JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Pesticide Residues in Food-Based Proficiency Test Materials, Spiking Values versus Consensus Assigned Values

Mark Sykes,*[,†] Michael Thompson,[‡] and Stewart Reynolds[†]

[†]The Food and Environment Research Agency, Sand Hutton, York YO41 1LZ, United Kingdom

[‡]School of Biological and Chemical Sciences, Birkbeck University of London, Malet Street, London WC1E 7HX, United Kingdom

ABSTRACT: We examine the differences among the three estimates of the true value of the measurand derived from routine proficiency testing of laboratories analyzing foodstuffs for pesticide residues. The three values are (i) the spike level (Sp), (ii) the mean result found by the laboratory conducting the test for sufficient homogeneity (Ho), and (iii) the consensus of the participants' results used as the assigned value (AV) in converting results into z scores. Data amounting to 205 examples were collected from successive rounds of three series of proficiency tests from the Food Analysis Performance Assessment Scheme (FAPAS): namely, series 05 (fats, oils, and animal products), series 09 (cereals and their products), and series 19 (fruits, vegetables, and their products). Irrespective of the class of test material, we found that the means of AV and Ho were almost identical, while the value of Sp was systematically higher than AV by a factor of 1.22. The dispersion of the individual values of both ratios, Ho/AV and Sp/AV, was examined by analysis of variance. A small part of the variance was attributed to the series, but a greater part, about 40%, was attributed to individual rounds within series. We discovered no connection between the ratios and the chemistry of the analyte.

KEYWORDS: *Pesticide residues, proficiency test materials, spiking value, assigned value*

INTRODUCTION

In the Food Analysis Performance Assessment Scheme (FAPAS), the test materials for laboratories undertaking the analysis of pesticide residues are prepared in one of two ways or as a combination of the two. The test material may contain incurred (native) levels of pesticide residues resulting from its agricultural production and/or postharvest treatment. Alternatively, the material may contain little or no incurred pesticide residue and is therefore spiked with pesticide standards. A native analyte seldom provides sufficiently high concentrations of analyte for a proficiency test (PT); therefore, the great majority of test materials are spiked. The material is then "homogenized", split into packaged units for distribution, and tested for sufficient homogeneity. Samples of the homogenized materials are analyzed by the participants and an assigned value for the measurand obtained as a consensus of the reported results.

This sequence of events gives rise to three separate possible estimates of the concentration of each analyte: namely, (i) the spiked level, (ii) the mean value obtained from homogeneity testing, and (iii) the assigned value. The second value is the mean of 20 analytical results produced by a laboratory of a status equivalent to a national reference laboratory. Only the assigned value, however, is used to convert the participants' results into z scores for assessing their performance. A comparison of the three estimates is important to throw light on the process of spiking and to reassure participants that the best possible estimate of the true value of the measurand is used as an assigned value.

Preparation and Homogenization of the Test Materials. FAPAS provides PTs for pesticide residue analysis in three classes of test material: series 05 (fats, oils, and animal products, since 1990), series 09 (cereals and their products, since 1991), and series 19 (fruits, vegetables, and their products, since 1997). Typically, each PT round from one of the three series encompasses one type of commodity spiked with between 5 and 15 pesticides. Occasionally, some pesticides will be already incurred in the material at a sufficient concentration for the PT. More often, the material is overspiked to raise the concentration to an appropriate analytical range.

A typical procedure for the preparation of a test material is cryogenic milling and blending of a bulk quantity of the chosen commodity, resulting in a friable powder or a puree. This material is screened for the presence of any incurred pesticide residues. The remaining bulk material is divided into two portions. The first portion is further mixed for a few hours and subsampled as the blank test material. This blank material may be used by participants for recovery experiments and/or for the preparation of matrix-matched calibration standards. The second portion is spiked with known amounts of pesticide standards and blended over a period of a few hours (for wet or dry vegetable matrices) or overnight (for animal product matrices). In the instance of dry, cereal (flour) test materials, the bulk material is left to stabilize for 4 weeks under cool, dark conditions, before a final blending and subsampling for the PT. Homogeneity testing according to the established procedure[1] is carried out on the subsamples for all of the pesticides (either spiked or incurred). Any spiking regime is a traceable procedure.

