STATISTICAL ANALYSIS

Contribution of Sampling to the Variability of Pesticide Residue Data

ÁRPÁD AMBRUS and EUGENIA SOBOLEVA
International Atomic Energy Agency, Agrochemicals Unit, Wagramerstrasse 5 A-1400 Vienna, Austria

The uneven distribution of pesticide residues among the treated objects leads to an inevitable variability of pesticide residue levels measured in the samples, which may significantly contribute to the combined uncertainty of the analytical results. A total of 8844 unit-crop residue data derived from 57 lots and 19 field trials were evaluated to determine the characteristic features of residue distribution in unit crops and composite samples. The average residue levels and the corresponding coefficient of variation (CV) values obtained for individual units taken from a given lot showed wide variation from lot to lot. There was no significant difference between the CVs of residue levels in sample sets of various unit crops or composite sample populations of different sizes taken from various crops. The CV values for levels of residues taken from individual lots followed normal distribution. Very good correlation was found between the CVs of the parent and sample populations. The experimentally obtained values were very close to those expected on the basis of the central limit theorem. The estimated typical relative standard uncertainties of sampling medium-size crops for pesticide residue analysis in the cases of sample sizes of 5, 10, and 25 were 37, 25, and 16%, respectively.

The distribution of pesticide residues among individual items of a treated crop/field may be influenced by several factors such as the application, the crop, and the environment, and the chemical, physical, and biological properties of the substance (1). The uneven distribution of pesticide residues among the treated objects leads to an inevitable variability of pesticide residue levels measured in the samples, which may significantly influence the combined uncertainty of the analytical results. Analysts have disregarded the effect of sampling on the combined uncertainty of the results for a long time, taking the very convenient but unscientific position that they are responsible only for the analysis of samples and not for the sampling process itself.

Received March 3, 2004. Accepted by GL May 25, 2004.
Corresponding author's e-mail: ambrus.arpad@ontsz.hu.

The importance of sampling was recognized practically from the beginning of residue analysis, and the various aspects to be considered during sampling were described long ago (2). Nevertheless, very limited information is available on the contribution of sampling to the combined uncertainty of the results; however, this information is necessary for correct interpretation of the results.

The objectives of this paper are (*1*) to present the results of the estimation of uncertainty of sampling and provide data for inclusion of sampling uncertainty in the combined uncertainty of the results, which is one of the basic requirements of the ISO 17025 Standard; (*2*) to identify the major factors affecting sampling uncertainty; and (*3*) to indicate the areas in which further work or information is required.

Summary of Theory–Definition of Terms

The pesticide residues in natural units or sample increments making up the sampled object form the *parent population*. The *primary sample* is ≥1 crop unit(s) taken from 1 position in a lot. The *sample size* is the number of primary samples in 1 sample. A *composite sample* consists of several primary samples. Samples taken repeatedly from the same parent population form the *sample population*. The pesticide residues measured in primary and composite samples form the primary and sampling distributions, respectively.

If the residue concentration levels (c_i) are measured separately in unit crops (primary samples), making up the composite sample, the residue concentration in the composite sample (R_j) is calculated as follows:

$$R_j = \frac{\sum_{i=1}^{n} g_i * c_i}{\sum_{i=1}^{n} g_i} \qquad (1)$$

where g_i is the mass of an individual unit i, c_i is the corresponding residue concentration, and n is the number of units or items making up 1 composite sample.

If the mass of the primary sample is not available, the residue concentration in the composite sample, of size n, is calculated as follows:

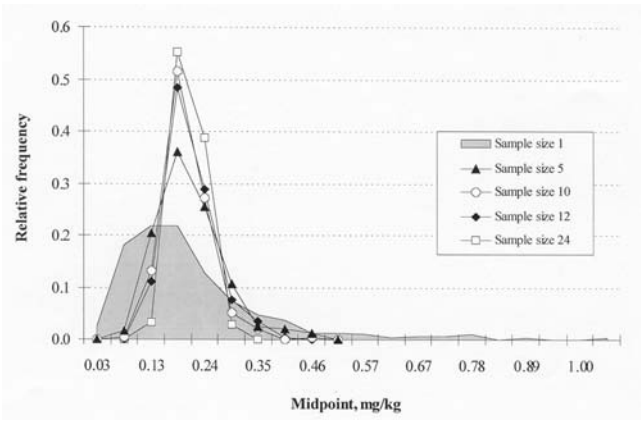

Figure 1. Distribution of chlorpyrifos-methyl residues in apple samples.

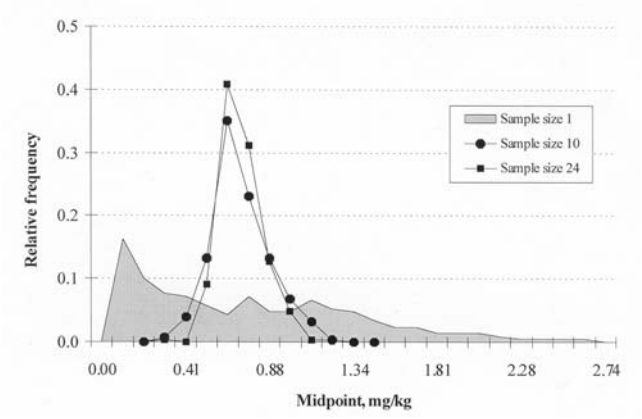

Figure 2. Distribution of vinclozolin residues in kiwi fruits.

$$R_j = \frac{\sum_{i=1}^{n} c_i}{n} \quad (2)$$

The average residue concentration in a set of primary samples, of size k, is calculated as follows:

$$R_k = \frac{\sum_{i=1}^{k} g_i * c_i}{\sum_{i=1}^{k} g_i} \quad (3)$$

The basic rules related to the parent and sampling distributions, provided that the errors of sample processing and analysis are negligible, are as follows (3):

(*1*) Whatever the shape of the frequency distribution among c_i values of the population with mean μ and standard deviation (SD) σ, the frequency distribution of R_j in repeated random composite samples of size n tends to become normal as n increases (Figures 1 and 2).

