

# A Modular Approach to the ECVAM Principles on Test Validity

**Thomas Hartung, Susanne Bremer, Silvia Casati, Sandra Coecke, Raffaella Corvi, Salvador Fortaner, Laura Gribaldo, Marlies Halder, Sebastian Hoffmann, Annett Janusch Roi, Pilar Prieto, Enrico Sabbioni, Laurie Scott, Andrew Worth and Valérie Zuang**

*ECVAM, Institute for Health & Consumer Protection, European Commission Joint Research Centre, Ispra, Italy*

**Summary** — The European Centre for the Validation of Alternative Methods (ECVAM) proposes to make the validation process more flexible, while maintaining its high standards. The various aspects of validation are broken down into independent modules, and the information necessary to complete each module is defined. The data required to assess test validity in an independent peer review, not the process, are thus emphasised. Once the information to satisfy all the modules is complete, the test can enter the peer-review process. In this way, the between-laboratory variability and predictive capacity of a test can be assessed independently. Thinking in terms of validity principles will broaden the applicability of the validation process to a variety of tests and procedures, including the generation of new tests, new technologies (for example, genomics, proteomics), computer-based models (for example, quantitative structure–activity relationship models), and expert systems. This proposal also aims to take into account existing information, defining this as retrospective validation, in contrast to a prospective validation study, which has been the predominant approach to date. This will permit the assessment of test validity by completing the missing information via the relevant validation procedure: prospective validation, retrospective validation, catch-up validation, or a combination of these procedures.

**Key words:** *in silico, in vitro, in vivo, modular approach, QSARs, test validity, validation.*

**Address for correspondence:** T. Hartung, ECVAM, Institute for Health & Consumer Protection, European Commission Joint Research Centre, 21020 Ispra (VA), Italy.  
E-mail: [thomas.hartung@cec.eu.int](mailto:thomas.hartung@cec.eu.int)

## Introduction

The validation of alternative tests is the process by which the reliability and relevance of a test are established for a particular purpose (1). Over the last 12 years, the European Centre for the Validation of Alternative Methods (ECVAM), in cooperation with international experts, has set up guidelines for validation, and five stages in the evolution of a new test were defined: *development, prevalidation, validation, peer review* and *regulatory acceptance*. These stages reflect the sequence of steps to be performed for a prospective validation exercise. They are well documented in a number of publications (for example, 2–5). The process proved to be successful and led to the regulatory acceptance of several alternative tests (summarised in 6). Furthermore, these principles can be applied to any type of test, including *in vivo* tests.

Advances in new technologies, such as pattern-based systems (for example, genomics and proteomics), and computer-based models (for example, quantitative structure–activity relationship [QSAR] models) and expert systems, impose the need to adapt the validation process to the generation of new tests and models. In addition, new approaches

are required, to make the best use of existing data, including aspects of scientific relevance, and to validate test strategies. These might include assessments of mechanistic basis in weight-of-evidence approaches. ECVAM therefore proposes to make the validation process more flexible, by breaking down the various steps in validation into independent modules, and defining for each module the information needed for assessing test validity. Once all the modules are satisfied, the suitability of a test to enter the peer-review process is determined. This information could be provided via prospective or retrospective validation, or a combination of both. In many cases, it will be necessary to carry out laboratory work to obtain the information required, whereas, in other cases, existing information might be sufficient. In addition, a previously validated test or model could be re-evaluated with respect to one or more modules at different times, according to technological progress. This approach would avoid the need to completely re-validate tests in cases where previously obtained information on validity was still applicable. However, care must be taken to ensure that a change in one module does not compromise the validity of the other modules. Further details and technical guidance need to be developed

to assist in the implementation of the modular approach.

Irrespective of whether a prospective or a retrospective approach is applied to complete the information for the individual modules, the reliability and relevance of a test ultimately have to be established in accordance with the previously defined validation principles. For this purpose, an assessment of each module will be conducted by the Validation Management Group (VMG), and recommendations will be made to proceed with independent peer review or to acquire additional information. Only when the information for all the modules is deemed by the VMG to be adequate, should the test enter the peer-review process.

Regarding prospective validation, ECVAM proposes to separately address the modules “between-laboratory variability” and “predictive capacity” in specially designed studies, with the aim of reducing the costs and time needed for validation studies. ECVAM is currently undertaking a pilot study to evaluate the feasibility of this approach. With regard to retrospective validation, ECVAM plans to establish principles for making assessments based on a weight-of-evidence approach, as well as technical guidance on how to implement these principles.

