Food Additives and Contaminants, 2000, Vol. 17, No. 7, 539–546

Residue variability and sampling—practical problems and consequences for residues monitoring

A. R. C. Hill*
Central Science Laboratory, Sand Hutton, York, YO41 1LZ, UK

(Received 26 November 1999; accepted 1 February 2000)

Data generated in the UK have indicated that pesticide residue levels can be highly variable between the individual fruit or vegetables from the same crop or lot in trade. Statistical experiments with these data showed that residues in composite samples, taken according to Codex recommendations, are unlikely to differ by more than a factor of 3–4 from the mean level in the lot. This was corroborated by results obtained from real composite samples. Many fruit or vegetables in trade are mixed after harvest to form combined lots. Analysis of composite samples provides a good indication of average residues but, where the lot has been mixed, such average values are potentially misleading. Residues monitoring is the best means available for general control of pesticide use and consumer exposure, but new strategies for sampling and analysis are required to address the combined effects of residues variability and mixing of lots.

Keywords : pesticide residues, residues variability, residues distribution, sampling, residues monitoring

Introduction

It has long been known that variability in the levels of pesticide residues in raw agricultural products is inevitable. A wide range of levels occurs in practice and, in addition to consumer implications, the implications for residues monitoring and enforcement actions must be considered. Maximum residue limits (MRLs) for plant materials and certain animal products are based on results from composite samples obtained from residue trials. The composite samples are intended to be representative of a population treated according to good agricultural practice (GAP). Testing for compliance with MRLs requires analogous samples to be taken from a lot in trade.

A lot is defined as the whole or part of a consignment that can be considered, from its identification markings, to have uniform characteristics.

Discernibly 'uniform characteristics' do not guarantee that a lot has been subjected to a single pesticide treatment regime, even if from a single producer. A grower may vary treatment regimes between parallel crops, for example to avoid development of pesticide resistance or to deal with a localized pest outbreak. Even with a single treatment regime, 'uniform characteristics' do not imply that residues are homogeneously distributed through the lot.

A heterogeneous distribution (or 'variability') of residues *within* each unit (i.e. a single fruit or vegetable) is usual and often intended, because it can maximize pest control and minimize residues. In contrast, variability *between* units is usually undesirable, because of adverse effects on pest control and the consequential need to apply more than the minimum amount of pesticide to achieve good control.

Residues monitoring, intended to help ensure consumer safety and good practice in pesticide use, must somehow take into account variability and this poses challenges to analysts and regulators alike. In order to respond logically to these challenges, some knowledge of the variability occurring in practice is required.

Background to the variability data generated in the UK

The present UK studies of unit-to-unit variability of residues originated almost unintentionally in 1991. Disregarding the specific problem of the UK MRL

* e-mail: alan.hill@csl.gov.uk

Table 1. Distribution of triazophos residues in carrots: 10-root samples from a single row in a commercial crop.

Sample	Part of root	% of root mass	Triazophos (mg/kg)	% of residue mass
A	Top 2–3 mm	2	9.1	48
	Next 1 cm	8	1.0	27
	Lower part	90	0.08	25
B	Top 2–3 mm	2	3.5	57
	Next 1 cm	10	0.3	26
	Lower part	88	0.02	17
C	Top 2–3 mm	2	6.1	56
	Next 1 cm	6	0.9	27
	Lower part	92	0.04	17
D	Top 2–3 mm	2	8.0	59
	Next 1 cm	7	1.1	26
	Lower part	91	0.05	15

Note: A and B root tops above ground, C and D root tops below ground.