(*2*) If the parent distribution is normal, the sampling distribution will also be normal for any sample size.

(*3*) The relation between the SD among primary samples (S) and composite samples of size n (S_{Rn}) is as follows:

$$S_{Rn} = S/\sqrt{n} \quad (4)$$

(*4*) There is less variation among composite sample means, R_j, than among individual concentrations, c_i. The larger the sample size, the smaller this variation (S_{Rn}) becomes.

(*5*) The frequency distribution (sampling distribution) of composite sample means, R_j, has a mean of μ and an SD of $\sigma/\sqrt{n}$ in the case of random sampling when $n/N < 0.02$ or sampling is performed with replacement. (N is the number of units (observations) in the parent population.)

(*6*) The average of c_i values in a composite sample (R_j) taken randomly from a parent population is an unbiased estimator of the mean residue (μ) of the population.

(*7*) If random samples are drawn with replacement from any population, the average value of the variance taken over all random samples is exactly equal to σ^2; thus, under the above conditions, S^2 is an unbiased estimate of σ^2.

The combined uncertainty of the test procedure (S_R) may be expressed as follows:

$$S_R = \sqrt{S_S^2 + S_L^2} \quad (5)$$

where S_S is the uncertainty of sampling and S_L is the combined uncertainty of the analysis (A) and sample processing (*Sp*), that is the process used to obtain a homogeneous matrix from the sample from which representative test portions are withdrawn for extraction. If the whole sample is analyzed, the mean residue remains the same, and Equation 5 can be written as follows:

$$CV_R = \sqrt{CV_S^2 + CV_L^2} \quad (6)$$

CV_L may be calculated as follows:

$$CV_L = \sqrt{CV_{Sp}^2 + CV_A^2} \quad (7)$$

If the whole sample is processed and analyzed, the uncertainty of sample processing, CV_{Sp}, is zero.

Residue Data Used for Estimation of Sampling Uncertainty

To estimate the uncertainty of sampling (S_S or CV_S), very large numbers of samples are required for analysis, which is expensive and time-consuming. Therefore, all available data suitable for estimation of sampling uncertainty have been evaluated. These included the results of 19 field trials specifically carried out to obtain information on the

Table 1. Summary of residues measured in unit crops/primary samples (*k*) and the coefficients of variation (CVs) of residues in random composite samples drawn with replacement from the primary sample populations

Pesticide	Country	Residues in primary samples, mg/kg[a]						CV of residues in composite samples		
		k	Minimum	Maximum	Average	Median	CV	*n* = 5	*n* = 10	*n* = 25
					Apple					
Carbaryl	United States	108	0.256	3.89	1.412	1.30	0.50	0.24	0.18	0.10
Carbaryl	New Zealand	95	0.005	1.90	0.358	0.28	0.91	0.38	0.28	0.18
Carbaryl	Argentina	100	0.02	0.68	0.152	0.12	0.83	0.38	0.27	0.17
Carbaryl	United Kingdom	100	0.33	2.26	1.055	1.02	0.39	0.17	0.12	0.08
Carbaryl	United Kingdom	100	0	1.62	0.507	0.40	0.67	0.28	0.21	0.13
Carbaryl	United Kingdom	100	0.01	2.73	0.977	0.94	0.64	0.29	0.19	0.13
Chlorpyrifos	United Kingdom	110	0.005	1.35	0.151	0.09	1.19	0.52	0.40	0.24
Chlorpyrifos	United Kingdom	110	0.005	0.26	0.056	0.04	0.78	0.33	0.26	0.15
Chlorpyrifos	United Kingdom	100	0.01	0.34	0.061	0.05	0.86	0.39	0.24	0.17
Chlorpyrifos-methyl	Hungary[b]	319	0.005	1.05	0.212	0.17	0.69	0.30	0.23	0.13
Chlorpyrifos-methyl	Hungary[b]	320	0.002	0.11	0.027	0.023	0.64	0.29	0.20	0.13
Diphenylamine	United States	108	0.089	1.82	0.473	0.38	0.63	0.29	0.20	0.13
Phosalone	France	100	0.0787	1.75	0.482	0.41	0.55	0.25	0.19	0.11
Phosphamidon	Hungary[b]	108	0.01	0.71	0.165	0.11	0.84	0.38	0.25	0.18
Thiabendazole	United States	108	0.347	2.97	1.021	0.87	0.50	0.23	0.15	0.10
Triazophos	United Kingdom	110	0.005	2.16	0.558	0.36	0.97	0.45	0.32	0.19
					Banana					
Chlorpyrifos-methyl	Jamaica	100	0.0028	0.091	0.008	0.01	1.18	0.51	0.34	0.26
Chlorpyrifos-methyl	Surinam	100	0.0015	0.075504	0.009	0.01	0.88	0.43	0.30	0.16
					Cherry					
Dimethoate	Hungary[b]	120	0.011	0.919	0.190	0.13	1.08	0.47	0.34	0.21
					Cucumber					
Vinclozolin	Hungary[b]	120	0.005	0.215	0.0656	0.0585	0.55	0.23	0.16	0.10
					Grape					
Chlorpyrifos-methyl	Hungary[b]	120	0.075	4.031	0.509	0.39	0.99	0.45	0.29	0.20
Metalaxyl	Hungary[b]	120	0.011	1.082	0.324	0.29	0.64	0.28	0.21	0.13
Vinclozolin	Hungary[b]	120	0.142	11.712	1.493	1.16	0.92	0.44	0.34	0.20
					Kiwi					
Chlorpyrifos	New Zealand[b]	209	0.005	0.72	0.170	0.12	0.96	0.42	0.30	0.21
Diazinon	New Zealand	100	0.001	0.035	0.011	0.01	0.58	0.26	0.19	0.11
Diazinon	New Zealand[b]	209	0.001	0.14	0.046	0.05	0.62	0.26	0.19	0.13
Parathion-methyl	Greece	100	0.0005	0.026	0.014	0.01	0.32	0.15	0.10	0.06
Parathion-methyl	Greece	100	0.002	0.019	0.009	0.01	0.36	0.16	0.11	0.07
Permethrin	New Zealand[b]	209	0.005	0.21	0.050	0.05	0.79	0.35	0.25	0.16
Phosmet	Chile	100	0.001	0.427	0.071	0.03	1.29	0.59	0.41	0.28
Pirimiphos-methyl	New Zealand[b]	209	0.005	1.09	0.151	0.09	1.10	0.48	0.33	0.21
Quinalphos	Italy	100	0.0005	0.103	0.022	0.01	1.02	0.45	0.32	0.19
Vinclozolin	New Zealand[b]	209	0.01	2.64	0.759	0.68	0.80	0.37	0.24	0.17
					Orange					
Bromopropylate	Cyprus	100	0.01	0.984	0.357	0.35	0.56	0.23	0.18	0.11
Chlorpyrifos	Spain	100	0.0035	0.301	0.074	0.06	0.83	0.34	0.25	0.17

