Proficiency test results from 5 countries involving 61 separate interlaboratory proficiency tests for pesticide residues were examined in this study. A total of 24 different matrixes and 869 relative standard deviations of the mean (or median) pesticide residue concentration were statistically evaluated in relation to the Horwitz function. The aim was to determine whether or not the concentration-dependent relationship described by Horwitz would hold for the much narrower range of chemicals and concentrations covered in routine pesticide residue analysis. Although for fatty (animal-derived) matrixes the variability increased as the concentration decreased in line with the Horwitz equation, the between-laboratories relative standard deviations for nonfatty matrixes (fruit, vegetables, and grain) remained at 25% over the entire concentration range of 1 mg/kg to 10 mg/kg for the pesticides studied. Given these findings, the Horwitz equation remains valid for calculating uncertainties involving pesticide residues in fatty matrixes. However, for pesticide residue analyses involving nonfatty matrixes, a constant relative standard deviation of 25% is more appropriate for calculating uncertainties, particularly when a reported result is assessed against a regulatory limit.

The fact that inherent uncertainties are associated with analytical results generated in quantitative analysis is one that few analytical chemists would argue. It is well known that replicate analyses of the same sample, analyzed either within- or between-laboratories, will generate different results. This is particularly true for the determination of chemical residues at trace (0.001–10 mg/kg) levels. It is not uncommon to have variations in analytical results of >50% at the lower (0.001 mg/kg) end of the range.

Consequently, knowledge of the uncertainty associated with any analytical result is essential. Responsible decisions regarding whether or not the reported result exceeds a regulatory limit can be made only with the full knowledge of the uncertainties associated with the result.

Furthermore, to avoid barriers in international trade, estimated uncertainties are also essential components for demonstrating the equivalence of analytical results generated by exporting and importing countries and compliance with international standards and guidelines such as EN 45001 or ISO Guide 17025.

The estimation of uncertainty has always been a necessary part of method validation. Often estimates were derived from repeated analyses of certified matrix reference materials or spiked matrix blanks. This approach has become known as the “top-down” or “holistic” approach to estimating measurement uncertainties in analytical chemistry. However, alternative approaches based on metrology are being discussed and advocated at an ever-increasing number of national and international conferences and meetings. The metrological approach, which is gaining acceptance in some circles, is also known as the “bottom-up,” “uncertainty budget,” “deconstructive,” and “component-by-component” approach and is based on the principles developed for estimating uncertainties in physical measurements.

However, a number of peculiarities are associated with the chemical analysis of pesticide residues that are different from those encountered in physical measurements. Although results of pesticide residue analysis are influenced by the uncertainties of the instrumentation used, they are mainly affected by a variety of other more important factors including the following: the stability of the analyte during the preparation of the analytical sample from the laboratory sample (e.g., decomposition of dithiocarbamates or carbosulfan); inappropriate/inefficient extracting solvents (e.g., acetone for acephate...
or methamidophos); influence of sample acidity on extraction efficiency of basic or acid pesticides (e.g., extraction of carbendazim from acidic apples); poor recoveries of some individual compounds in cleanup steps (e.g., very early elution of pyrethroids in gel permeation chromatography [GPC]); losses of polar compounds by adsorption onto glass surfaces (e.g., glyphosate or chlormequat); losses of volatile analytes during solvent removal (e.g., loss of HCB during rotary evaporation); analyte adsorption or decomposition in the gas chromatographic (GC) injection port (e.g., DDT); analyte adsorption or decomposition in the analytical column (e.g., loss of iprodione in GC); and matrix effects on GC response, i.e., many polar pesticides in extracts of plant materials exhibit an increased response due to the presence of the matrix (e.g., organophosphorus pesticides).

The magnitude of these specific effects in pesticide residue analysis is influenced by a number of factors. These include the purity of the reagents, the cleanliness and type of glassware, the method or instrumentation used, and the technical skills of the analyst. There is no a priori way of determining if they contribute to bias or precision. Their contribution must be determined experimentally and undoubtedly will consist of a confounded mixture of precision and bias interactions.