Equilibration of Spiked Test Material and Recovery during Analysis. The production and testing of PT materials

Received: January 24, 2013
Revised: April 8, 2013
Accepted: April 15, 2013
Published: April 30, 2013

differs from the procedure applied to validation of analytical methods. In general, validation (recovery experiments) involves the spiking of the blank sample with standard solution, and then extraction/analysis follows effectively immediately.[2,3] Sometimes, recovery measurements from fruit and vegetable matrices take into account an equilibration period, but this varies between 15 min,[4] 1 h,[5] or 30 min.[6] For other matrices, this may be longer, for example, 2 h in beef matrix.[7] The length of this equilibration period can have a significant effect on the degree to which the analyte becomes "bound" to the matrix of the test material (particularly for cereal-based test materials), and therefore, there is an increase in the amount of analyte that is non-solvent-extractable. Often, the spiking protocol reported does not specify at all whether or not there is an equilibration period.[8−12] The collaborative study following method validation reported in ref 2 had a necessarily imposed equilibration period of about 2 days, while the samples were in transit to the participating laboratories.[13]

It is essential that the concentrations of the analytes in the test materials remain stable during transport to the participating laboratories and for the period of time allowed by the PT organizer for the exercise to be completed. PT schemes other than FAPAS have their own protocols for checking analyte stability. The European Union Reference Laboratory (EU RL) scheme[14] carries out level checks during preparation and mixes the bulk material for 12−24 h before cryogenic milling in liquid nitrogen. Subsamples are analyzed before dispatch to participants and at the conclusion of the PT. One PT scheme had the wheat crop grown and sprayed especially to produce incurred residues.[15] Additional spiking was required for some pesticides, after which the test material was left to equilibrate for 4 weeks.

Data. Here, we examine data collected from FAPAS and compare the three estimates of concentration (spiking level, homogeneity mean, and consensus assigned value). The data were compiled from nine rounds of series 05 (between July 2010 and January 2012), 18 rounds of series 09 (between May 2008 and May 2011), and 17 rounds of series 19 (between March 2010 and April 2011). These series can all be regarded as mature schemes, that is, with stabilized properties. All spiking was fully traceable to the international standard (SI) and to pure standards of the analytes with negligible uncertainty. Homogeneity results and participant results were not corrected for recovery, although the recoveries were determined and found to be within the acceptable range (70−120%).[16]

RESULTS AND DATA ANALYSIS

The assigned value was used as the reference value for two-way comparisons to homogeneity means and spiked values. The corresponding ratios were used for statistical appraisal. Figure 1 shows a dot plot of the two ratios. The homogeneity mean/assigned value ratio (Ho/AV) was close to symmetrical and centered close to 1.000, with a median of 0.971. The spiked value/assigned value ratio (Sp/AV) was clearly centered at a higher level than 1.000 (median = 1.226) and positively skewed. The burden of the statistical analysis was therefore focused on discovering, if possible, the reason behind the discrepancy in Sp/AV by examining factors that might have influenced this ratio.

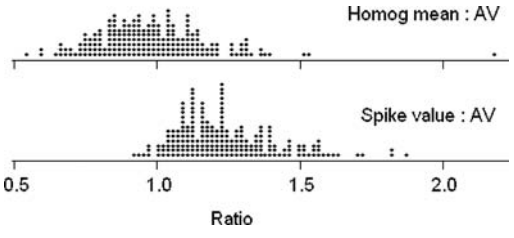

Figure 1. Dotplots of the ratios Ho/AV and Sp/AV (all results).

A scatter plot of ratio Ho/AV against AV (Figure 2) shows no obvious trend in the mean value, in either the combined data or within separate series. There was no apparent heteroscedasticity (higher dispersion of ratios with higher concentration). A plot of Sp/AV against the assigned values (Figure 3) showed a very similar outcome, except that the ratio was systematically higher than 1.000. The apparently greater dispersion among the ratios for series 09 is brought about by the skew. The discrepancy in Sp/AV seems to be independent of the concentration of the analyte.

