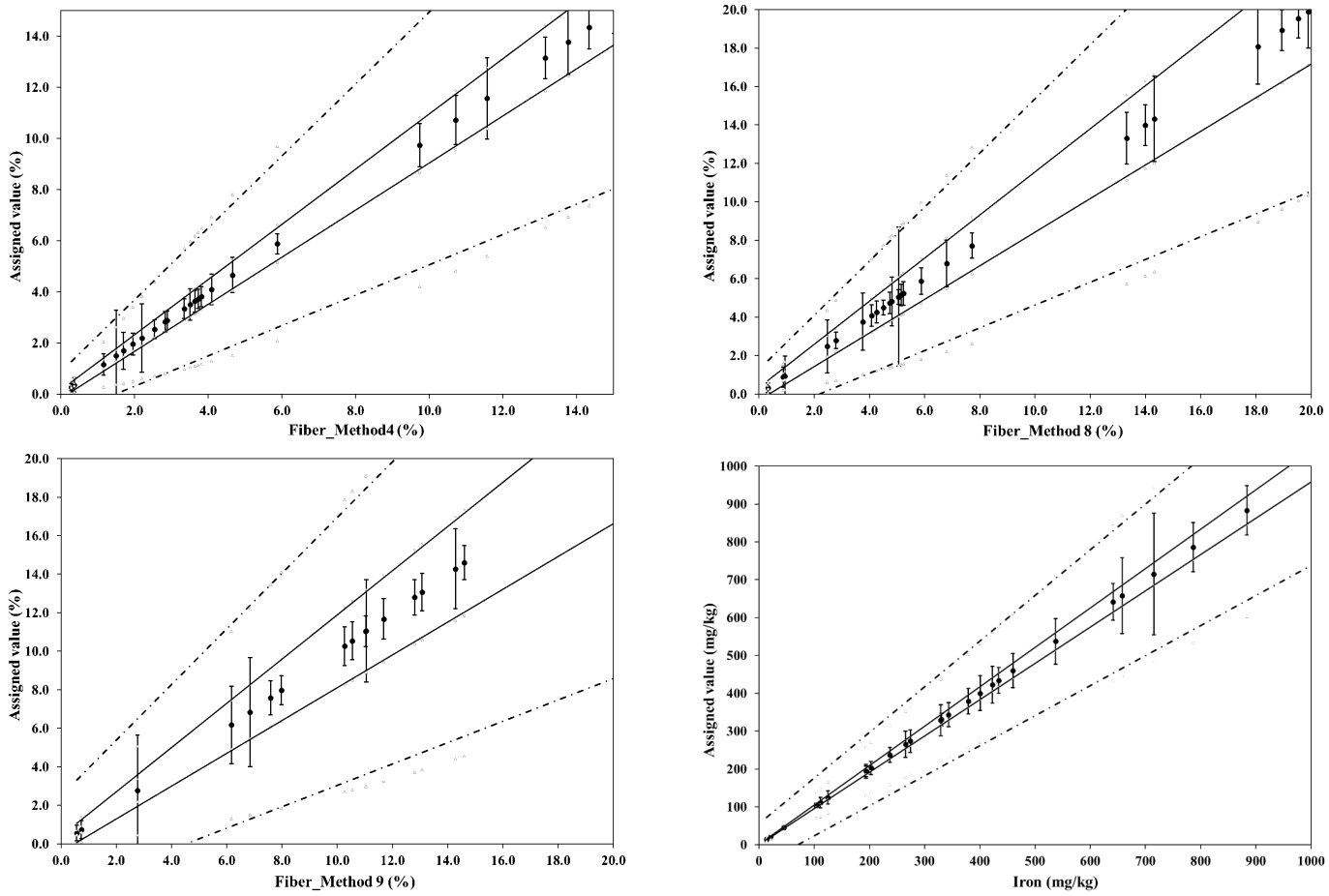
Assigned value (%)
Fiber_Method4 (%)
Assigned value (%)
Fiber_Method 8 (%)
Assigned value (%)
Fiber_Method 9 (%)
Assigned value (mg/kg)
Iron (mg/kg)

Figure S2(B)

