Guidelines for in-house validation of analytical methods for pesticide residues in food and animal feeds†‡

A. R. C. Hill* and S. L. Reynolds

Pesticides Group, Central Science Laboratory, Sand Hutton, York, UK YO41 1LZ

Received 21st January 1999, Accepted 30th March 1999

Criteria are presented by which analytical methods may be judged to have been validated for the determination of pesticide residues. All stages of analysis are addressed, from initial preparation of samples to the production of results, but with a focus on simplicity and cost-effectiveness of the requirements. Criteria are provided for both quantitative and qualitative (screening) methods and they may be applied to single- or multi-residue methods.

Introduction

Valid analytical data are essential for satisfactory monitoring and control of pesticide residues. The analyst must generate information to show that a method intended for these purposes is capable of providing adequate specificity, accuracy and precision, at relevant analyte concentrations and in appropriate matrices. This information is known collectively as validation and provides the basic evidence to support the validity of the results subsequently generated using the method. In practice, validation of a method cannot encompass the whole range of analytical variables encountered in its use. Accordingly, performance validation data1–4 (a sub-set of method validation data and often referred to as internal/analytical quality control data) will be required, to provide evidence of the on-going performance of the method and analyst. Performance validation also provides a continuing check on the effects of minor modifications and on method transfer between analysts.

Concepts of method validation continue to evolve and are currently under consideration, for food analysis, in the European Community and by the Codex Alimentarius Commission. Comprehensive overviews of validation requirements have been published,5–10 which identify many parameters by which method performance may be judged. Other authors11,12 have contributed specific statistical and computational techniques to assist with the process of method validation. However, although the parameters to be assessed are clearly defined, few criteria (e.g., specified limits for accuracy or precision) are provided to define the acceptability of a method. In part, this may be because acceptability is determined by the purpose served by the method and thus a broad overview of validation cannot address the differing requirements of each specific area of analysis. This paper specifically addresses pesticide residues analysis but we suggest our proposals may have value for some other trace analyses.

Validation of methods by inter-laboratory study has become impractical in most cases. Even where it is practical, it is usually impossible to validate all combinations of analyte, analyte concentration and sample matrix to which the method may be applied. Published methods may be supported by validation data but the information is usually limited in scope and, in most cases, further in-house validation data will have to be generated. Method validation, whether in-house or inter-laboratory, has rarely incorporated rigorous investigation of sample processing, extraction efficiency or specificity. These omissions are serious, as the procedures may have a profound influence on the validity of the results obtained.

Users of methods or analytical procedures (see Glossary) should have ready access to validation data, to ensure they do not unknowingly exceed the boundaries of validation. Knowingly exceeding the boundaries of validation may be undesirable but, if unavoidable, the analyst must provide this information with the results.

In a perfect world, all analytical methods would provide quantitative results of negligible uncertainty and at low cost. In reality, cost and practicality mean that some degree of compromise must be accepted. Methods are often loosely termed ‘quantitative’, ‘semi-quantitative’ or ‘qualitative’, depending on the accuracy and precision achievable. The looseness of these terms is useful and we propose that methods should be allocated an appropriate status on the basis of the method validation criteria satisfied. Other authors have made somewhat similar distinctions based on the limits of quantification and detection, estimated at method validation, but these limits may give a false impression of what is achievable routinely, as they may vary considerably when the method is in use.

Some analytical procedures are more difficult and expensive to characterise than others, but all validation is costly. Within a single laboratory it is likely that, over a period of time, certain reagents, equipment, and so on, will be changed from those used at method validation. It is impracticable and probably unnecessary to re-validate a method to take account of all minor changes or minor extensions made. We propose that these should be checked through performance validation and that re-validation of methods be limited to major changes or extensions. This flexible approach places considerable responsibility on analysts and laboratory managers in classifying changes as major or minor.

Method validation information may be needed to support accreditation or publication of the method, or to defend results generated from its use, but there is little international agreement on the exact requirements. There is increasing emphasis on international standards as a means of removing trade barriers, and thus increasing reliance on analytical data to determine compliance with the standards. Against this backdrop, we propose minimum requirements for method validation—in terms of the parameters studied and criteria for acceptability—with the aim that the requirements should be simple, rational and affordable. We hope our proposals will help external assessors of methods to decide whether sufficient information has been produced to support the validity and claimed status of the method.