Although some of these parameters can be minimized, for example, by the use of standard or official methods, it is virtually impossible to reduce all influences to a negligible amount if multiresidue techniques are involved. Invariably, when multiresidue techniques are used, a compromise regarding the optimization of recoveries and precision will have to be made, at least for some analytes, in order that all compounds remain part of the list of analytes covered by the analytical regimen.

Some practical problems and difficulties are associated with the determination of individual uncertainties in each laboratory. For example, the precision and bias determined as a result of one series of measurements can provide only a snapshot of the likely uncertainties obtained in the same laboratory at a different time.

Factors that may influence the uncertainties of a separate sequence measurements include the following: the nature of samples analyzed before the sequence to determine uncertainties, because these samples may influence the stability of the analyte in the injector or change its rate of transmission through the analytical system (injector, column, detector), and the difficulty of keeping all components of a multiresidue method constant over several months, particularly when samples with different matrixes are analyzed (e.g., additional

Figure 1. Plot of standard deviation, $s_R$, versus mean analyte concentration for 869 proficiency test data (logarithmic scale). The dashed line represents the Horwitz function, and the solid line represents the calculated regression derived from proficiency test results.
cleanup steps may be required for particular samples, GC column type/brand may be changed, the temperature program may vary, and the integrity/inertness of the GC detector and/or column may deteriorate).

The aim of this study was to examine the interlaboratory proficiency test results from a number of countries in order to determine whether or not the concentration-dependent relationship described by Horwitz et al. (1) also holds for the narrower range of concentrations covered in routine pesticide residue analysis.

Determination of Analytical Uncertainty

The Bottom-Up Concept

The basic process of the “bottom-up” concept (also called “uncertainty budget” or “component-by-component approach”) as described in the EURACHEM Guide (2) involves identifying all possible sources of random and systematic variability, assigning a (relative) standard uncertainty to each parameter in the form of a standard deviation, calculating the square root of the sum of squares of individual standard deviations (uncertainty propagation), taking covariance into consideration, and estimating uncertainties due to chemical sources (“Type B” effects) by professional judgment based on experience. This approach has been successfully used in physics and engineering and may have application in physical chemistry.

Advantages of this approach include the fact that because some uncertainties can be estimated on the basis of method validation data and analytical experience, additional experiments are seldom necessary. This approach saves time and resources and is consequently more cost effective and realistic than interlaboratory comparison studies. Furthermore, it provides estimates of uncertainty that are more realistic than those derived from within-laboratory repeatability studies.

Disadvantages of the bottom-up concept, in addition to the problem related to the frequent unavailability of complete validation data of “old” methods, include the fact that such quantitative models of measurement uncertainty are often unable to include all “chemical sources” of variability. It is not uncommon to find that the largest contributions to uncertainties arise from Type B effects. Among these are the least predictable effects such as fluctuations of analyte stability caused by different amounts of relevant matrix components and by impurities in different lots of chemicals used for extraction and cleanup; changes in the break-through volumes in adsorption chromatographic cleanup steps, which are influenced by unknown coextractives or unknown amounts of known coextractives; the actual inertness of the GC injector; “matrix effects,” which may be observed as increased or decreased responses, compared with those produced by simple solvent solutions of the

\[
\text{fatty (x): } \log_{10} s_R = 0.856 \log_{10} c - 1.658
\]
\[
\text{non-fatty (○): } \log_{10} s_R = 1.025 \log_{10} c - 0.457
\]

Figure 2. Plot of standard deviation, \(s_R\), versus mean analyte concentration of nonfatty (solid line) and fatty (dashed line) material subsets of proficiency test results.

ALDER ET AL.: JOURNAL OF AOAC INTERNATIONAL Vol. 84, No. 5, 2001 1571
analyte; and the well-known, but unmanageable, fluctuations in the sensitivity of a GC nitrogen–phosphorus detector.

Often the influence of such variability on the analytical precision is small or negligible within an individual laboratory over a period of some days or weeks and cannot be observed when an uncertainty budget is determined by a single laboratory. This is consistent with the conclusion, described by Horwitz (3), that “this approach is likely to (1) overlook important variables and double count others, (2) avoid considering unknown and unknowable interactions and interference, and (3) adjust missing variables with an uncontrollable Type B component.” Evidently, physical and chemical measurements often have entirely different error patterns that behave differently on replication, especially under reproducibility conditions (4). Consequently, precision parameters determined in a specific laboratory and used for the calculation of uncertainty within the uncertainty budget concept may vary significantly with a new analyst or with time.