Table 1. (*continued*)

Pesticide	Country	Residues in primary samples, mg/kg[a]						CV of residues in composite samples		
		k	Minimum	Maximum	Average	Median	CV	*n* = 5	*n* = 10	*n* = 25
Imazalil	Spain	102	0.0035	0.389	0.117	0.11	0.68	0.30	0.20	0.13
Imazalil	Cyprus	100	0.0232	3.622	1.719	1.67	0.35	0.16	0.11	0.07
Imazalil	Uruguay	110	0.01	1.026	0.410	0.38	0.42	0.20	0.13	0.08
Imazalil	Morocco	100	0.0035	1.642	0.593	0.61	0.58	0.27	0.18	0.12
Malathion	Cyprus	100	0.01	1.016	0.191	0.15	0.77	0.34	0.25	0.16
Methidathion	Cyprus	100	0.01	2.916	0.674	0.53	0.89	0.43	0.28	0.18
Methidathion	Morocco	100	0.035	1.47	0.438	0.40	0.63	0.26	0.21	0.13
Parathion-methyl	Cyprus	100	0.01	2.11	0.459	0.12	1.28	0.54	0.39	0.24
				Peach						
Acephate	Italy	100	0.010	2.60	0.477386	0.382	0.92	0.41	0.31	0.18
Dimethoate	Italy	100	0.010	1.22	0.286716	0.229	0.79	0.35	0.22	0.15
Phosalone	Italy	100	0.01	0.9862	0.23035	0.17475	0.92	0.45	0.28	0.20
				Pear						
Carbaryl	Italy	110	0.0025	0.561	0.100	0.07	1.00	0.46	0.32	0.20
Carbaryl	The Netherlands	100	0.0025	0.087	0.020	0.01	0.94	0.41	0.30	0.20
Phosalone	France	100	0.076	1.548	0.526	0.44	0.62	0.29	0.20	0.12
				Plum						
Acephate	Spain	101	0.012	1.112	0.244	0.21	0.73	0.33	0.22	0.14
Acephate	Spain	100	0.003	0.12	0.044	0.04	0.66	0.29	0.20	0.13
Acephate	Portugal	100	0.01	0.53	0.131	0.10	0.82	0.36	0.26	0.16
Fenitrothion	Spain	100	0.0035	0.137	0.029	0.02	0.80	0.32	0.23	0.15
Methamidophos	Spain	101	0.002	0.184	0.042	0.03	0.78	0.35	0.25	0.16
Methamidophos	Spain	100	0.003	0.023	0.007	0.01	0.67	0.30	0.22	0.13
Phosalone	United Kingdom	100	0.0045	3.586	0.247	0.08	2.34[c]	1.14	0.79	0.47
Phosalone	Italy	100	0.07	1.895	0.390	0.25	0.96	0.43	0.30	0.20
Pirimiphos-methyl	Spain	100	0.0015	0.263	0.035	0.02	1.36	0.58	0.42	0.28
				Potato						
Aldicarb	United Kingdom	100	0.0045	0.45	0.08	0.05	1.10	0.51	0.36	0.23
Aldicarb	United Kingdom (main crop)	100	0.0045	0.54	0.07	0.05	1.14	0.46	0.33	0.18
Aldicarb	South Africa	100	0.01	0.49	0.08	0.06	0.94	0.41	0.29	0.19
Aldicarb	South Africa	100	0.02	0.40	0.05	0.04	0.82	0.38	0.25	0.18
Aldicarb	South Africa	100	0.02	0.22	0.06	0.04	0.57	0.24	0.19	0.11
Aldicarb	Jersey Island[d] (UK)	100	0.02	0.67	0.15	0.13	0.66	0.31	0.22	0.13
Aldicarb	United Kingdom	100	0.0045	0.60	0.05	0.03	1.54[c]	0.67	0.45	0.27
Aldicarb	United Kingdom[d]	100	0.0045	0.20	0.04	0.03	1.00	0.54	0.38	0.25
Aldicarb	United Kingdom[e]	100	0.0045	0.16	0.04	0.03	0.60	0.26	0.18	0.11
Aldicarb	United Kingdom[e]	100	0.0045	0.31	0.06	0.04	0.99	0.47	0.31	0.21
Aldicarb	United States[b]	79	0.01	0.74	0.12	0.08	1.14	0.54	0.34	0.22
Aldicarb	United States[b]	100	0.02	0.54	0.09	0.07	0.89	0.37	0.27	0.17
Aldicarb	United States[b]	100	0.02	0.37	0.09	0.08	0.67	0.31	0.22	0.12
Aldicarb	United States[b]	100	0.01	0.32	0.07	0.05	0.82	0.35	0.26	0.17
Aldicarb	United States[b]	100	0.04	0.34	0.14	0.13	0.52	0.23	0.17	0.11
Aldicarb	United States[b]	100	0.05	0.93	0.29	0.24	0.66	0.28	0.21	0.12

Table 1. (*continued*)

Pesticide	Country	Residues in primary samples, mg/kg[a]						CV of residues in composite samples		
		k	Minimum	Maximum	Average	Median	CV	*n* = 5	*n* = 10	*n* = 25
Aldicarb	United States[b]	100	0.01	0.39	0.11	0.10	0.63	0.28	0.22	0.12
Aldicarb	United States[b]	100	0.03	0.71	0.17	0.15	0.62	0.27	0.19	0.13
				Tomato						
Formetanate	Spain	100	0.0035	0.542	0.050	0.03	1.40	0.63	0.40	0.27
Methamidophos	Spain	100	0.00335	0.6732	0.061	0.04	1.44	0.65	0.44	0.28

[a] *n* = 1.
[b] Supervised field trials.
[c] The sampled commodity likely consisted of >1 lot.
[d] First harvest of new potato as fully maturated main crop.
[e] Main crop.

distribution of residues among unit crops or primary samples (4) and 57 lots sampled at markets (5).