for triazophos which did not reflect GAP at that time, there was concern over why certain carrot crops exceeded MRLs, even though the evidence available indicated that growers had adhered to GAP. The samples had been obtained directly from growing crops at harvest. No cultural, climatic, varietal, treatment or producer factors were connected with the variations observed between crops. Indeed, two samples taken from notionally identical, adjacent crops might contain either similar, or greatly differing, residue levels. Non-systemic pesticides were the main problem in carrots and, because these tend to be located mainly in the crowns, it was originally thought that inconsistent removal of carrot tops could have been responsible (Hill *et al.* 1991, PR 819, unpublished). However, as indicated in table 1, although this might lead to a difference of a factor of two or so in residue levels, it was not sufficient to explain the very large differences found between crops. The real origins of the problem emerged after we were unable to produce consistent data from experiments on the effects of domestic processing (table 2). Initial suspicions that the inconsistencies were due to bias in the selection of carrot sizes, before and after processing, were disproven when the individual roots were analysed (table 3) (Hill *et al.* 1993, PR 934, unpublished). The unit-to-unit variation in residue levels was sufficient to explain the conflicting data from both the laboratory experiments and the field crops. The ranges of residue levels (maximum/mean or maximum/minimum) in the nine-carrot samples seem unrelated to the agreement, or otherwise, between the corresponding samples of nine and ten carrots.

The data in table 3 suggested that unforeseen risks might arise for a small proportion of consumers, in those cases where a commodity treated with acutely toxic pesticides could be consumed as a single large unit in one meal.

Data have since been generated from large samples of apples, bananas, celery, kiwi fruit, nectarines, oranges, peaches, pears, plums, potatoes and tomatoes. The samples have included produce from various parts of the world, in addition to UK products. Unlike the carrots, samples of these other commodities were obtained from retail or wholesale sources. However, care was taken to ensure that all units came from one or more cartons bearing identical marks of origin, grower codes, etc.

Residues in individual units were determined only where analysis of a 10-unit composite sample revealed the presence of organophosphorus or carbamate pesticides. In many cases, even the highest residue levels found were low.

Attention was focused on acutely toxic pesticides because residue variability has little or no influence on safety margins for other pesticides. Residue levels of all pesticides, in all products, were found to be variable (Hill *et al.* 1996, FD95/161, unpublished, Reynolds *et al.* 1998, FD98/16, unpublished). The frequency distributions of organophosphorus and carbamate pesticide residues did not appear to be

Table 2. Effects of domestic processing on residues (mg/kg) in samples of carrots from seven commercial crops, 1993.

Pesticide	Processing	Residues (mg/kg) in sample						
		A	B	C	D	E	F	G
Chlorfenvinphos	None	0.04			0.06	0.07	0.17	0.33
	Boiling	0.05			0.006	0.02	0.31	0.47
	Microwave	0.05			0.06	0.21	0.31	0.21
	Stir-frying	0.03			0.05	0.02	0.06	0.51
Pirimiphos-methyl	None			0.41				
	Boiling			0.56				
	Microwave			0.40				
	Stir-frying			1.27				
Quinalphos	None		< 0.003				0.05	0.21
	Boiling		0.04				0.26	0.41
	Microwave		< 0.003				0.14	0.05
	Stir-frying		< 0.003				0.13	0.25
Triazophos	None				0.28	0.05		
	Boiling				< 0.01	0.04		
	Microwave				< 0.01	0.12		
	Stir-frying				0.18	0.02		

Note: data corrected for change in weight on cooking.

Table 3. Residues in carrots of different sizes: residues in individual roots and composite samples from seven commercial crops, 1993.

No.	Root Size	Residues (mg/kg) in sample										
		A CHL	B QUI	C PIR	D CHL	D TRI	E CHL	E TRI	F CHL	F QUI	G CHL	G QUI
1	Small	0.02	0.25	1.43	0.51	0.14	0.06	0.01	0.44	0.21	0.24	< 0.0032
2	Small	0.02	< 0.003	7.53	0.91	0.14	0.01	< 0.01	0.73	0.10	0.81	0.26
3	Small	0.10	0.24	0.15	0.07	< 0.01	0.20	< 0.01	0.26	0.45		
4	Medium	0.04	0.08	0.36	0.06	< 0.01	0.04	< 0.01	0.49	< 0.003	0.79	0.83
5	Medium	0.03	0.02	2.54	< 0.005	< 0.01	0.66	0.40	0.59	0.19	0.18	< 0.003
6	Medium	0.12	< 0.003	2.49	0.49	< 0.01	0.13	1.88	0.36	0.26	0.54	0.33
7	Large	0.03	0.10	0.12	0.15	0.01	0.12	< 0.01	0.17	0.14	0.61	0.11
8	Large	0.13	0.04	1.48	0.11	< 0.01	0.53	2.28	0.29	0.31	0.64	0.47
9	Large	0.09	0.02	0.08	0.02	0.05	0.37	0.12	0.38	0.17	0.66	2.30
Mean (weighted)		0.08	0.06	1.05	0.18	0.03	0.26	0.60	0.36	0.19	0.55	0.74
Real composite		0.13	0.06	0.86	0.13	0.25	0.05	0.04	0.34	0.28	0.36	0.19
	Max/mean	2	4	7	5	5	2	3	2	2	1	3
	Max./min.	7	> 83	63	> 182	> 14	66	> 228	4	> 150	5	> 767