In general, the validation of an *in vitro* test or *in silico* model by ECVAM will begin with an assessment of the extent to which the validity principles

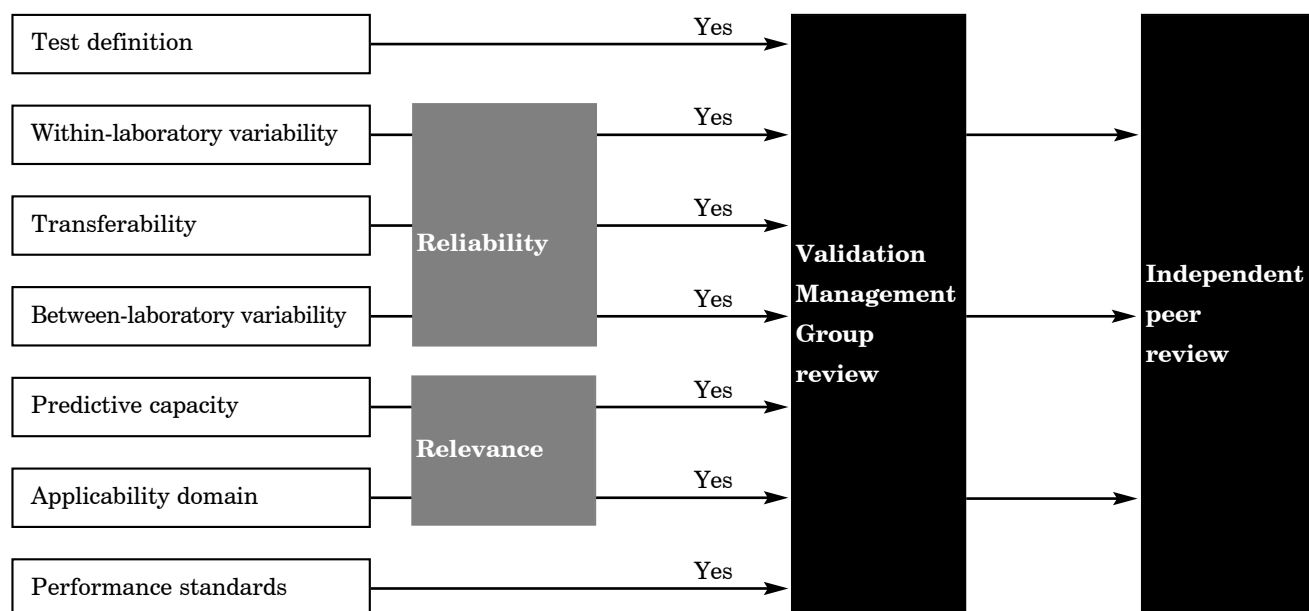
for each module have already been addressed. In appropriate cases, this will be followed by the conduct of the additional experimental and/or statistical analysis needed to satisfy all the validity principles. Thus, in general, a combined approach of retrospective validation and prospective validation will be adopted. This approach should be applicable to different types of tests and models. For example, the application of the validity principles to an experimental test is analogous to their application to an *in silico* model (Table 1).

By adopting this approach, ECVAM will be able to provide a dossier of information to the peer-review panel, European Commission services and regulatory bodies, on the extent to which a given method meets well-defined validity principles. This dossier will help the appropriate body to make a decision on the acceptability of the method within the context of a particular responsibility or regulatory programme.

## The Modules

The seven modules proposed for the validity assessment of a test are defined below, and are schematically depicted in Figure 1. Information should be provided for all of the modules, unless a particular module is not applicable to a specific test/method,

**Figure 1: The modular approach for applying the ECVAM principles on test validity**



A “yes” indicates that the appropriate information for the module is adequate for entrance into the peer-review process. All seven modules have to be satisfactorily completed, as judged by the Validation Management Group, before a method can enter the peer-review process.

ECVAM = European Centre for the Validation of Alternative Methods.

**Table 1: Comparison of information required for each module for experimental tests and *in silico* models to enter the peer-review process**