To obtain reliable estimates of sampling uncertainty, only those data sets were used which contained >50 units and in which <20% of the samples had residues below the limit of quantitation (LOQ). The LOQ values were replaced with 0.5 LOQ, because it was shown with a model experiment (4), that the true coefficient of variation (CV) of the population was less affected by this replacement than by the use of LOQ values.

A total of 8844 crop units/primary samples were taken for consideration. The residue data represented 9 medium-size (unit weight, 50–250 g), 1 small-size (unit weight, <50 g), and 2 large-size (unit weight, >250 g) commodities, 26 different active ingredients, and 78 commodity-pesticide combinations (Table 1).

A limited number of field trials in which replicate composite samples were taken were also considered. The results are summarized in Tables 2 (6) and 3 (7).

Modeling Random Sampling

The Codex Maximum Residue Limits (MRLs) for fruits and vegetables apply to a composite sample, which should consist of a minimum of 5 or 10 natural units, *n*, for large- and medium-size crops, respectively (8). By applying a specifically designed computer program, 300 random composite samples of sizes 5, 10, 12, and 24 or 25 were drawn with replacement from each of the primary residue populations consisting of *k* crop units. The sample sizes were selected to estimate the uncertainty of sampling according to the Codex Sampling Procedure ($n = 5$ and 10) and the practice in supervised field trials in which large composite samples ($n = 24$–25) are usually taken, as well as to illustrate the effect of sample size. The sampling procedure is illustrated in Figure 3.

The average residues in the composite samples were calculated according to equation 1 from the residues measured in the primary samples. The SD and CV values of residues were calculated from the residues in the primary and sample populations.

Results and Discussion

Distribution of Residues in Primary and Composite Samples

The ranges of residue levels in unit crops and their CVs and the CVs of residue levels in composite samples are shown in Table 1. The relative frequency distributions of residues in primary and composite samples are illustrated in Figures 1 and 2. The distribution of residues in crop units is strongly skewed in a positive direction. The log-normal transformation of residues provided more symmetric distribution, but only about 50% of the data sets could be considered normal ($P = \geq 0.05$).

The test for normality of residues measured in composite samples indicated that the distribution of residues in samples of size 5 was still slightly skewed and generally could not be considered normal. The distribution of residues in samples of sizes 10–12 was normal or close to normal, whereas residues in samples of sizes ≥25 followed normal distribution.

The average residue levels and the corresponding CV values of residues among individual units taken from a given lot showed wide variation from lot to lot. There was no significant difference between the CVs of residues in sample sets of various unit crops with a wide range of average residues (Figure 4) or composite sample populations of various sizes as illustrated by sample size 5 in Figure 5.

The logarithm of the variances of the residues and the average residue levels in crop units taken from a lot gave a linear relationship (Figure 6) for the 76 data sets representing single lots. The good correlation between ln V and ln R or $V = 0.5001R^{1.9221}$ ($R^2 = 0.94$) confirms that the SD is proportional

Table 2. Folpet residues (mg/kg) in replicate composite samples[a] taken from 1 site at supervised field trials in various countries

	Folpet residues in				
Country	Replicate 1	Replicate 2	Replicate 3	Average	CV[b]
France	0.9	0.6	0.7	0.73	0.21
France	0.7	0.8	0.5	0.67	0.23
France	0.7	1.4	0.7	0.93	0.43
France	0.8	0.8	0.6	0.73	0.16
France	1.8	1.2	1.8	1.60	0.22
France	1.1	1.5	1	1.20	0.22
France	1.2	1.4	0.8	1.13	0.27
France	0.7	0.7	1.4	0.93	0.43
Hungary	5.4	4.4	5.1	4.97	0.10
Hungary	6.5	5.9	8	6.80	0.16
Portugal	2.7	2.8	2.6	2.70	0.04
Portugal	3	3.2	2.3	2.83	0.17
Spain	1.7	2	3.1	2.27	0.33
Spain	2.2	2.3	1.7	2.07	0.16
Switzerland	2.2	3.1	2.8	2.70	0.17
Switzerland	2.7	3.4	3.3	3.13	0.12
Average CV					0.21

[a] Sample size was not specified, but usually samples containing 20–25 apples are taken at supervised field trials carried out to support a petition for registration.

[b] CV = Coefficient of variation.

to the average residue level and, consequently, the CV of the residues is independent of the residue concentration.

The minimum, average, and maximum CV values of residue levels found in each crop are summarized in Table 4, which shows around 3-fold differences in the variability of residue levels among lots or fields. The 2 sets of tomato data indicate somewhat higher variability than do the other commodities. However, 1-way analysis of variance confirmed that the average CV values obtained for different crops were not significantly different (P = 0.0945). The above findings on the nature of residue distributions made it possible to estimate the typical sampling uncertainty for composite plant samples of various sizes.