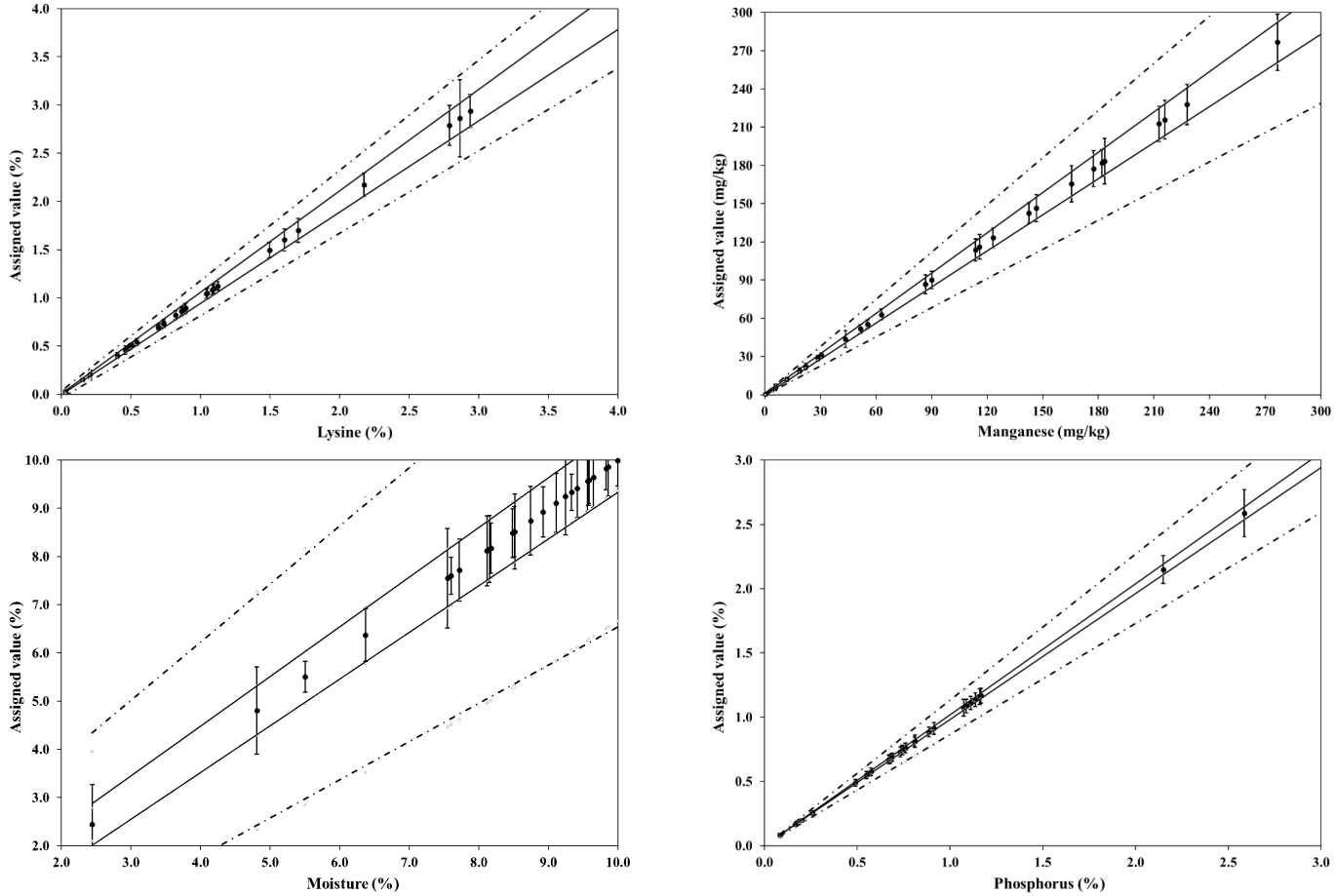

Figure S2(C)

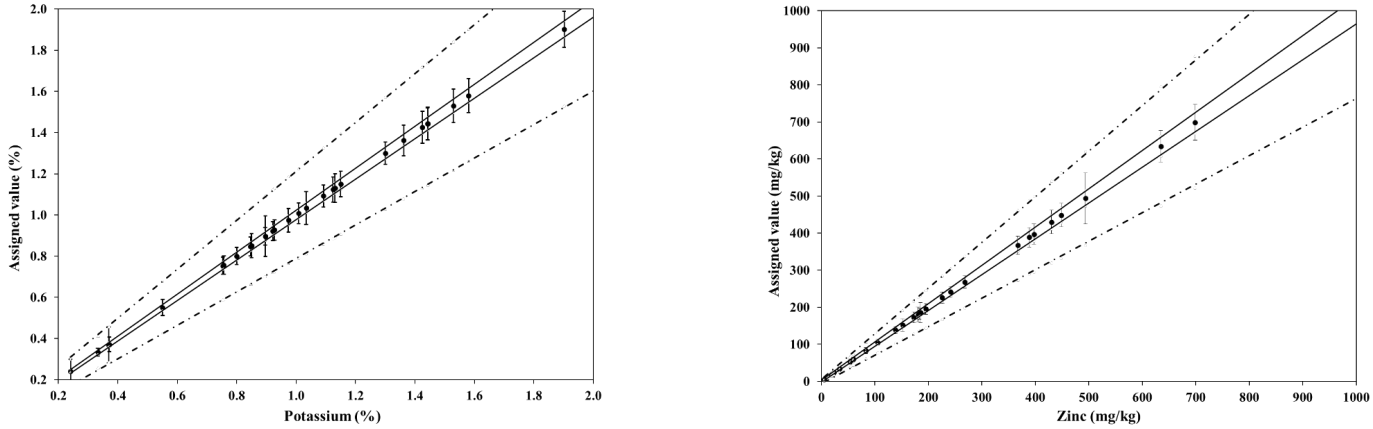

Figure S2 A-D : A scatter plot of the assigned value and its standard deviation obtained from the validation data and a regression line of the expanded uncertainty (*dashed line*) and 2 CV (*solid line*) models for select analytes. Error bars represent standard deviations of the assigned value.