CHL—chlorfenvinphos, PIR—pirimiphos-methyl, QUI—quinalphos, TRI—triazophos.
Note: The 'real composites' were separate 10-carrot sub-samples, taken from the same sample as the nine individual carrots.

characteristically different from those of other types of pesticides, although few data sets were generated for the latter. Some crops, notably celery, were associated with lower variability than others, such as the top fruit. This may reflect the relative ease with which pesticides can be applied uniformly to a low-growing crop but it could also be an artefact of the rather small number of data sets.

About 12 000 individual unit data have been generated, encompassing many different pesticides. Many of the data relate to multiple residues.

Issues arising from data generated in the UK

Analytical effects

Recovery experiments were carried out routinely in all analysis batches and, in most cases, showed that analytical precision did not contribute significantly to the variance found in the units. The exceptions were in the few cases where the variability was particularly low, as in certain celery samples. Otherwise, the variability found was primarily due to the variability of residues between units.

Distributions of residues

The data on variability in carrots were derived from 10 units only and are not suitable for estimating frequency distributions. Data for the other commodities were derived from 100-unit samples, providing a good indication of frequency distributions.

There was no correlation between unit mass and residue level, in any commodity. Figure 1, showing carbaryl in apples, is typical. The lack of a correlation—even for pesticides applied post-harvest—indicates that unit size in these commodities is not an important determinant of pesticide deposition or retention.

Over many years, workers have observed that the frequency distribution of residue levels in treated plants or crops is usually positively skewed, often approximating a log-normal distribution. This kind of distribution was frequently observed in the present studies, including the residues of post-harvest pesticides. Figure 2 shows the frequency distribution of thiabendazole in a sample of apples. The few other data sets obtained for post-harvest treatment agents also showed more or less a log-normal frequency distribution. Superficially, a distribution nearer normal might be expected with post-harvest applications. Inadequate mixing (or depletion) of active ingredient in the tank, or higher deposition rates on surface layers in fogging (or on 'high points' with electrostatic applications) could produce the skewed distribution.

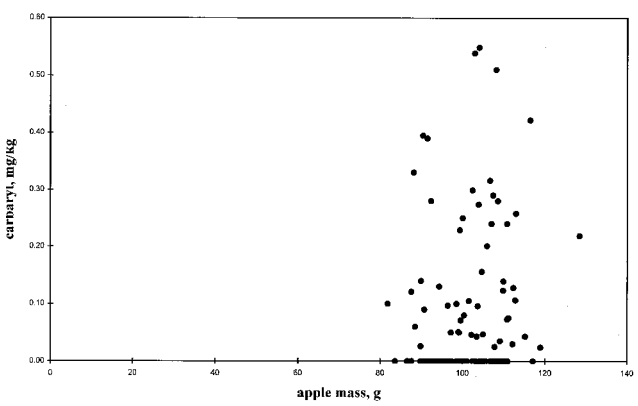

Figure 1. Unit mass and residue level: carbaryl in a 110-unit sample of apples.