Additional problems encountered in the calculation of uncertainty budgets relate to the several thousand relevant pesticide/crop combinations encountered, e.g., those identified by the Codex Committee on Pesticide Residues (CCPR), as well as the dozens of analytical methods used in pesticide residue analysis. The large number of combinations makes the individual estimation of all uncertainty budgets impossible. On the other hand, accepted rules for the grouping of pesticides with similar physical and, in particular, chemical properties do not exist. Finally, to identify systematic errors in an individual laboratory, traceability to a stated reference value is necessary. However, of the thousands of pesticide/crop combinations analyzed, probably <5% are available as standard reference materials. Consequently, in the absence of a suitable reference material, method and/or laboratory bias cannot be determined.

The Top-Down (Holistic, etc.) Concept of the Analytical Methods Committee (5)

In the top-down concept, which is now addressed in the revised draft of the EURACHEM Guide (6), interlaboratory method performance tests and their between-laboratories data are the starting points for estimating measurement uncertainty. Many relevant sources of uncertainty are reflected in the between-laboratories standard deviations of such tests. This is different from the bottom-up concept, in which the total uncertainty is estimated by starting at the lowest level (i.e., with one aspect of the method), then considering all aspects of the method, and finally assessing the influence (bias) of a laboratory.

In the top-down model, the deviation between the mean or median of the interlaboratory method performance test (assigned value) and the analytical result of 1 laboratory may be caused by the following: the systematic error of the laboratory (laboratory bias), the error of the method used (method bias), the systematic error within a run (run bias), and the random measurement error.

Consequently, many error components within a laboratory are systematic. However, viewed from the higher level of a method performance test, these total biases often behave as random variables with small deviations occurring more often.

<table>
<thead>
<tr>
<th>Category</th>
<th>Data used</th>
<th>Matrix type</th>
<th>Pesticide type</th>
<th>No. of observations (n)</th>
<th>Slope (a)</th>
<th>Bias (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>869</td>
<td>0.937</td>
<td>-1.072</td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>Fatty</td>
<td>All</td>
<td>414</td>
<td>0.855 ± 0.020</td>
<td>-1.660 ± 0.135</td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>Nonfatty</td>
<td>All</td>
<td>455</td>
<td>1.025 ± 0.022</td>
<td>-0.457 ± 0.148</td>
</tr>
<tr>
<td>4</td>
<td>All</td>
<td>Fatty</td>
<td>Organochlorines</td>
<td>218</td>
<td>0.864 ± 0.023</td>
<td>-1.581 ± 0.159</td>
</tr>
<tr>
<td>5</td>
<td>All</td>
<td>Nonfatty</td>
<td>Organochlorines</td>
<td>83</td>
<td>0.962 ± 0.042</td>
<td>-0.921 ± 0.293</td>
</tr>
<tr>
<td>6</td>
<td>All</td>
<td>Fatty</td>
<td>Other pesticides</td>
<td>196</td>
<td>0.879 ± 0.038</td>
<td>-1.528 ± 0.254</td>
</tr>
<tr>
<td>7</td>
<td>All</td>
<td>Nonfatty</td>
<td>Other pesticides</td>
<td>372</td>
<td>1.028 ± 0.029</td>
<td>-0.431 ± 0.188</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Equation generated by linear regression of proficiency test data examined in this study</th>
<th>RSD_R at 1 mg/kg</th>
<th>RSD_R at 0.1 mg/kg</th>
<th>RSD_R at 0.01 mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty</td>
<td>( \log_{10} s_R = 0.855 \times \log_{10} c - 1.660 )</td>
<td>16%</td>
<td>23%</td>
<td>32%</td>
</tr>
<tr>
<td>Nonfatty</td>
<td>( \log_{10} s_R = 1.025 \times \log_{10} c - 0.457 )</td>
<td></td>
<td>ca 25% over entire concentration range</td>
<td></td>
</tr>
</tbody>
</table>
and large deviations being rare (which is typical for a normal distribution of random variables).