Table 3. Chlorpyrifos residues (mg/kg) in 7 replicate composite samples taken randomly from a commercial orchard treated according to normal agriculture practice

Sample size = 12		Sample size = 24	
Avg. residue, mg/kg	CV[a]	Avg. residue, mg/kg	CV
0.16	0.18	0.1636	0.37
0.095	0.37	0.0779	0.40
0.036	0.24	0.0324	0.25
0.026	0.24	0.0289	0.28
0.021	0.17	0.0246	0.21
		0.16	0.33
		0.022	0.29
Avg. CV_R = 0.24; Avg. CV_S = 0.20		Avg. CV_R = 0.30; Avg. CV_S = 0.27	

[a] CV = Coefficient of variation.

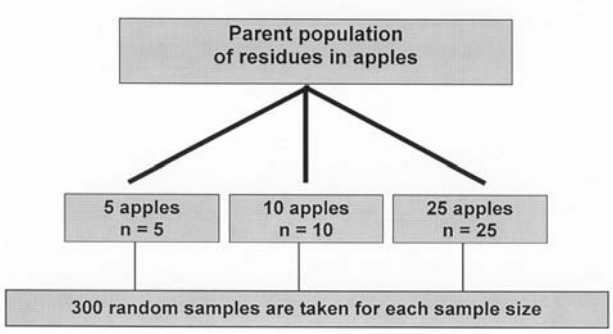

Figure 3. Schematic illustration of drawing random samples from apples.

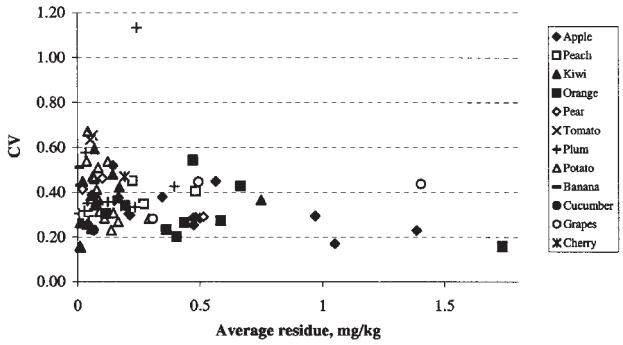

Figure 5. The coefficient of variation (CV) of pesticide residues in 78 sets of composite samples of size 5.

Estimation of Typical Sampling Uncertainty

To obtain the best estimate of the sampling uncertainty, the random error of analysis was taken into account on the basis of Equations 6 and 7. Because whole individual units were analyzed, the random error of sample processing is equal to zero in Equation 7. The random error of analysis was not reported. Therefore, the uncertainty of the analysis was calculated with the Horwitz formula (9) from the mean residue levels measured, with the assumption that the within-laboratory reproducibility is about 2/3 of the among-laboratories uncertainty ($CV = 0.66*0.01*2C^{-0.1505}$). The correction for the analytical random error had no practical effect on the estimated typical CV values for medium-size commodities (the average CV of 0.806 was reduced to 0.798); therefore, it was concluded that the variability among residues in crop units reflects the sampling uncertainty alone.

A very good correlation was found between the CVs of the parent and sample populations. Figure 7 shows a linear relationship, with a slope of 1.0005 ($R^2 = 0.969$), between the CV values of the sample populations of size 5 and the $CV/\sqrt{5}$ value of the parent population. Figures 8 and 9 illustrate the relationships between the CVs of the parent populations and the sample populations of sizes 10 and 25. The linear regression equations and the regression coefficients are summarized in Table 5. The tabulated data indicate that the slopes of the linear regression equations obtained from the experimental data are very close to those theoretically expected on the basis of Equation 4 ($1/\sqrt{n}$). It can be seen that the nature of the crops or the chemical composition of the residues did not affect the relationship between the CVs of the parent and sample populations.

The SD of the residues in unit crops is roughly proportional to the mean; therefore, the CV is a fairly stable statistic over the range of residue values, and we may expect that its distribution is close to "normality." The statistical tests performed with Statgraphics 5.0 Plus for normality indicated (Kolmogorov $P = 0.101$; Chi square $P = 0.384$) that the combined population of all 76 CV values can be adequately modeled with normal distribution. The frequency distribution of the CV values of the residues in 76 data sets is illustrated in Figure 10. On the basis of the above findings, the typical

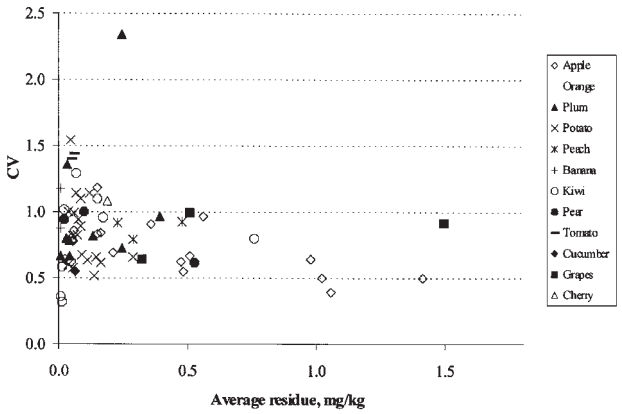

Figure 4. The coefficient of variation (CV) of pesticide residues in 78 sample sets consisting of 90–320 crop units.

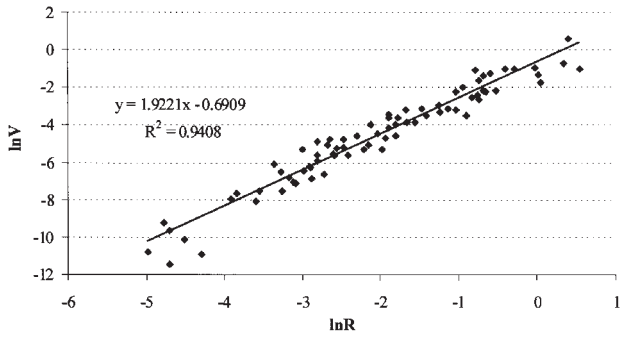

Figure 6. Relationship of the logarithm of variance and average residue levels in 76 pesticide residue data sets.