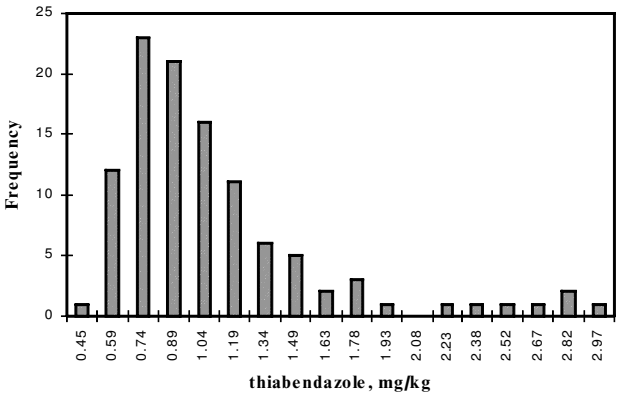

Figure 2. Frequency distribution of thiabendazole in the indivudal apples of a 108-unit sample.

Sample size experiments

As described by Ambrus (2000), statistical sampling experiments were carried where a majority of the units contained measurable residues. In each case, 20 random 'samples' of 5, 10 and 20 units were withdrawn (with replacement) from the 100-unit data sets. The residues in the composite samples obtained were adjusted for the mass of the units selected. An example is shown in figure 3. Predictably, the 5-, 10-, and 20-unit samples had a frequency distribution of residue levels that tended to approach normal, irrespective of the original frequency distribution of the individual units. A summary of these sampling data for apples, bananas, nectarines, oranges, peaches, pears and tomatoes is presented in table 4. Inevitably the 5-unit samples were associated with a considerably greater spread of results than the 10-unit

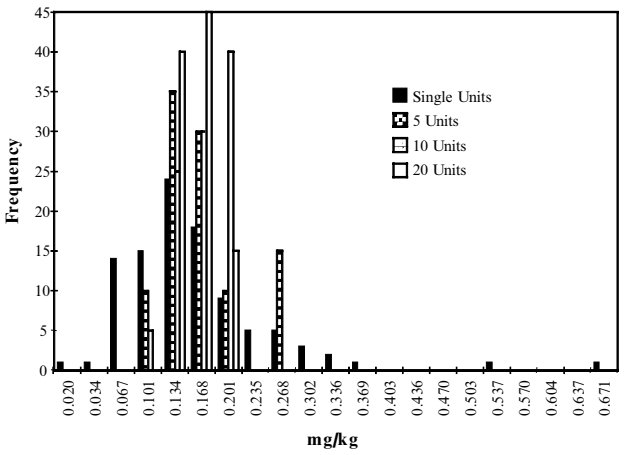

Figure 3. Frequency distributions of aldicarb residues in samples of 1, 5, 10 and 20 potatoes withdrawn statistically at random (with replacement), utilising the data from a 100-unit sample.

samples, but the difference in spread between 10- and 20-unit samples was small.

In the case of celery (where only 40 heads were analysed in each case and where residue levels were rather uniform), 5-unit samples provided a spread of results (table 5) roughly comparable to the 10-unit samples of fruit. It is not known if the relatively uniform levels of residues between heads are a general characteristic of celery crops. Most pesticides tended to be at highest levels in the outer sticks (petioles) of the celery heads, so it might be expected that residues would be made more variable by the vagaries of trimming for market. As noted for carrots above, the influence of trimming on variability between units in celery appears to be small.

The results support current procedures recommended by the Codex Alimentarius Commission (1993). These require a minimum of 10 units or 1 kg (whichever is the greater) where the unit weight is < 250 g, or a

Table 4. Sample size and variation in residues in apples, bananas, nectarines, organes, peaches, pears, tomoatoes (40 data sets).

	Sample size (units)				'Real'[b]
	1	5[a]	10[a]	20[a]	10-units
Mean level, ratio to 100-unit mean	1.0	1.0	1.0	1.0	1.2
% CV	104	48	34	23	59
Highest level, ratio to 100-unit mean	11.1	3.5	2.4	2.5	4.0
Lowest level, ratio to 100-unit mean	< 0.01	< 0.01	0.3	0.4	0.4
Highest max./min. result ratio for any batch	432	110	8	6	–
Number of data points	4090	800	800	800	40

[a] Data from 20 random selections, made with replacement, taken from each of forty 100-unit data sets.
[b] Data from screening, i.e. real 10-unit composite samples.
All values are relative to the weighted mean level (mg/kg) calculated for each 100-unit batch.

Table 5. Sample size and variation in residues in celery (10 data sets).