One popular application of the top-down concept is the Horwitz equation. Horwitz and coworkers combined approximately 7000 between-laboratories variability data from method performance tests into one data pool. This data set was then used to derive an empirical relationship for between-laboratories precision and concentration of the analyte (1, 7, 8).

The Horwitz equation, \( s_R = 0.02 \times c^{0.8495} \), is equivalent to the linear function that describes the regression line between the logarithm of the between-laboratories standard deviation, \( s_R \), and the logarithm of the concentration, \( c \) (in g/g). Equivalent expressions of the formula include the following:

\[
\log_{10} s_R = 0.8495 \times \log_{10} c - 1.6990
\]

\[
\text{RSD}_R \%, = 2^{1.05\log_{10} c}
\]

\[
\text{RSD}_R \%, = 2 \times c^{-0.1505}
\]

The Horwitz equation is in agreement with the well-accepted rule that high concentrations may be determined more precisely than low concentrations. The equation predicts that lowering the concentration by 2 orders of magnitude will double the between-laboratories relative standard deviation (RSD\(_R\)) associated with the result.

The overall effect of the combination of data from many different interlaboratory method performance tests is a more reliable estimate of uncertainty because of the increased degrees of freedom and the random selection of all significant inputs.

Advantages of the top-down concept, based on the Horwitz equation, include the following: the incorporation of laboratory biases in the uncertainty calculation because laboratory (or analyst) variability is also randomized (this is in contrast to the “bottom-up” approach); increased confidence in surveillance monitoring results in relation to consumer protection and assessment against established regulatory limits (maximum residue limit, tolerance, or similar limit) because deviations generated by different laboratories have been included; the fact that this approach requires traceability to a stated reference value of an incurred reference material to a lesser extent because laboratory bias is already considered, and only the influence of the method bias is unknown without a reference material; the fact that the estimation of uncertainty is very simple, because the Horwitz equation is applicable to all concentrations, methods, and analytes; and the fact that the influence of methods and/or analytes on precision may be tested.

The problems/drawbacks associated with the top-down concept based on the Horwitz equation include the facts that appropriate and sufficient data are needed as the basis for the estimation of a valid relation between concentration and uncertainty (the precision data used by Horwitz came from a diverse range of collaborative trials involving not just pesticide residue methods, and the range of concentrations considered

<table>
<thead>
<tr>
<th>Pesticide</th>
<th>Apple</th>
<th>Carrot</th>
<th>Tomato</th>
<th>Orange</th>
<th>Wheat</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acephate</td>
<td>45</td>
<td>33</td>
<td>33</td>
<td>35</td>
<td>24</td>
<td>34</td>
</tr>
<tr>
<td>Demeton-S-methyl</td>
<td>64</td>
<td>50</td>
<td>50</td>
<td>69</td>
<td>48</td>
<td>56</td>
</tr>
<tr>
<td>Demeton-S-methylsulfone</td>
<td>47</td>
<td>35</td>
<td>35</td>
<td>39</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Dimethoate</td>
<td>25</td>
<td>19</td>
<td>18</td>
<td>21</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>Ethoprophos</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Fenpropimorph</td>
<td>54</td>
<td>46</td>
<td>46</td>
<td>55</td>
<td>24</td>
<td>45</td>
</tr>
<tr>
<td>Fenthion</td>
<td>24</td>
<td>32</td>
<td>32</td>
<td>36</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>Iprodione</td>
<td>42</td>
<td>31</td>
<td>31</td>
<td>49</td>
<td>47</td>
<td>40</td>
</tr>
<tr>
<td>Malathion</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Methamidophos</td>
<td>49</td>
<td>34</td>
<td>32</td>
<td>32</td>
<td>34</td>
<td>36</td>
</tr>
<tr>
<td>Methidathion</td>
<td>22</td>
<td>19</td>
<td>21</td>
<td>22</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>Mevinphos</td>
<td>24</td>
<td>21</td>
<td>20</td>
<td>27</td>
<td>30</td>
<td>24</td>
</tr>
<tr>
<td>Omethoate</td>
<td>59</td>
<td>31</td>
<td>31</td>
<td>37</td>
<td>30</td>
<td>38</td>
</tr>
<tr>
<td>Paraoxon</td>
<td>21</td>
<td>17</td>
<td>17</td>
<td>31</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Paraoxon-methyl</td>
<td>31</td>
<td>20</td>
<td>20</td>
<td>31</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Parathion-methyl</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>17</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Phorate</td>
<td>26</td>
<td>18</td>
<td>20</td>
<td>29</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>Phosmet</td>
<td>30</td>
<td>24</td>
<td>28</td>
<td>37</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>Phosphamidon</td>
<td>26</td>
<td>28</td>
<td>28</td>
<td>31</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>Terbufos</td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>24</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 3. Between-laboratories RSDs (%) resulting from the matrix effect of some nonfatty samples
Table 4. Calculation of expanded uncertainty based on between-laboratories precision data from proficiency tests