† The opinions expressed in the following article are entirely those of the authors and do not necessarily represent the views of either The Royal Society of Chemistry or the Editor of The Analyst.
‡ © Crown copyright.
Scope of validation

We define an analytical method as the series of procedures used, from receipt of a sample to calculation of the final result. Analysis of reference materials is often said to provide the ideal means for validation but it is important to recognise that this approach may ignore procedures that can have a critical effect on the results.

Not all procedures can be validated. Sample preparation (see Glossary)—as distinct from sample processing (see Glossary)—may be impossible or unnecessary to validate. For example, preparation procedures such as washing or brushing to remove soil, or the removal of sample parts which are not to be analysed, cannot be validated. Similarly, sub-sampling without sample processing (such as taking the outer leaves from cabbages, without comminution and mixing of the whole cabbages) may be required for certain purposes but the procedure is unlikely to be validated as part of a quantitative method. Certain calculations, such as uncertainty or correction procedures is unlikely to be validated as part of a quantitative method. Certain calculations, such as uncertainty or correction may not be validated and the validation approach may ignore procedures that can have a critical effect on the results.

Validation requirements

A summary of minimum requirements for method validation is presented in Table 1.

Sample processing and storage

Sample processing, sub-sampling and storage can profoundly influence the results of analysis. These processes may be forgotten in method validation or considered too difficult to validate. They are not impossible to validate and the validation is transferable, with the procedures, to other methods (for the same analytes and matrices). Extrapolation of the validation to similar commodities or chemically similar analytes may be possible but the analyst should take a cautious approach. There are two issues involved in the validation of these procedures, potential loss of the analyte and the homogeneity of its distribution. Where homogeneity has been demonstrated for any group of physically similar commodities (see Extension to new sample matrices, below), the data can be considered to support sub-sampling of any other commodity in the group. Sample storage can be validated by analysis of reference materials, if these are in a similar physical state to those of ‘real’ samples, whereas processing and homogeneity cannot. The storage conditions tested should include the maximum likely storage period and temperature.

Checks of analyte loss and homogeneity require an unprocessed sample of each representative commodity, containing trace levels of analyte(s). The trace levels may be incurred through normal food production practices but this may lead to relatively high costs and may require additional apparatus, capable of extracting large analytical samples. The alternative of adding the analyte(s) to whole commodity units (e.g., whole fruit or vegetables), prior to processing or storage, should be less costly and can provide a more stringent test.

To test the homogeneity achieved, initially the analyte must be heterogeneously distributed in the sample. This may be achieved by, for example, processing a sample of 10 apples in which only one contains (or has been dosed with) the analyte. Validation of sample processing with respect to analyte stability (i.e., potential loss) may be performed in various ways. The simplest approach is to measure recovery of a known mass of analyte deposited very carefully on whole commodity units (producing a known nominal concentration) before they are processed. The analyte may also be co-deposited with an internal standard (known to be stable under the conditions of processing) prior to processing and the ratio of analyte to internal standard post-processing may be compared with the ratio before deposition. In this case, it may not be necessary to know the mean concentration prior to processing. These are stringent tests, because superficial deposits may be more prone to evaporation, or adsorption of analyte onto the equipment used, than residues distributed within the commodity. The deposits may resemble residues from post-harvest treatments. Where incurred residues are used for this test, replicate analyses are required before and after processing. The former necessitates extraction from intact commodity units and may require large-scale extraction equipment.