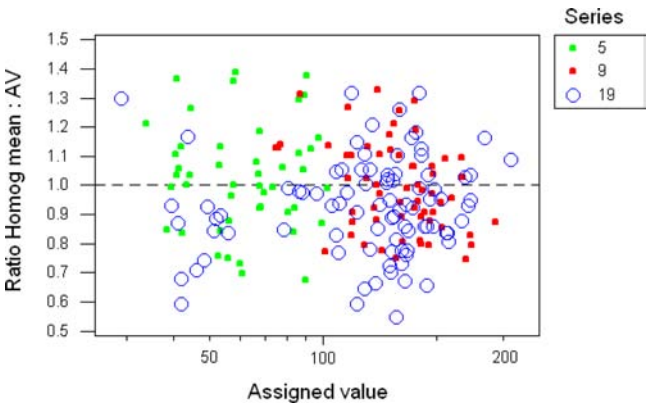

Figure 2. Scatterplot of the Ho/AV ratio versus assigned value (μg kg^{-1}) showing series 05 values (green solid circles), series 09 values (red solid circles), and series 19 values (open blue circles) (note that the logarithmic x axis is used simply to display a more uniform scatter of points).

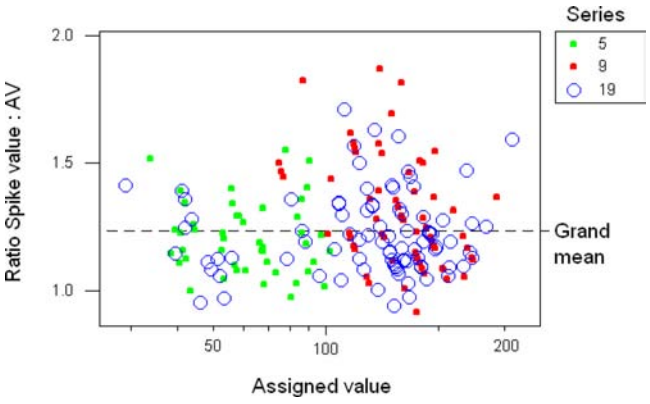

Figure 3. Scatterplot of the Sp/AV ratio versus assigned value (μg kg^{-1}) showing series 05 values (green solid circles), series 09 values (red solid circles), and series 19 values (open blue circles) (note that the logarithmic x axis is used simply to display a more uniform scatter of points).

The ratios, organized by series and by round within series, are shown in boxplots (Figures 4 and 5). There are no obvious time trends within series (round numbers are plotted sequentially by date). Small trends were found by regression

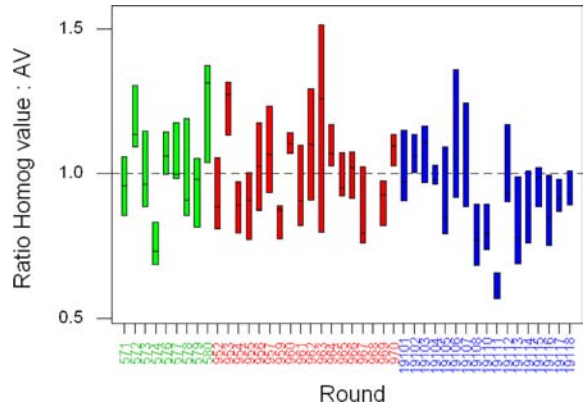

Figure 4. Boxplot of all Ho/AV ratios organized by series and round.

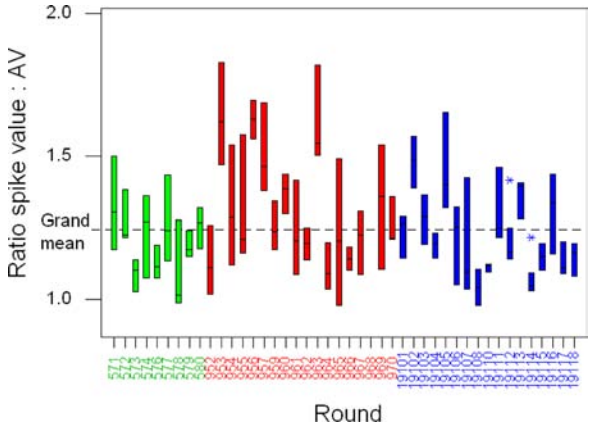

Figure 5. Boxplot of all Sp/AV ratios organized by series and round.

but accounted only for a trivial proportion of the total variance. Some of these trends were significant but only by virtue of the large number of degrees of freedom. Analysis of variance (ANOVA) between series showed small variations among series means, as shown in Table 1, all significant by virtue of the large number of degrees of freedom.