Table 4. Minimum, average, and maximum CV values of residues obtained in primary data sets (*n* = 1) of various crops

Crop	No. of data sets	CV[a]		
		Minimum	Average	Maximum
Apple	16	0.39	0.72	1.19
Banana	2	0.88	1.03	1.18
Cherry	1	1.08	1.08	1.08
Cucumber	1	0.55	0.55	0.55
Grape	3	0.64	0.85	1.66
Kiwi	10	0.32	0.78	1.29
Orange	10	0.35	0.70	1.28
Peach	3	0.79	0.88	0.92
Pear	3	0.62	0.85	1.00
Plum	9	0.66	0.85	1.36
Potato	18	0.52	0.85	1.54
Tomato	2	1.40	1.42	1.44

[a] CV = Coefficient of variation.

sampling uncertainty was calculated as the average of the CV values found for 76 data sets.

Most of the data sets contained 90–108 samples, and a few consisted of 120 and 320 unit-crop data. Therefore, the simple arithmetic mean was calculated instead of the weighted mean. Because the number of random primary samples taken from a lot was large (90–320), we can assume that the calculated variances and mean values and, consequently, the CV values, give reasonably accurate and unbiased estimates of the mean and the variance, μ and σ^2, respectively, of the residue populations in unit crops.

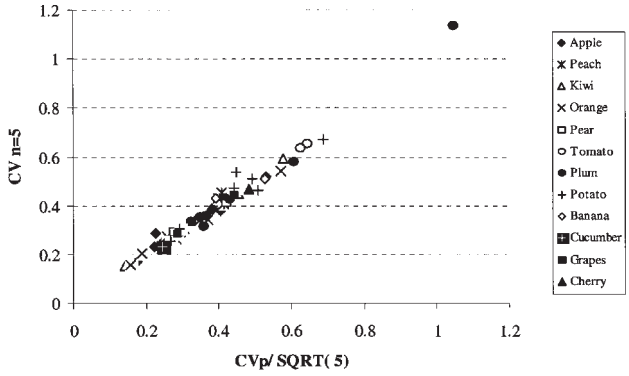

Figure 7. Correlation of coefficient of variation (CV) values of residues in composite samples of size 5 (*n* = 5) and square root 5 of CV values of residues in corresponding primary samples (*CVp*).

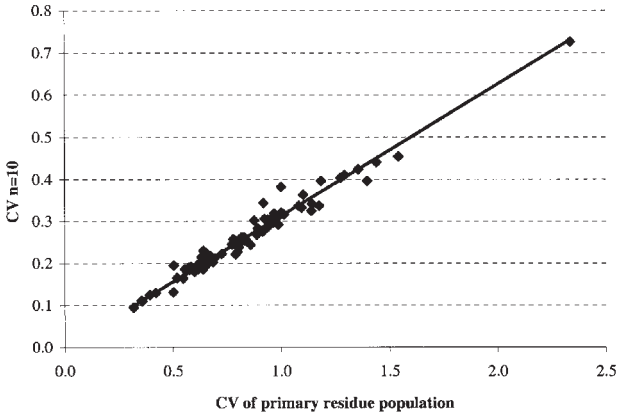

Figure 8. Correlation of coefficient of variation (CV) values of residues in primary samples (*CVp*) and in corresponding composite samples of size 10 (*n* = 10).

The calculated average CV values for the primary and composite samples are summarized in Table 6. The arithmetic mean CV of the residues of 26 pesticides in 12 different commodities obtained from 76 data sets is 0.81, which can be considered as the typical variability of residues in crop units. The typical CV values for composite samples are somewhat smaller than those predicted earlier, based on a limited number of data sets (10, 11).

The expected lower and upper tolerance limits of the CV values for individual data sets, at 95 and 99% probability and confidence levels, are shown in Table 7, which indicates that the CV of the residue values obtained for a single field may vary within a very large range.

The CV values of 1 data set each from plums (CV = 2.34) and from potatoes (CV = 1.54; Figure 4 or Table 1) were outside the 95 and 99% tolerance limits, indicating that the sampled commodities were probably a mixture of crops derived from several fields/lots. Therefore, these sets were not included in the calculation of the typical values reported above.

When the uncertainty of estimated SDs based on a few measurements is considered, it appears that the residues measured in composite samples taken from treated fields (Tables 2 and 3) fit well within the predicted range.

Because the slope obtained from the linear regression of corresponding CV values of residues in unit crops and composite samples was very close to the theoretically expected values (Table 4), the typical sampling uncertainty (CV_{Sn}) values for various sample sizes (n) can be calculated from the average CV_{1typ} of the residues in unit crops: $CV_{Sn} = CV_{1typ}/\sqrt{n}$.

Number of Unit Crops Required for Satisfying the Minimum Mass of the Composite Sample Specified in the Codex Sampling Guide

The Codex sampling procedure for medium-size crops requires a minimum of 10 crop units and a minimum sample

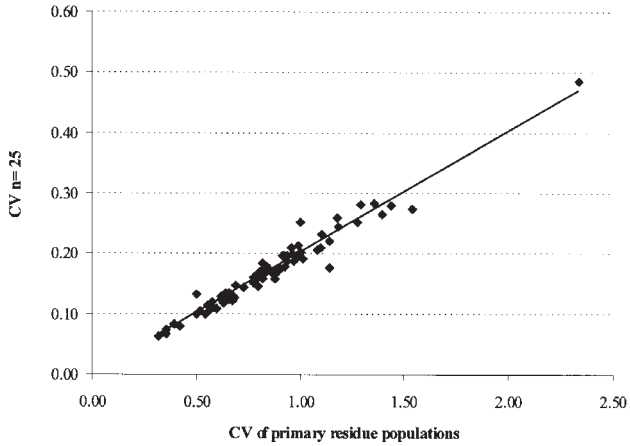

Figure 9. Correlation of coefficient of variation (CV) values of residues in primary samples (*CVp*) and in corresponding composite samples of size 25 (*n* = 25).