Extraction

Extraction efficiency. With the exception of low moisture products, bound residues of intact analytes do not usually form a major proportion of residues in foods. Nonetheless, validation of extraction efficiency remains important. In cases where extraction is independent of subsequent and previous procedures, changes in the other procedures will not require re-validation of extraction. Rigorous validation of extraction efficiency of organic analytes can only be performed with samples containing analyte(s) incurred by the route by which the trace levels would normally be expected to arise. Suitable certified reference materials containing incurred residues of pesticides are rarely available, so that validation normally requires identification and quantification of radio-labelled analyte(s), metabolites and all other degradation products. This is beyond the capability of most laboratories involved in routine trace-level monitoring and thus extraction is rarely validated. Alternative approaches involve (i) comparison with extraction of samples using a procedure which has previously been validated rigorously; (ii) comparison with extraction of samples
Table 1  Summary of parameters and criteria for in-house validation of the procedures involved in a method

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Parameter</th>
<th>Level(s)</th>
<th>No. analyses required</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quantitative method</td>
</tr>
<tr>
<td>Sample processing</td>
<td>Analyte stability</td>
<td>About 5 × LCL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>≥5 Replicates of each representative commodity, post-processing, and ≥5 pre-processing, if the notional analyte level is not known at that stage</td>
<td>No significant loss of analyte during processing (P = 0.05)</td>
</tr>
<tr>
<td></td>
<td>Analyte homogeneity</td>
<td>About 5 × LCL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>≥5 Replicates of a representative commodity, post-processing. Use stability data, if analyte proves to be stable</td>
<td>RSD ≤15% (not including analytical contribution)</td>
</tr>
<tr>
<td>Sample storage</td>
<td>Analyte stability</td>
<td>About 5 × LCL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>≥5 Replicates at each time point, including time zero</td>
<td>No significant loss of analyte during storage (P = 0.05)</td>
</tr>
<tr>
<td>Extraction</td>
<td>Extraction efficiency</td>
<td>About 5 × LCL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>≥5 Replicates of reference material with incurred residues, or ≥5 replicates by the reference procedure and ≥5 by that under test, using any material with incurred analytes</td>
<td>Mean from test procedure within 95% confidence intervals of consensus value (reference material); or ≥5 by the lower 95% confidence interval of the reference procedure</td>
</tr>
<tr>
<td>Clean-up determination</td>
<td>Specificity of analysis detection</td>
<td>At LCL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ensure measured response is solely due to the analyte. One analysis of each of ≥5 separate blanks of each representative commodity</td>
<td>Identify by mass spectrometry, or most specific technique available</td>
</tr>
<tr>
<td>Calibration and range</td>
<td>LCL&lt;sup&gt;a&lt;/sup&gt; to the required maximum level</td>
<td>≥3 Replicates at ≥3 levels (≥5 levels for non-linear systems), on a minimum of 2 occasions</td>
<td>Sufficiently repeatable response and fit of calibration line to enable accuracy and precision criteria to be achieved</td>
<td>Sufficiently repeatable response and fit of calibration line to enable accuracy and precision criteria to be achieved</td>
</tr>
<tr>
<td>Accuracy and precision</td>
<td>LCL&lt;sup&gt;a&lt;/sup&gt; and accepted limit (see Glossary)</td>
<td>≥5 Replicates at each level for each analyte/representative commodity combination</td>
<td>Mean recovery 7–110% with RSD ≤10%. Reference materials, all results within 99% confidence intervals</td>
<td>Mean recovery 50–120% with RSD ≤25%, or mean recovery 20–150% with RSD ≤10%. Reference materials, mean result within 99% confidence intervals</td>
</tr>
<tr>
<td>Analyte stability in extracts and standard solutions</td>
<td>LCL&lt;sup&gt;a&lt;/sup&gt; and accepted limit (see Glossary)</td>
<td>≥5 Replicates at each appropriate point in time (including zero) and for each representative commodity</td>
<td>No significant change in analyte concentration (P = 0.05)</td>
<td>No significant change in analyte concentration (P = 0.05)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Lowest calibrated level for the intended application.
by a very different extraction technique; for example, using a very different solvent (especially if coupled with a more effective sample disintegration technique); or (iii) analysis of a certified reference (or proficiency test) material containing incurred analyte, where the consensus analyte level has been determined by a number of laboratories using either a rigorously validated extraction technique or several different extraction techniques. Whilst option (iii) is preferred, it is usually impractical because of the lack of availability, or high cost, of suitable materials.