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Standard deviation, mg/kg</th>
<th>Expanded uncertainty, mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty\a</td>
<td>( S_R = 2 \times 10^4 \times (10^{-6} \times \bar{X})^{0.8495} )</td>
<td>( U = 4 \times 10^4 \times (10^{-6} \times \bar{X})^{0.8495} )</td>
</tr>
<tr>
<td>Nonfatty\b</td>
<td>( S_R = 0.25 \times \bar{X} ) (equivalent to ( \bar{X} ))</td>
<td>( U = \bar{X} \times 0.5 \times 2 \times 25% )</td>
</tr>
</tbody>
</table>

\a Based on the Horwitz equation.
\b Based on data of proficiency tests with nonfatty materials; valid for concentrations in the range 0.03–3.0 mg/kg.

was very large [0.05 \( \mu g/kg \text{–} 60\% \)]; the Horwitz equation was derived from method validation studies (collaborative trials) using prescribed methods, not from proficiency tests for which methods were not prescribed (the estimation of uncertainty from method validation studies excludes the bias associated with the method [9]); the resulting estimates of uncertainty are based on a distribution of the between-laboratories standard deviation, \( s_R \), that is pertinent to the “typical” laboratory (the relevance of this uncertainty to all routine laboratories [particularly bad-performing laboratories] is questionable).

**New Straightforward Approach Based on Proficiency Test Data**

A number of international organizations and working groups, including the Pesticides Working Group of the “Lebensmittelchemische Gesellschaft” [part of the Society of German Chemists (10)], the Analytical Methods Committee The Royal Society of Chemistry (5), the Codex Committee on Methods of Analysis and Sampling (CCMAS; 11), and the Nordic Committee on Food Analysis (NKML; 12), have expressed some reservations regarding the use of the “bottom-up” approach for calculating uncertainties in food laboratories.

The approach is seen to be too complicated and probably unnecessary if the method has been validated, is used in interlaboratory proficiency tests, and incorporates appropriate internal quality assurance procedures. In addition, significantly different uncertainties may be determined by laboratories with comparable precision and trueness, caused by the differences in available data and the estimates based on judgment.

The revision of the EURACHEM Guide (6) supports the use of in-house validation study data and provides assistance for the estimation of measurement uncertainty based on quality assurance data and proficiency testing schemes. The latter option for the top-down estimation of uncertainty based on proficiency test results is discussed in the following sections.

Uncertainty derived from proficiency tests data.—In the straightforward Proficiency Test (PT)-based approach, the “top-down” concept was used to estimate general uncertainties in pesticide residue analysis based on data from 61 interlaboratory proficiency tests organized by federal government agencies in Australia, Canada, and the United Kingdom as well as Germany and The Netherlands (under the supervision of the European Commission), and by industry (Nestlé, Inc., The Netherlands); 1700 participants (median: 9 laboratories per test); 13 029 individual concentrations combined to give 869 relative standard deviations (RSDs) of mean (or median) residue concentrations; 15 nonfatty (horticultural) matrices including apple homogenate fresh and dry, carrot puree, cucumber, flour, grain, grape extract, grape pulp, lettuce extract, lettuce, pear, pepper, potato powder, strawberry pulp, and tomato; 9 fatty (animal) matrices including butterfat, cod liver oil, cod liver tissue, fish oil, meat, milk powder, beef drippings, lard, and trout powder; and 173 different pesticides determined in the proficiency tests.