Table 1. Results of ANOVA by Series for Ho/AV and Sp/AV Ratios

ratio	series 05	series 09	series 19
Ho/AV	1.028	1.002	0.931
Sp/AV	1.208	1.305	1.220

A nested ANOVA (rounds within series) of the two sets of ratios was performed to show the standard deviations of the separate contributions (Table 2). The outcome is very similar

Table 2. Component Standard Deviations from Nested ANOVA of Ho/AV and Sp/AV Ratios

source of variation	Ho/AV (standard deviation of component)	Sp/AV (standard deviation of component)
between series	0.039	0.037
between rounds (within series)	0.107	0.117
within rounds (between observations)	0.142	0.133
total	0.182	0.181

for the two sets of ratios. The differences among the means of the series makes a relatively small contribution (about 4%) of the total variance. The variation among rounds within a series makes a more substantial contribution of about 40% of the total variance. The remainder (about 56%) probably represents the unique behavior of individual analytes.

To assess this last idea, Figures 6 and 7 show boxplots for individual analytes in instances where three or more values of

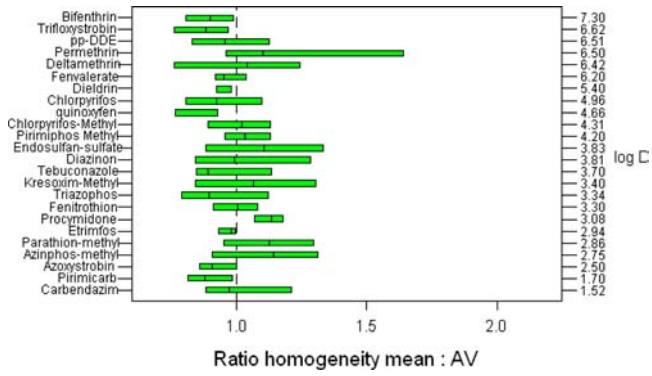

Figure 6. Boxplot of Ho/AV ratios for analytes with three or more values, over all series, organized by analyte in order of log *D* value.

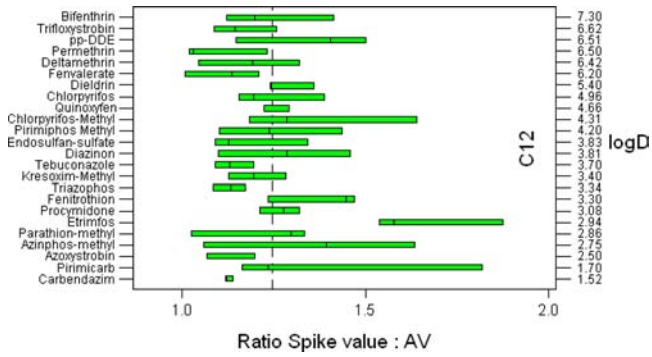

Figure 7. Boxplot of Sp/AV ratios for analytes with three or more values, over all series, organized by analyte in order of log *D* value, with the grand mean (dashed line).

the ratios were available, regardless of the series from which the result originates. Combining results of the three series seems justified because of the relatively small contribution to the variance. In the event, no significant variation between analytes could be detected by one-way ANOVA.