**Analyte stability in standard solutions, sample extracts, etc.** The stability of the analyte in solvent and sample extracts (etc.) should be investigated during method development. However, where analyte stability in solvent solution has been validated in one laboratory, the validation data may be applied to the same analyte/solvent combination elsewhere, subject to strict adherence to the conditions adopted at validation. This investigation contributes to tests for ruggedness and the time intervals chosen will depend on the requirements expected in practice.

**Recovery**

Where the nature of methods allows determination of recovery, it forms an important means to determine the accuracy and precision of transmission of the analyte through the extraction and clean-up (and derivatisation, etc.) procedures. The minimum requirements and acceptance criteria for recovery by quantitative, semi-quantitative and qualitative (screening) methods are given in Table 1.

Recovery should normally be determined at two levels: (i) at or about the lowest calibrated level (LCL, see Glossary and below); and (ii) at or about the accepted limit (see Glossary), where this is significantly higher. Where these two levels are similar (for example, because the accepted limit is set at or about the limit of determination), the second level for recovery may be set at 5–10 times the accepted limit.

**Methods which cannot measure recovery.** There are methods with which it is impossible to determine recovery, because the recovery determination is identical to the calibration procedure. For example, in the direct analysis of liquids, in solid-phase micro-extraction (SPME), or in certain headspace analyses. The inability to determine recovery does not imply that the methods cannot be validated but, unless reference materials are analysed or comparisons made with an alternative method, the accuracy and precision data will refer only to the calibration.

**Detection and determination**

**Specificity.** The detected response must be demonstrated as attributable to the analyte, at levels down to the LCL and preferably by the least equivocal technique available. Specificity usually decreases with the level of analyte concentration and cut-off levels may be identified for differing degrees of specificity. Potentially interfering responses can be controlled by ensuring that the clean-up and determination procedures in the method provide adequate selectivity. Chromatographic separation provides a powerful tool for increasing selectivity when linked to the detection system. Full-scan mass spectrometry (MS) provides a robust approach to high specificity for analytes amenable to the technique but, where few ions (or few characteristic ions) are generated, a combination of MS techniques may be required to provide the supporting evidence.

Potential interference should be estimated by means of analysis of one each of ≥5 separate notionally ‘blank’ samples of each representative commodity (see Glossary). Where the analyte occurs naturally, or where it is otherwise impractical to obtain ‘blank’ material, the background levels should be rigorously confirmed in the same way, comparing the adopted technique against the most specific technique available. In most cases, specificity may be validated where an analyte degradation product (etc.) is detected. However, validation can be extremely difficult where the detected species can be generated from compounds other than the target analyte, or if it is a degradation product of the analyte which is not included in the accepted limit. In these cases, the specificity, and perhaps the accuracy of the method, cannot be validated.

**Matrix effects.** The effect, if any, of co-extractives on the analyte response obtained should be assessed by comparing standards prepared in extracts (‘matrix-matched’) of representative commodities (see Glossary) with those in solvent. This may be assessed by incorporation into the calibration carried out for recovery determination, but it may not be possible to test the effect in methods which cannot determine recovery. The presence or absence of matrix effects should be demonstrated over the concentration range of interest. Such effects may not be consistent and the default should be to use matrix-matched calibration unless it is demonstrated to be unnecessary.

Where matrix effects could occur and ‘blank’ sample material is not available for matrix-matching, calibration should utilise standard addition, with the background level representing the LCL (see below).

**Detection range.** It is not essential to establish the highest and lowest possible levels of analyte that could be determined by the method, and the validated range should be chosen on the basis of intended purpose. The validation must demonstrate that measured responses are within the dynamic range of the detection system. If the limits of the detector dynamic range are approached, the method should incorporate a suitable concentration or dilution procedure. Stability of detection system response, and the consequently required frequency of calibration, is primarily an issue of ruggedness but it should be established from the replicate analyses undertaken for validation. For screening methods, the only requirement is to demonstrate that the presence and absence of analyte can be differentiated at the limit required.