All proficiency test standard deviations were calculated after the removal of outliers. Given that outliers were most often removed by the organizers, different approaches were used. The mean RSD of all 869 proficiency test results was 26%, with most RSDs (>95%) below 50%.

Only small differences between the means of RSDs for fatty and nonfatty matrices were observed. The mean RSD of results for nonfatty (horticultural) matrices was 27%, and the mean RSD of results for fatty (animal) matrices was 24%.

The complete data set is presented in Figure 1 and shows the standard deviations for each pesticide/matrix combination in this study as a function of concentration.

The data in Figure 1 show that the slope of the regression line generated by our data is closer to 1 than is the slope generated by the Horwitz equation. A slope of 1 indicates that the standard deviation is directly proportional to concentration and that the RSD is independent of concentration.

In order to identify the most important factors contributing to the observed correlation between \( \log_{10} \) \( S_R \) and \( \log_{10} \) \( c \), the entire data pool was divided into subsets based on matrix types (fatty or nonfatty) and pesticide/matrix combinations. For example, Figure 2 presents the data for all pesticides divided into fatty and nonfatty matrices. All categories considered are shown in Table 1.

Table 1 clearly shows a difference between the Category 2 and Category 3 proficiency test results. The regression for the nonfatty matrix data has a slope that is very close to 1, indicating that the RSD is independent of concentration. In contrast, the slope found in proficiency tests for the fatty matrices is similar to the slope derived from the Horwitz equation (as shown in Figure 1) and indicates a concentration dependence.

Given that >75% of all organochlorine pesticides were added to fatty matrices, the difference in observed behavior could simply be a reflection (hidden effect) of the fact that organochlorines are easier to determine in fatty materials.

However, inspection of the Categories 4 and 5 results in Table 1 suggests that the effect is not due to the types of analytes, but rather to the matrix. Again no concentration dependence of the standard deviation is observed for proficiency tests involving nonfatty matrices.

A similar relationship is evident in Categories 6 and 7 of Table 1 for pesticides other than organochlorines. The uncertainty of results is less concentration dependent for nonfatty matrices than for fatty matrices.
Statistical Significance of the Different Behavior of Fatty and Nonfatty Samples in Residue Analysis

The equations calculated for the regression lines for fatty and nonfatty matrixes were $\log_{10} s_R = 0.855 \times \log_{10} c - 1.66$ and $\log_{10} s_R = 1.02 \times \log_{10} c - 0.46$, respectively (see Table 1, Categories 2 and 3).

The null hypothesis that there is no difference between slopes of 0.855 and 1.02 was tested in accordance with the procedures outlined by Sachs (13). Given that the calculated $t$-value of 2.36 was larger than the critical value of 1.96, the null hypothesis was false.

In a further statistical test, the matrix type was coded with a dummy variable that was set to “0” for nonfatty matrixes and to “1” for fatty matrixes.

This additional variable allowed a 2-factor regression between standard deviation, matrix type, and concentration to be run in accordance with the process described by Dawson-Saunders and Trapp (14).

From the 2-factor regression, a regression coefficient of 0.067 between matrix type (“0” or “1”) and logarithmic concentration was calculated with a corresponding $t$-value of 4.49. Given that the $t$-value of 4.49 was much larger than the critical value of 1.96, a significant correlation exists between $\log_{10} s_R$ and $\log_{10} c$ as well as standard deviation and matrix type.

It should be noted that when large amounts of data are tested, even very small differences between 2 subsets of data can appear to be significantly different, despite the fact that there may not be a practical difference between the 2 groups. However, the rather surprising finding (an RSD independent of concentration) was also described in a paper by Thompson and Lowthian (15). Thompson and Lowthian re-evaluated the method validation data of collaborative trials used by Horwitz; they showed that the Horwitz function holds at RSDR values of about 30% at concentrations below $1 \times 10^{-8}$ g/g. They stated that the better precision achieved by laboratories at very low levels may be because analysts take more time and care to avoid RSDs of $\geq 40\%$, because results, under those circumstances, are likely to become meaningless.