Finally, it was thought possible that the differences in behavior among the analytes might be related to an index of their chemical natures. The logarithm of the distribution coefficient of the analyte between octan-1-ol and water (log *D*) seemed to be a reasonable proxy for chemical processes that affect recovery in the analysis of pesticides. It acts as an indicator of the relative positions of the analytes along the hydrophilic/hydrophobic spectrum and, therefore, possibly the strength of chemisorption in food substrates. In Figures 6 and 7, the analytes are shown organized in decreasing order of log *D* measured at pH 7. However, no trend in the means was visible or detected by regression of the individual ratios. In short, there was no perceptible effect related to log *D*.

dx.doi.org/10.1021/jf4003363 | J. Agric. Food Chem. 2013, 61, 4205−4209

DISCUSSION

There is systematically good agreement between the assigned value and the homogeneity mean. The mean ratio Ho/AV is close to unity, and there seems to be no concentration-related change in the ratio or its variance. This implies that, on average, either both are correct or that both suffer from a common problem. However, the variance of the ratio has been shown to comprise components derived from the specific series and specific rounds within a series, as well as variance within a round. Series-specific variance suggests differences caused by the nature of the test material, bearing in mind that the matrices are quite different in chemical composition and unequally demanding to analyze accurately. Round-specific variance might reflect the effects of different matrices within the series specification or simply the fact that the identity of the analytes present tends to be different in successive rounds; there might be variation in the difficulty of determining them. However, no effect related to the distribution coefficients of individual analytes was found.

The spiking level, however, is systematically higher than the assigned value, with a median ratio Sp/AV of 1.226. Broadly speaking, the variance of the ratio can be separated into components very similar to those found for the Ho/AV ratio and with the same causes. The systematic component, however, suggests a single general effect beyond those found by ANOVA, that is, regardless of test material or analyte. *A priori*, we see that causes could include (a) impure spikes, (b) overestimated recoveries, or (c) loss of the spiked analytes during the preparation, storage, or distribution of the test materials. Loss of analyte could be brought about actually by chemical reaction, such as photolysis, oxidation or hydrolysis, and volatilization, or effectively by irreversible adsorption onto the matrix of the test material or separation media.

The purity of the analytes used to spike the test materials cannot account for the size of the discrepancy observed, that is, about 18% of the spiked value. The reference standards used all come from reputable suppliers with certificates of analysis, indicating that purity is in excess of 98%.

It is standard practice for pesticide residues to be reported without correction for recovery,[16] provided that the batch recovery is in the range of 70−120%. FAPAS requires that participants in pesticide PTs also report recovery values when submitting their results. The mean recovery reported by the laboratories testing for homogeneity is 92%. This is, of course, a marginal recovery based on spiking without any details of how the blank test material was spiked and the amount of time elapsed between spiking and extraction. Given the closeness of the mean ratio Ho/AV to unity, it seems that a closely similar value would be representative of the participants. Correction for a recovery of that magnitude would bring the mean of the ratio Sp/AV up to only about 1.09, insufficient to account for the observed discrepancy. Marginal recovery is always likely to overestimate the recovery of native analyte, but in FAPAS, nearly all of the analyte is spiked rather than native. Even so, there is undoubtedly more time for the spiked analyte to become irreversibly adsorbed during the preparation and storage of test materials than during participants' analyses. In that case, overestimation of recovery cannot be ruled out as a contributing factor in the observed discrepancy between spiking and consensus assigned values. However, it is conceptually indistinguishable from the loss of the spike during test material preparation.

The preparation of test materials for FAPAS pesticide in cereals PTs has always included a stabilization period, following spiking but before homogeneity testing is carried out. Analytes can bind to the matrix and become "non-solvent extractable". This tendency has been known for many years and is particularly strong in low-moisture-containing commodities, such as cereals and dried fruits.[11,12,16−20] It is noteworthy that, in the present study, the mean ratio Sp/AV (and its variance) is significantly greater in series 09 results, where the test materials are cereals and cereal products. Extraction efficiency is dependent upon many factors; the physicochemical properties of the analyte, polarity of the solvent, pH, temperature, particle size of the test material (surface area), length of time of shaking, and adsorption onto active surfaces of glassware or plastics can all have an influence. This is acknowledged in reports of other PT schemes,[14,15] although in the present study, no correlation was found between Sp/AV ratios and a proxy for the molecular polarity of the analyte. Even where a medium is used to afford some stability to the sample prior to analysis, such as a solid-phase extraction disk,[21] losses can still occur by volatilization or hydrolysis from residual water. In the instances of method validation and quality control, official guidance[16] does not specify the timeliness of spiking and subsequent analysis. Any losses, therefore, will not usually be noticed in a laboratory's routine work. The official guidance document does recommend the addition of water to samples of low-moisture content prior to extraction. No actual time between water addition and extraction is specified, however, and for some analytes, significant losses as a result of enzyme hydrolysis may occur as this period of time extends.