**Adoption of a lowest calibrated level (LCL).** Limits of detection or quantification almost always vary with time, equipment, etc., which tends to make them irreproducible. Thus ‘method validation’ of these parameters may be of little value. The LCL should be decided by the purpose for which the method is to be used. The requirements are to demonstrate that the LCL can be detected consistently and, for quantitative and semi-quantitative methods, that the response is adequately consistent across a batch of determinations. Internal reproducibility will be determined by performance validation. If the analyte response is superimposed on a ‘background’ signal, great caution is required in determining the true LCL and it should normally be at least three times the maximum background level.

**Extending, modifying or transferring methods**

Ideally, a method would be validated before use for every analyte commodity combination that will be analysed but this is unlikely to be practicable. In some cases, further validation of the method will be essential but, in general, the validation may take the form of on-going performance validation. There are no
simple rules for deciding between a requirement for method validation and a requirement that can be met by performance validation. We provide here some outline guidance but the laboratory manager must take responsibility for ensuring that the validation provided is adequate.

**Extension to new sample matrices**

Matrices should be differentiated sufficiently but not unnecessarily. Validation of changes to most procedures can be incorporated into performance validation. Whilst we recognise that they are all too rarely investigated in practice, sample processing and storage may require specific validation, prior to analysing samples of a new commodity. For commodities available in a wide range of manufactured variants, cultivated varieties, or parts, etc., generally (though not invariably) a single variant may be considered to represent approximately similar matrices. However, where a particular variant is known (or suspected) to differ from others in its effects on method performance, it should be the subject of separate validation of the method. Decisions on the limits to method validation may be based on broad categories such as water, sugar, fat, acid, fibre content of the samples and/or on the similarity of the plants or animals involved. Examples of extensions of method validation and their limits are: data for whole grains should not be taken to apply to bran or beer but data for wheat may be applicable to other whole grains; data for muscle should not be taken to apply to fat, offal or eggs but data for chicken fat may be applicable to other animal fats; data for a cooked product should not be taken to apply to the fresh product (although the reverse may be acceptable); data for one brassica vegetable, citrus fruit, etc., may be applicable to similar products from the same group but should not be taken to apply to all vegetables or fruit.

**Extension to new analytes**

The introduction of new analytes may be addressed by performance validation but, again, this does not encompass procedures for sample processing, storage and extraction. Where the physico-chemical characteristics of a new analyte resemble closely those of analytes for which method validation data already exist, these procedures may be utilised without further specific validation. The analyst must be aware that this specific validation is lacking and, if intermittent or persistent problems are experienced with recovery, calibration, extract storage, etc., further method validation must be undertaken before proceeding further. Analytes of widely differing characteristics must each be subjected to full method validation at the outset.

**Extension to other analysts and other laboratories**

Performance validation data should provide adequate support for transfer to other analysts (during training) within a laboratory, but requires strict adherence to the SOPs relating to sample processing, storage and extraction. The many possible differences in equipment, consumable supplies and working practices adopted in different laboratories tend to make procedures peculiar to a laboratory. Where the details of operation differ, or where there is insufficient information to show that procedures are identical, validation should not be transferred between laboratories but may serve as guidance.

**Validation of modifications to methods**

Methods or procedures may be modified intentionally or unintentionally. Even in the most well-regulated laboratories, it is common for methods to be modified with time, through changes in the supply of reagents, equipment, etc. Minor modifications to methods may be addressed through performance validation. Examples of major modifications, likely to require method validation, are as follows.

**Processing.** (i) A marked decrease in the fineness of comminution (or a marked increase in the fineness of comminution of biochemically active samples); (ii) an increase in mean temperature > 10 °C, or an increase in time > 2-fold.

**Extraction.** A much less efficient disintegration process; a change of solvent where the polarity and/or chemistry is altered significantly; a change in temperature sufficient to affect solubility and/or partition; a major decrease in time to < 50%; or a major decrease in solvent to test portion ratio.

**Clean-up.** A significant change in clean-up technique, such as the use of a completely different adsorbent or solvent; changing to a procedure based on a different principle; or omitting the procedure entirely.

**Determination.** A major change in the detection technique, especially if a different detection principle is involved; adoption of a very different chromatographic separation system; or a ≥5-fold reduction in the lowest calibrated level.