	Sample size (units)			'Real'[b]
	1	5[a]	10[a]	5-units
Mean level, ratio to 40-unit mean	1.0	1.0	1.0	1.3
% CV	75	31	24	57
Highest level, ratio to 40-unit mean	5.6	2.3	2.1	2.4
Lowest level, ratio to 40-unit mean	< 0.05	0.08	0.1	0.6
Highest max./min. result ratio for any batch	66	26	13	–
Number of data points	400	200	200	10

[a] Data from 20 random selections, made with replacement, taken from each of ten 40-unit data sets.
[b] Data from screening, i.e. real 5-unit composite samples.
All values are relative to the weighted mean level (mg/kg) calculated for each 40-unit batch.

minimum of five units or 2 kg (whichever is the greater) where the unit weight exceeds 250 g. Nearly all results were within a factor of 3–4 of the mean residue level. As an approximation, there was a 95% probability of providing a sample with a residue level within about ± 70% of the mean. These statistical results were corroborated by those obtained from the independent 5-unit samples of celery and 10-unit samples of the other commodities, taken from the same large samples.

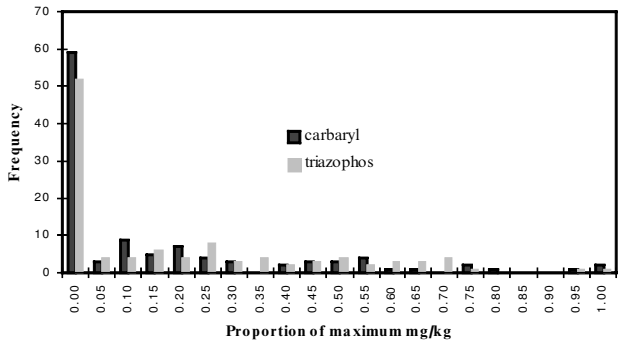

Figure 4. Frequency distributions of carbaryl and triazophos residues in a 110-unit sample of apples.
Note: units labelled as zero residue levels contained < 0.01 mg carbaryl/kg, < 0.05 mg triazophos/kg. The maximum levels were 0.55 mg carbaryl/kg and 1.75 mg triazophos/kg.

Complex frequency distributions

In approximately half of the pesticide/product combinations studied, the frequency distributions were not simple. In some cases, this may have been an artefact produced by the low level of pesticide present and a corresponding inability to detect the residues in a large number of the units. In many cases, the distribution was real. In a simple example, the frequency distributions of carbaryl and triazophos in a sample of apples (figure 4) appear to be bimodal. When these data are ranked according to carbaryl and then triazophos levels (figure 5), it is obvious that this sample contains apples subjected to at least two different field treatments. Only two fruit contained both pesticides: it is not known if they were carbaryl-treated fruit that became contaminated with triazophos from adjacent fruit in packing and storage, or if they represented a third field treatment. Similarly, the three fruit that contained neither compound could represent any or none of these treatments. Chlorpyrifos was also present in a large proportion of these fruits, with an apparently log-normal frequency distribution, but the levels showed no correlation with either of the other two pesticides. All three pesticides must surely have been applied in the orchard. The chlorpyrifos may have been a single treatment applied to two groups of apple trees—one treated with carbaryl and the other with triazophos. Alternatively, its frequency distribution may an artefact of mixing two completely unrelated treatments.

In some cases, the picture was more complex, suggesting the mixing of several different populations of fruit. In figure 6, the residues data for a 100-unit sample of oranges are ranked according to the levels of dicofol, ethion, mecarbam and dimethoate. These oranges were clearly a mixed population, with no fruit containing both ethion and mecarbam. However, the