A final inference may be important. If, as assumed, the discrepancies between the estimates of true value are due to the loss or sequestration of analyte, it must all occur at the preparation stage of the test material. Despite the time interval between homogeneity testing and analysis by the participants, there is virtually no systematic difference between the two results.

The motivation behind this study was to address the discrepancies, expressed here as the ratio Sp/AV between the spiked values and the corresponding assigned values (the consensus of participants' results) as used in FAPAS. The dispersion of this ratio was found to contain contributions from the nature of the test material and possibly the identity of the analyte. The systematic part of the discrepancy was attributed to differences in recovery between analyte spiked at the time of preparation of the test materials and analyte spiked by a participant during their analysis. Given the large number of analytes involved, this conjecture could not be confirmed by any practicable experiment. Moreover, the recovery of naturally incurred (native) analyte in routine samples is essentially unknowable given currently available technology; that is one of the main reasons why pesticide residue analysis is effectively treated as empirical (method-dependent), and therefore, results are not corrected for recovery. It is conceivable that analytical methods involving more rigorous extraction procedures could recover more of the analytes than represented here. However, such methods would necessarily be costly to execute and *ipso facto* not fit for the purpose in routine analysis. Given these limitations, it is clear that the participants' consensus is the appropriate assigned value for FAPAS proficiency testing.

AUTHOR INFORMATION

Corresponding Author

*Telephone: +44-0-1904-462697. Fax: +44-0-1904-500440. E-mail: mark.sykes@fera.gsi.gov.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Round co-ordinators responsible for the preparation and validation of the test materials and compilation of the participants' results were Simon Hunter, Michael Knaggs, and Laura Prenton.

ABBREVIATIONS USED

FAPAS, Food Analysis Performance Assessment Scheme; Sp, spike; Ho, homogeneity; AV, assigned value; PT, proficiency test; EU RL, European Union Reference Laboratory; SI, international standard; AVONA, analysis of variance

REFERENCES

(1) Fearn, T.; Thompson, M. A new test for "sufficient homogeneity". *Analyst* **2001**, *126*, 1414−1417.

(2) Lehotay, S. J.; de Kok, A.; Hiemstra, M.; van Bodegraven, P. Validation of a fast and easy method for the determination of residues from 229 pesticides in fruits and vegetables using gas and liquid chromatography and mass spectrometric detection. *J. AOAC Int.* **2005**, *88*, 595−614.

(3) Wong, J.; Hao, C. Y.; Zhang, K.; Yang, P.; Banerjee, K.; Hayward, D.; Iftakhar, I.; Schreiber, A.; Tech, K.; Sack, C.; Smoker, M.; Chen, X. R.; Utture, S. C.; Oulkar, D. P. Development and interlaboratory validation of a QuEChERS-based liquid chromatography−tandem mass spectrometry method for multiresidue pesticide analysis. *J. Agri. Food Chem.* **2010**, *58*, 5897−5903.

(4) Lehotay, S. J.; Son, K. A.; Kwon, H.; Koesukwiwat, U.; Fu, W. S.; Mastovska, K.; Hoh, E.; Leepipatpiboon, N. Comparison of QuEChERS sample preparation methods for the analysis of pesticide residues in fruits and vegetables. *J. Chromatogr., A* **2010**, *1217*, 2548−2560.

(5) Bolaños, P. P.; Moreno, J. L. F.; Shtereva, D. D.; Frenich, A. G.; Vidal, J. L. M. Development and validation of a multiresidue method for the analysis of 151 pesticide residues in strawberry by gas chromatography coupled to a triple quadrupole mass analyzer. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2282−2294.

(6) Jiang, Y. P.; Li, Y. J.; Jiang, Y. T.; Li, J. G.; Pan, C. P. Determination of multiresidues in rapeseed, rapeseed oil, and rapeseed meal by acetonitrile extraction, low-temperature cleanup, and detection by liquid chromatography with tandem mass spectrometry. *J. Agric. Food Chem.* **2012**, *60*, 5089−5098.