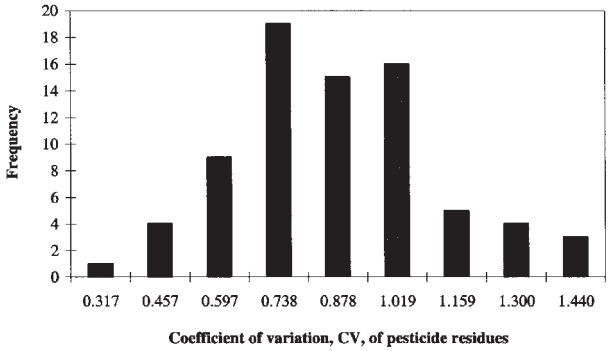

Figure 10. Frequency distribution of coefficient of variation (CV) values of residues in individual crop units.

of 1 kg; both criteria must be met. The weight ranges of randomly generated composite samples of size 10, based on the mass of crop units, are summarized in Table 8. For instance, the smallest mass values found in the 16 sets of 300 composite apple samples of size 10 were selected. The average, minimum, and maximum mass values of the 16 samples were 1248, 754, and 1796 g, respectively. The average of the minimum sample weights indicates that the Codex requirements would generally be satisfied with 10 units from apple, banana, orange, and pear; 14 units from kiwi and peach; and 20–38 units from plum, potato, and tomato. However, a cucumber or a bunch of grapes may be a medium- or large-size crop, depending on its variety, and the number of primary samples should be taken accordingly.

If the smaller crop units are taken randomly from different positions, and each unit can be considered to be a primary sample, then the expected sampling uncertainty can be calculated as $0.81/\sqrt{n}$, which can be used instead of the typical uncertainty estimated for medium-size crops. The calculated minimum number of crop units to be taken in a composite sample and the corresponding CV_S values are shown in Table 8. Naturally, in such cases the minimum number of primary sampling positions specified in the Codex Sampling Procedure is not applicable.

Table 5. Parameters of linear regression equations between the CVs[a] of residues measured in sample and parent populations

Sample size	Regression equation	R^2	$1/\sqrt{n}$
n = 5	$CV_5 = 0.4508CV_1$	0.975	0.447
n = 10	$CV_{10} = 0.3128CV_1$	0.9672	0.316
n = 25	$CV_{25} = 0.2007CV_1$	0.9575	0.2

[a] CV = Coefficient of variation.

Conclusions

The distribution of residues in crop units can be considered to be continuous and largely skewed in the positive direction. The log-normal transformation resulted in normal distribution only in about 50% of the 78 data sets. Therefore the log-normal transformation should be used only if sufficient data points are available for verifying the normality of the transformed data.

There was no difference between the 12 crops and or 26 pesticide active ingredients concerning the relationship of average residues and the CVs of the residue populations, or between the CVs of the parent and sample populations. There were around 3-fold differences in the variabilities of the residues, CVs, among lots or fields of 1 crop, but the difference was insignificant between the average CV values obtained for the crops and the pesticides examined.

The CV values of residues in crop units from 76 data sets were normally distributed. The calculated tolerance limits for the CV values indicate that larger variability may occur within individual fields, which should be taken into account when residues measured in replicate samples are compared, and the low or high values are discarded as outliers.

By taking into consideration the nature of residue distributions, it was possible to estimate a typical relative standard uncertainty of sampling of medium-size crop units, which is independent from the crop, pesticide, and residue level. The CV_{1typ} is 0.81 or 81%. The 95% confidence limits for the estimated typical relative uncertainty are 0.75 and 0.87.

The typical relative standard uncertainty of composite random samples of sizes 5, 10, and 25 are 0.37, 0.25, and 0.16, respectively. The distribution of residues in samples of sizes ≥25 was normal, and it was normal or close to normal in

Table 6. Average coefficients of variation (CVs) of residues in primary and composite samples and the recommended typical sampling uncertainty for medium-size crops

Crop	No. of data sets	Average CV			
		$n = 1$[a]	$n = 5$[b]	$n = 10$[b]	$n = 25$[b]
Apple	16	0.72	0.32	0.23	0.15
Banana	2	1.03	0.47	0.32	0.21
Cherry	1	1.08	0.47	0.34	0.21
Cucumber	1	0.55	0.23	0.16	0.10
Grape	3	0.85	0.39	0.28	0.18
Kiwi	10	0.78	0.35	0.24	0.16
Orange	10	0.70	0.31	0.22	0.14
Peach	3	0.88	0.40	0.27	0.17
Pear	3	0.85	0.39	0.28	0.17
Plum	8	0.85	0.37	0.26	0.17
Potato	17	0.81	0.37	0.26	0.16
Tomato	2	1.42	0.64	0.42	0.27
Avg. of 76 data sets		0.806	0.359	0.252	0.161
Typical uncertainty		0.81	0.37[c]	0.25	0.16
Confidence interval (95%)		±0.06	±0.03	±0.02	±0.01

[a] Average of CV_S of residues measured in crop units.
[b] Average of CV_S obtained from 300 composite samples generated with model calculations.
[c] CVp/√5 gave 0.3651; therefore 0.37 is considered the typical average uncertainty for sample size 5.

samples of sizes 10–12; therefore, normal statistics may be used for the evaluation of residue data based on samples of sizes ≥10.

The slopes of the linear regression of corresponding CV values of residues in unit crops and composite samples were very close to the theoretically expected values. Therefore, the typical sampling uncertainty (CV_{Sn}) values for various sample sizes (n) can be calculated from the average CV_{1typ} of the residues in unit crops: $CV_{Sn} = CV_{1typ}/\sqrt{n} = 0.81/\sqrt{n}$. This relationship can be used to estimate the sampling uncertainty if the numbers of crop units taken randomly from the sampled commodity are larger than the minimum sample size specified in the Codex Sampling Procedure (8).

It is emphasized that all data except cherry, cucumber, and grape (5 data sets in 76) were obtained with medium-size crops, for which the calculated values can be considered the best possible estimate of sampling uncertainty. The number of degrees of freedom of the estimated sampling uncertainty, df, is 75; thus, it may be considered infinite in further calculations.

The estimated values are valid only for single lots. The sampling uncertainty can be much higher and unpredictable in the case of consignments consisting of several lots.