**Acknowledgements**

The authors are very grateful for constructive suggestions provided by W. Korth, National Residues Survey, Kingston, Australia; A. Ambrus, IAEA, Vienna, Austria; L. Alder, BGVV, Berlin, Germany; and R. Fussell of the Central Science Laboratory.

---

**Glossary**

- **Accepted limit** A concentration value used as a limit for trading standards, safety, action purposes, etc., whether or not it has any legal status.
- **Analyte** The chemical species to be determined, including any isomers, metabolites, breakdown products or bound forms as defined by the accepted limit, and including any derivatives or degradation products generated for the purpose of analysis.
- **Analytical portion** (Test portion) A portion of the analytical sample taken for analysis. The test portion should be representative of the analytical sample for quantitative and semi-quantitative methods. For qualitative methods (and potentially for some semi-quantitative methods), the test portion may be positively biased (e.g., peel from citrus fruit).
- **Analytical sample** (Test sample) The laboratory sample after sample preparation, if required, and from which the analytical portion is withdrawn. It may or may not be subject to sample processing.
- **Blank** (i) A sample which contains no detectable level of the analyte. (ii) A complete analysis made without the
Commodity group A group of foods or animal feeds sharing sufficient characteristics as to make them similar for the purposes of analysis by a method. The characteristics may be based on major constituents (e.g., water, fat, sugar, acid content) or other compositional relationships.

Laboratory sample The sample as received at the laboratory but not including the packaging.

Lowest calibrated level (LCL) The lowest nominal concentration of analyte detected and measured in calibration of the detection system. The corresponding level in the sample is dependent on the accuracy of the method, at that level, but for pesticide residues analysis the LCL is not corrected for recovery.

Matrix-matched calibration Calibration using standards prepared in an extract of the commodity analysed (or of a representative commodity). The extract concentration must be similar to that of the samples analysed. The objective is to compensate for the effects of co-extractives on the determination system. Such effects are often unpredictable, but matrix-matching may be unnecessary where co-extractives invariably prove to be without effect.

Method The series of procedures from receipt of a sample for analysis through to the production of the final result.

Method validation A set of data indicating performance characteristics of a method, achieved for specified analytes and sample types.

Negative result A result indicating that the analyte is not present at or above the lowest calibrated level.

Performance validation Sets of data generated during the analysis of batches of samples to support the validity of on-going analyses. A sub-set of method validation requirements.

Positive result A result indicating the detected presence of the analyte at or above the lowest calibrated level.

Procedure An individual step, stage or process, manual or automated, undertaken as part of the method of analysis.

Qualitative method (Screening method) A method capable of detecting the presence of an analyte but giving little or no quantitative information, other than the specified level may have been exceeded. If the results are actionable, they should be confirmed using a quantitative method, if possible (see Table 1).

Quantitative method A method capable of producing quantitative results, with good accuracy and precision (see Table 1).

Reference material A certified reference material or an in-house reference material, for which the mean and uncertainty of analyte concentration has been determined.

Reference procedure A procedure previously established as fully effective or efficient. Where this is not available, a reference procedure may be one that, in theory, should be highly efficient and is fundamentally different from that under test.

Representative commodity A single food or feed, used to represent a commodity group for method validation purposes. A commodity may be considered representative on the basis of the objective is to compensate for the effects of co-extractives on the determination of the analyte distribution, prior to removal of the analytical portion.

Sample preparation The procedure, if required, to convert the laboratory sample into the analytical sample, by removal of parts (soil, stones, bones, etc.) not to be included in the analysis.

Sample processing The procedure(s) (e.g., cutting, grinding, mixing) used to make the analytical sample acceptably homogeneous with respect to the analyte distribution, prior to extraction.

Semi-quantitative method A method capable of producing quantitative results with only moderate accuracy and precision (see Table 1). If the results are actionable, the quantities should be confirmed using a quantitative method if possible.

References

7. Validation of Analytical Methods, NMKL Procedure No. 4, Nordic Committee on Food Analysis, FIN-02044 VTT, Finland, 1996.