(7) Wu, G.; Bao, X. X.; Zhao, S. H.; Wu, J. J.; Han, A. L.; Ye, Q. F. Analysis of multi-pesticide residues in the foods of animal origin by GC−MS coupled with accelerated solvent extraction and gel permeation chromatography cleanup. *Food Chem.* **2011**, *126*, 646−654.

(8) Kim, B.; Ahn, S.; Mitani, Y. Interlaboratory comparison for the determination of pesticide residues in Chinese cabbage. *Accredit. Qual. Assur.* **2011**, *16*, 499−505.

(9) Pizzutti, I. R.; de Kok, A.; Zanella, R.; Adaime, M. B.; Hiemstra, M.; Wickert, C.; Prestes, O. D. Method validation for the analysis of 169 pesticides in soya grain, without clean up, by liquid chromatography−tandem mass spectrometry using positive and negative electrospray ionization. *J. Chromatogr., A* **2007**, *1142*, 123−136.

(10) Sack, C.; Smoker, M.; Chamkasem, N.; Thompson, R.; Satterfield, G.; Masse, C.; Mercer, G.; Neuhaus, B.; Cassias, I.; Chang, E.; Lin, Y.; MacMahon, S.; Wong, J.; Zhang, K.; Smith, R. E. Collaborative validation of the QuEChERS procedure for the determination of pesticides in food by LC−MS/MS. *J. Agric. Food Chem.* **2011**, *59*, 6383−6411.

(11) Walorczyk, S. Development of a multi-residue screening method for the determination of pesticides in cereals and dry animal feed using gas chromatography−triple quadrupole tandem mass spectrometry. *J. Chromatogr., A* **2007**, *1165*, 200−212.

(12) Walorczyk, S. Development of a multi-residue method for the determination of pesticides in cereals and dry animal feed using gas chromatography−tandem quadrupole mass spectrometry II. Improvement and extension to new analytes. *J. Chromatogr., A* **2008**, *1208*, 202−214.

(13) Lehotay, S. J. Determination of pesticide residues in foods by acetonitrile extraction and partitioning with magnesium sulfate: Collaborative study. *J. AOAC Int.* **2007**, *90*, 485−520.

(14) Medina-Pastor, P.; Rodriguez-Torreblanca, C.; Andersson, A.; Fernandez-Alba, A. R. European Commission proficiency tests for pesticide residues in fruits and vegetables. *TrAC, Trends Anal. Chem.* **2010**, *29*, 70−83.

(15) Poulsen, M. E.; Christensen, H. B.; Herrmann, S. S. Proficiency test on incurred and spiked pesticide residues in cereals. *Accredit. Qual. Assur.* **2009**, *14*, 477−485.

(16) European Union (EU). *Method Validation and Quality Control Procedures for Pesticide Residues Analysis in Food and Feed*; EU: Brussels, Belgium, 2011; SANCO/12495/2011.

(17) QuEChERS. http://www.quechers.cvua-stuttgart.de (accessed July 31, 2012).

(18) Matthews, W. A. An investigation of the non-solvent extractable residues of [^{14}C]chlorpyrifos-methyl in stored wheat. *Pestic. Sci.* **1991**, *31*, 141−149.

(19) Skidmore, M. W.; Paulson, G. D.; Kuiper, H. A.; Ohlin, B.; Reynolds, S. Bound xenobiotic residues in food commodities of plant and animal origin. *Pure Appl. Chem.* **1998**, *70*, 1423−1447.

(20) Kimihiko, Y.; Yasuhide, T.; Hitoshi, U.; Katsuhhiko, N. Malathion residue in wheat kernels is degraded by thion organophosphorus pesticide-specific carboxylesterase. *J. Health Sci.* **2006**, *52*, 221−227.

(21) Cobb, J. M.; Mattice, J. D.; Senseman, S. A.; Dumas, J. A.; Mersie, W.; Riley, M. B.; Potter, T. L.; Mueller, T. C.; Watson, E. B. Stability of pesticides on solid-phase extraction disks after incubation at various temperatures and for various time intervals: Interlaboratory study. *J. AOAC Int.* **2006**, *89*, 903−912.