In summary, both statistical tests applied had the same result. Although for fatty materials, the variability increased as the concentration decreased in line with the Horwitz equation, for nonfatty materials between-laboratories RSDs remained almost constant over the entire concentration range studied, as shown in Table 2.

Possible Rationales for Constant Uncertainty of Nonfatty Matrixes

On average the uncertainty is somewhat greater in studies involving horticultural products than in those involving fatty/animal products. The absence of an improvement in pre-
cision with increasing pesticide concentrations in the case of nonfatty materials may be due to several factors, including the following: (1) the lack of appropriate certified reference materials in horticultural matrixes, compared with the reference materials available in a fatty (animal) matrix, makes it more difficult to determine sources of error; (2) the range of pesticides tested in nonfatty food matrixes is frequently wider than that for fatty materials, e.g., the proficiency tests investigated in this study included 131 individual pesticides in nonfatty samples and 60 pesticides in fatty matrixes; (3) proficiency tests are often conducted at pesticide concentrations near the limit of quantitation (LOQ; e.g., maximum of 5–10 times the LOQ); thus, precision parameters reported at higher concentrations exclusively may reflect the precision of compounds that are inherently more difficult to determine; (4) extraction of animal materials is more straightforward and harmonized than is extraction of fruits and vegetables, and larger deviations in extraction efficiency between laboratories may be the consequence for nonfatty samples; (5) the instability of pesticides in standard solutions may cause some deviations (in animal test materials, less stable pesticides are often not used because they are metabolized in the animal and do not give rise to incurred residues); and (6) the matrix effect of extracts of fruits, vegetables, and grains on the response of pesticides may be more variable than the matrix effect of animal extracts.

Factors (4; 6) contribute Type B effects to the uncertainty, which might be greater and more variable in the analysis of horticultural products than in the more standardized analysis of animal products.

An unpublished German collaborative study indicated that the last hypothesis may be a significant contributor to variability. In this study the matrix effect was examined by 20 laboratories, using a total of 47 GC instruments. Ethyl acetate extracts of control samples of 5 matrixes (1 mL) were prepared by 1 laboratory, and each contained the extractable compounds from 4.7 g sample after solvent partitioning and GPC cleanup. The concentrations of the pesticides added to these matrix extracts were measured against a pure ethyl acetate standard containing identical amounts of the same spiking solution. Although the observed matrix effect differed between laboratories and instruments, the means of the apparent concentrations of all pesticides in the matrix extracts were generally higher than the actual concentrations added. Furthermore, perhaps of even greater importance in the context of measurement uncertainty was the extent of variation of the apparent concentrations. Table 3 shows the RSDs of the mean concentrations measured for each pesticide/matrix combination by using all 47 GC instruments. It should be noted that outliers were excluded before the calculations.

The observed variation of the apparent concentrations is rather high with an overall mean value for the RSD of 29%.

![Figure 4. Precision results of 455 pesticides tested in 37 proficiency tests with nonfatty materials, compared with the expanded uncertainty calculated with a concentration-independent RSD of 25%](image-url)
This “matrix effect standard deviation” is higher than the overall standard deviation of the proficiency tests, most likely because very polar pesticides were chosen. Despite the fact that the pesticides are polar, this study provides good evidence of some sources of measurement uncertainty that have traditionally been underestimated. Although this phenomenon has been discussed recently (16–20), it has rarely been seriously considered, let alone taken into account, in the previous 20 or 30 years of pesticide residue analysis.

With respect to the observed constant uncertainty in the case of nonfatty matrices, it should be noted that the proficiency test database contains a very small set of data that are <0.03 and >3 mg/kg. Consequently, at concentrations of <0.03 mg/kg, RSDs of >25% cannot be excluded and are probable. On the other hand, at concentrations of >3 mg/kg, better precision is likely, and RSDs of <25% may be generally achievable, especially if liquid chromatographic methods are used. The data presented in this paper for the pesticide residue determinations in nonfatty matrixes should not be seen as fundamentally opposing the Horwitz equation.