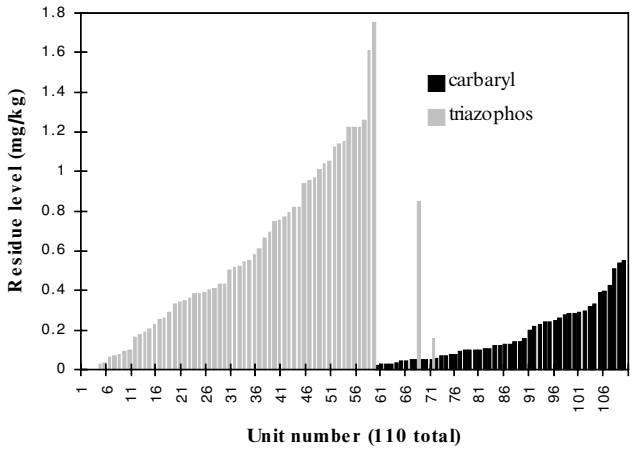

Figure 5. Carbaryl and triazophos residues in apples (same sample as Figure 4), arrnaged in order of the residue levels of carbaryl and triazophos present in each fruit.

residue distribution picture was more complex than can be shown in figure 6, because the fruit also contained residues of chlorpyrifos, omethoate (from dimethoate), imazalil, malathion and tetradifon. Mixing becomes more evident the more residues there are present in the lot and, in some samples, it appeared that a single post-harvest treatment had been applied to units from very different field treatments.

The multiple origins (or pesticide treatment histories) of the units in these samples would not be evident if only a single pesticide had been detected, nor if a similar treatment regime had been used for all fruit.

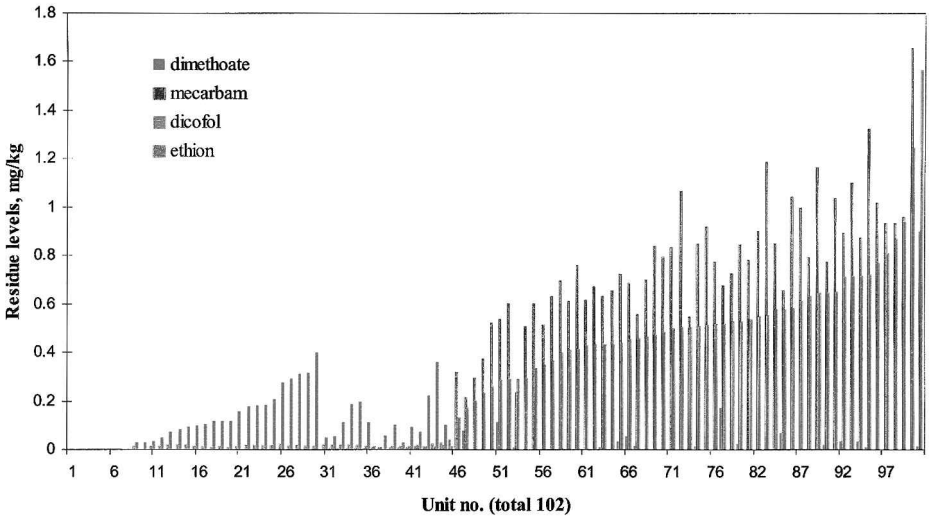

Figure 6. Ranked distribution of dicofol, dimethoate, ethion and mecarbam in a 100-unit sample of oranges.

Table 6. Extreme results from analysis of large samples.

Sample	Pesticide	Residue in 10-unit composite (mg/kg)	Maximum residue in a unit (mg/kg)	Proportion of units containing residues
Apple A	Propargite	0.20	0.06	2/100
Apple B	Chlorpyrifos	0.006	0.083	6/100
Peach A	Ethiofencarb	< 0.05	1.17	9/105
Pear A	Carbaryl	0.04	0.28	6/102
Plum A	Pirimicarb	0.02	< 0.01	0/100
Plum B	Chlorpyrifos	0.01	0.006	1/100
Plum C	Chlorpyrifos	0.07	0.69	2/100

Although the tests described under (c), above, support existing Codex sampling recommendations, there are some limitations. The most significant limitation is that the tests did not include data sets where only a minority of the units contained measurable residues. In some of pesticide/product combinations studied, a pesticide detected in the composite sample was found in none (or very few) of the 100 individual units, or *vice versa*. In some cases, the maximum found in the 100 units was too low to account for the residue found in the 10-unit composite sample. Some examples where less than 10% of units contained residues are presented in table 6. Some of the results show, or imply, relatively high residue levels in one or two units. Of course, we do not know if these 'infrequent' units represented unusual, high extremes from a single treatment, or if they were part of another treatment population that was mainly distributed elsewhere in the lot. Where only 1–10% of units contain residues, the probability of a sample containing two such units is 0.01–1.0%. Thus nearly all samples of 10 units would contain either no units containing a residue, or just one. This is borne out by the data from the real composite samples.