The uncertainty of sampling small crops will be determined by the number of primary sample positions, and the typical CV_S estimated for the medium-size crops will probably be

Table 7. Expected CV[a] values at 95 and 99% probabilities, β_p, and confidence levels, β_t, for residues in samples taken from 1 field (lot)

n	Average	$\beta_p = 0.95$, $\beta_t = 0.95$		$\beta_p = 0.99$, $\beta_t = 0.99$	
		Minimum	Maximum	Minimum	Maximum
1	0.81	0.24	1.38	0.01	1.61
5	0.37	0.11	0.62	0.01	0.72
10	0.25	0.07	0.43	0.00	0.61
25	0.16	0.05	0.28	0.00	0.52

[a] CV = Coefficient of variation.

Table 8. Mass range of 10 crop units, the number of fruits required to satisfy the 1 kg minimum mass requirement of the Codex Sampling Procedure and the corresponding sampling uncertainty

	Mass of 10 crop units, g[a]					
		Range				
	Average	Minimum	Maximum	No. of crop units/kg[b]		CV_S[c]
			Apple			
Minimum	1248	754	1796	8.0	13.3	0.22
Maximum	1445	887	1961	6.9	11.3	
Mean	1348	840	1869	7.4	11.9	
			Banana			
Minimum	1140	1120	1161	8.8	8.9	
Maximum	1547	1401	1692	6.5	7.1	
Mean	1338	1261	1415	7.5	7.9	
			Cucumber			
Minimum	2299	2299	2299	4.3	4.3	
Maximum	4760	4760	4760	2.1	2.1	
Mean	3516	3516	3516	2.8	2.8	
			Grape			
Minimum	1405	1338	1518	7.1	7.5	
Maximum	3477	3474	3482	2.9	2.9	
Mean	2362	2340	2378	4.2	4.3	
			Kiwi			
Minimum	768	707	790	13.0	14.1	0.22
Maximum	904	791	980	11.1	12.6	
Mean	834	755	863	12.0	13.2	
			Orange			
Minimum	1723	1435	2197	5.8	7.0	
Maximum	1965	1617	2498	5.1	6.2	
Mean	1844	1539	2321	5.4	6.5	
			Peach			
Minimum	788	729	900	12.7	13.7	0.22
Maximum	991	913	1144	10.1	11.0	
Mean	889	823	1021	11.2	12.2	
			Pear			
Minimum	1545	1318	1759	6.5	7.6	
Maximum	1937	1744	2237	5.2	5.7	
Mean	1745	1529	2007	5.7	6.5	
			Plum			
Minimum	525	383	786	19.0	26.1	0.16
Maximum	633	462	968	15.8	21.6	
Mean	579	422	871	17.3	23.7	

Table 8. (*continued*)

Mass of 10 crop units, g[a]						
		Range				
	Average	Minimum	Maximum	No. of crop units/kg[b]		CV_S[c]
			Potato			
Minimum	948	262	1429	10.5	38.2	0.13
Maximum	2055	571	3359	4.9	17.5	
Mean	1470	383	2427	6.8	26.1	
			Tomato			
Minimum	430	379	480	23.3	26.4	0.16
Maximum	538	455	620	18.6	22.0	
Mean	478	419	536	20.9	23.9	

[a] The average and range of minimum, maximum, and mean masses of 10 units obtained from the 300 random composite samples generated from the residues measured in the crop units.

[b] Number of crop units required to obtain a minimum of 1 kg composite sample based on the average and minimum mass of 10 units.

[c] Sampling uncertainty corresponding to the largest number of crop units if they are taken randomly.

applicable. However, for large crops (sample size 5), the estimated value can be considered only as temporary because only cucumber and grape represented this group.

No extrapolation can be made for other types of crops, such as leafy vegetables and cereal grains, because no experimental data have been available for consideration. Further field trials are required to generate data for those crops.

Acknowledgments

The model calculations and statistical tests were performed by Mariana Schweikert and Philipp Klaus. Their valuable assistance is sincerely appreciated. The authors are grateful to Caroline Harris and the Pesticide Safety Directorate of the United Kingdom, Patrick Holland, New Zealand, Derek Gale, Dow Chemical, and P. Rao, Rhone-Poulenc Agro, for providing the residue data measured in unit crops, which made the estimation of the typical uncertainty of sampling possible.

References

(1) Ambrus, Á. (1979) in *Pesticide Residues*, H. Frehse & H. Geissbühler (Eds), Pergamon Press, Oxford, UK, pp 6–18

(2) Lykken, L. (1963) in *Residue Reviews*, Vol. 3, F.A. Gunther & J.D. Gunther (Eds), Springer-Verlag, Berlin, Germany, pp 19–34

(3) Snedecor, G.W., & Cochran, W.G. (1980) *Statistical Methods*, 7th Ed., The Iowa State University Press, Ames, IA

(4) Ambrus, Á. (2000) *Food Addit. Contam.* **17,** 519–537

(5) Harris, C. (2000) *Food Addit. Contam.* **17,** 491–495

(6) Food Agriculture Organization (FAO) (1997) *Pesticide Residues in Food: Evaluations 1996*, FAO, Rome, Italy, pp 485–486

(7) Ambrus, Á., & Lantos, J. (2002) *J. Agric. Food Chem.* **50,** 4846–4851

(8) Codex Secretariat (2003) *Recommended Method of Sampling for the Determination of Pesticide Residues for Compliance with MRLs*, ftp://ftp.fao.org/codex/standard/en/cxg_033e.pdf

(9) Horwitz, W. (2000) in *Principles and Practices of Method Validation,* A. Fajgelj & Á. Ambrus (Eds), Royal Society of Chemistry, Cambridge, UK, pp 1–9

(10) Ambrus, Á. (1996) *J. Environ. Sci. Health B* **31,** 435–442

(11) Ambrus, Á. (1999) in *Pesticide Chemistry and Bioscience: Food and Environment Challenge,* G.T. Brooks & T.R. Roberts (Eds), Royal Society of Chemistry, Cambridge, UK, pp 339–350