Reporting Uncertainty

It is generally accepted that an analytical result should be expressed with an expanded uncertainty quoted. The expanded uncertainty of a measured value “X,” calculated by using a coverage factor of 2 (which is related to a level of confidence of approximately 95%), is shown below:

\[ w_P = X \pm U \]

where \( w_P \) is the mass ratio (concentration) of the pesticide, \( X \) is the measured value (in mg/kg), and \( U \) is the expanded uncertainty.

If all concentrations in this study are expressed in mg/kg, the equations in Table 4 can be used to calculate the expanded uncertainty (\( U \)) for a fatty matrix for any analyte concentration and for a nonfatty matrix between 0.03 and 3 mg/kg.

The expanded uncertainty (\( U \)) calculated in this way reflects most sources of variability in pesticide residue analysis. As expected, very few proficiency test results show RSDs higher than those calculated for the expanded uncertainties in Table 4 (see also Figures 3 and 4).

Although the expanded uncertainties for some particular pesticide/matrix combinations may be significantly lower than those in Table 4, the only way to demonstrate such a lower expanded uncertainty is by an interlaboratory method performance test involving identical pesticide(s) and matrices. In all other cases, estimates derived by using the equations in Table 4 would provide reliable values of expanded uncertainty, based on the average proficiency test results of many countries.

Conclusions

The between-laboratories variability (expressed as the coefficient of variation), within the practical concentration range of a working pesticide laboratory, clearly shows a significant difference between fatty and nonfatty matrices.

Although it may be argued that the concentration range examined in this paper was too narrow, it was, however, realistic if pesticide residue analysis is the prime function of the laboratory. Consequently, it is valid to examine this relationship while at the same time recognizing that it focuses on a subset of the concentration range covered by Horwitz.

In this study, we examined the data generated by 61 separate interlaboratory proficiency tests run in 5 different countries and found that, in relation to the Horwitz equation, the concentration dependence is (1) virtually identical for fatty materials and (2) significantly different for nonfatty materials.

Until better estimates are available, it is proposed that the corresponding RSDs of the Horwitz equation be used as estimates of standard uncertainty with respect to pesticide analysis of fatty (animal) matrixes. For nonfatty (horticultural) samples a constant RSD of 25% seems more appropriate for the uncertainty approximation. Many systematic and random effects such as repeatability within the laboratory as well as randomized components of laboratory and method bias have already been incorporated in this estimate.

This practice may sometimes result in unexpectedly “pessimistic” estimates of uncertainty, but it has the advantage that no detailed calculations are required to determine the uncertainty budget based on various contributions from numerous sources of the analytical process.

When the PT-based approach is implemented, it should be noted that the estimation of uncertainty cannot simply be used by each individual laboratory unless the laboratory first demonstrates better (or at least equivalent) analytical performance.

The relevance of the performance data used in this paper to a single laboratory should be demonstrated by (1) achieving a within-laboratory standard deviation smaller than the between-laboratories standard deviation in Table 2 (in the case of difficult pesticides this has to be demonstrated for each individual chemical); (2) successfully participating in proficiency test programs with Z-scores of >2 kept to a minimum (about 5%); (3) having no large method and/or laboratory bias for recovery tests; and (4) obtaining acceptable analytical results after the analysis of suitable certified matrix reference materials.

If not all of these requirements are fulfilled, separate determinations of uncertainty based on a concept different from the proposed PT-based approach using proficiency test data would be necessary.

Acknowledgments

We thank Julie Fillion of the Pest Management Regulatory Agency, Laboratory Services, Canada, and Henri Diserens of the NESTEC Ltd. Research Centre, Switzerland, for providing the Canadian and Nestlé data used in this study.
References

(12) Estimation and Expression of Measurement Uncertainty in Chemical Analysis, NKML Procedure No. 5 (1997) Nordic Committee on Food Analysis (NKML); General Secretariat of NKML, c/o VTT Biotechnology and Food Research, PB 1500, FIN-02044 VTT, Finland
(15) Thompson, M., & Lowthian, P.J. (1997) J. AOAC Int. 80, 676–679