Implications for interpretation of monitoring data

Firstly, compliance with MRLs can be tested satisfactorily where the entire lot has been subjected to the same treatment. Examples in trade might be post-harvest treatments used for storage and shipping, or fruits and vegetables (e.g. certain lettuce, fresh raspberries) that may have to be packed at the point of harvest and cannot be mixed subsequently.

Secondly, where a lot combines several pesticide treatment regimes, composite samples are unlikely to show whether or not the producer adhered to good agricultural practice. Compliance with MRLs may therefore be misleading. Multiple misuse of pesticides could be obscured by mixing lots and even gross misuse might go undetected if many lots were to be mixed. In practice, mixing is much more likely to result from the desire to provide large lots of uniform grade and quality. Where mixing is likely, samples should be taken at point of harvest (or withdrawal from store for post-harvest treatments) to test for compliance with MRLs.

Thirdly, residues of some pesticides will be missed altogether in mixed lots, due to dilution.

Fourthly, the occurrence of multiple residues may be over-estimated in mixed lots, in that there may be fewer pesticides in any one unit than in the composite sample.

Fifthly, although the residues in composite samples are diluted and are more representative of average consumer exposure to residues, the maximum residue levels per unit remain unaffected by mixing of lots. Thus the hazard presented by a single large unit containing a high residue of an acutely toxic pesticide is not affected by mixing, although the probability of detecting its presence is diminished.

Conclusions

Monitoring and enforcement analyses are the only viable means for controlling residues. They provide an incentive for adherence to good agricultural practice and an incentive to discontinue using the more persistent or hazardous pesticides.

Where there is no mixing of lots, present day monitoring data provide a sound means for controlling the use of pesticides. However, in many cases, we do not know if mixing has occurred. For commodities where mixing of lots is frequent (and this includes products from different treatments produced by a single grower), new strategies are required to test compliance with GAP. With adequate mixing of sufficient lots by packers, the probability of detecting gross misuse of pesticides could be made very low. We have no evidence that this occurs deliberately, but we do not have a lot of evidence to the contrary either. Where average consumer exposure to residues is sought from monitoring data, present approaches may be acceptable, but few monitoring studies are large enough to enable such averages to be computed with any certainty.

Currently, these problems are not easy to resolve. Taking samples at the point of harvest, or at packhouses prior to mixing, is costly. As an alternative, 'high' residues detected in composite samples could be followed-up by analysis of individual units. Apart from greatly increasing costs, this risks the loss of perishable commodities in trade and it may be difficult to interpret the data.

In the longer term, new strategies are required for sampling and analysis for monitoring. Pesticides will be essential in feeding the world for the foreseeable future, and monitoring will have an important part to play in controlling residues. What is needed are cheap, rapid and minimally disruptive sampling techniques for use in the field, coupled with equally cheap, rapid and small-scale analytical testing on-site. When these become available, the residues profile of a lot might be assessed without transporting large samples to the laboratory, followed by lengthy analysis. This may seem an impossible dream at present but it could start to become practical within a few years, if some of the current developments in analytical miniaturisation can be harnessed for residues monitoring.

Acknowledgements

Funding was provided through the UK Pesticides Safety Directorate and the staff of its Pesticide Chemistry Branch provided advice and support throughout. Careful and dedicated analysts, too numerous to name individually, contributed to this work at CSL. The studies of variability in celery, kiwifruit, plums and potatoes were managed by my colleague, Mr S. L. Reynolds.

References

AMBRUS, A., 2000, Within and between field variability of residue data and sampling implications. *Food Additives and Contaminants* (in press).

CODEX ALIMENTARIUS COMMISSION, 1993, Recommended methods of sampling for the determination of pesticide residues. *Codex Alimentarius*, volume 2 (Rome: FAO), pp. 369–386.