Experimental test methods (including alternatives)	<i>In silico</i> models (including QSARs)
<p><i>Test (method/model) definition</i>  Test protocol and SOPs</p> <p>Definition of endpoint predicted</p> <p>Training set (chemical structures + complete set of data)</p> <p>Definition of prediction model/data interpretation procedure (developed from training set)</p> <p>(Provisional) domain of applicability</p> <p>Explanation of mechanistic basis (known or putative)</p>	<p>Definition of endpoint predicted</p> <p>Training set (chemical structures + complete set of data)</p> <p>Definition of prediction model (developed from training set)</p> <p>(Provisional) domain of applicability</p> <p>Explanation of mechanistic basis (known or putative)</p>
<p><i>Within-laboratory variability</i>  Assessment of reproducibility of experimental data in same laboratory (different operators, different times)</p>	<p>Within-laboratory assessment not normally necessary, since this source of variability is not normally present</p> <p>For some types of model, which are defined by mathematical routines based on the randomised searching of optimal solutions, different models can sometimes be developed</p> <p>For these types of model, an assessment will need to be made on whether the differences in model definition are biologically significant</p>
<p><i>Transferability</i>  Assessment of reproducibility of experimental data in second laboratory (different operator)</p> <p>Ease of transferability (practicability)</p>	<p>Confirmation of descriptor data by different operator (in the case of descriptors that are themselves estimated)</p> <p>Confirmation of model definition by a different operator</p> <p>Confirmation of goodness-of-fit and robustness of statistics</p> <p>Confirmation of reproducibility of predictions by different operator (same version of model [software] should be used)</p> <p>Assessment of adequateness of documentation on development and application of model</p>
<p><i>Between-laboratory variability</i>  Assessment of reproducibility of experimental data in 2–4 laboratories (different operators, different times)</p>	<p>Between-laboratory assessment not directly relevant, since major source of error is between different operators</p>
<p><i>Predictive capacity</i>  Assessment of predictive capacity of the prediction model associated with the test system (by using test chemicals that were not used in the development of the prediction model)</p>	<p>Indication of predictive capacity by performing suitable cross-validation (usually leave-many-out)</p> <p>Assessment of predictive capacity by performing external validation (i.e. by using test structures that were not used in the development of the model)</p>
<p><i>Applicability domain</i>  Definition of chemical classes and/or ranges of test method endpoints for which the model makes reliable predictions</p>	<p>Definition of chemical classes and/or ranges of molecular descriptors for which the model makes reliable predictions</p>
<p><i>Performance standards</i>  Definition of reference chemicals that can be used to demonstrate the equivalence in performance between a new test and a previously validated test</p>	<p>Definition of reference structures that can be used to demonstrate the equivalence in performance between a new model and a previously validated model</p>

QSAR = quantitative structure–activity relationship; SOP = Standard Operating Procedure.

as could be the case for information on within-laboratory variability for *in silico* methods.

1. Test definition.
2. Within-laboratory variability.
3. Transferability.
4. Between-laboratory variability.
5. Predictive capacity.
6. Applicability domain.
7. Performance standards.

### Test definition

For prospective validation, ECVAM defined a set of criteria which need to be met before a test can enter the prevalidation process (3). This document aims to define in a similar way the information that should be available before a test enters the peer-review process.

An adequate test definition for the purpose of peer review comprises the following information.

1. A definition of the scientific purpose of the test.
2. A description of the mechanistic basis of the test in view of the broader current scientific knowledge of the test endpoint.
3. A definitive protocol compliant with Good Laboratory Practice and Good Cell Culture Practice (7), including: any necessary Standard Operating Procedures (SOPs); specification of endpoints and endpoint measurements; derivation and expression of results; interpretation of the results via a prediction model; and the inclusion of adequate controls.

As a result of information obtained during the validation study or once the test has been fully validated, it may be necessary to redefine the test definition (for example, when there is an agreed modification to the SOP or a refinement of the prediction model is needed). In such cases, the possible consequences for all other modules must be considered by the VMG. In any event, during the assessment of the predictive capacity of a test, the test definition should not be changed.

### Within-laboratory variability

Within-laboratory variability will typically follow the International Conference on Harmonisation procedures (8, 9), addressing variability over time and for different operators, but using the same laboratory set-up.

### Transferability

The transferability of a test is a crucial parameter for robustness. The information should demonstrate that the test can be successfully repeated in a laboratory different from the one which has developed or which was involved in the optimisation of the test (3). This step is necessary to evaluate the practicality of the test and to provide an estimation of the amount of training that will be necessary to successfully transfer the test to a naive laboratory, as well as to identify possible sources of within-laboratory and between-laboratory variability.

The “transferability” of an *in silico* model is interpreted here as the ability to reproduce the model definition and model predictions. If the model definition is reproducible, a different operator (i.e. an expert other than the developer of the model) should be able to derive the same model or an equivalent model from the training set of data. If the model predictions are reproducible, a different operator should obtain the same predictions for a set of reference structures as the model developer had previously reported. Thus, the transferability of an *in silico* model will depend on the adequacy of the documentation explaining how the model was developed and how it should be applied.

### Between-laboratory variability

According to the current ECVAM validation process, the between-laboratory variability of a test is usually evaluated in three or four well-trained laboratories with a relatively large number of test substances. A validation study does not necessarily benefit from a larger number of laboratories; on the contrary, this can contribute to organisational and logistical problems.

If the assessment of the between-laboratory variability and predictive capacity (see below) of a test are considered to be two separate aspects, in principle, they can be addressed in different laboratories or at different time-points.

In general, the between-laboratory variability of a test can be assessed with a limited number of test substances. However, the set of test substances evaluated must cover the full range of toxic effects, relevant chemical classes and physicochemical properties. For a sufficiently well-established test, there is no obvious reason why this variability should change when a different subset of test substances is used. This approach will provide the opportunity to assess the predictive capacity later and with a larger number of substances in, for example, one laboratory only. This could result in a substantial reduction in cost and an increased throughput of validation studies.

The necessary number of participating laboratories and test substances should be defined on a case-

by-case basis and with the help of a biostatistician. This evaluation should be based on within-laboratory variability, transferability, and the availability of experience with the test.

### Predictive capacity

By applying a prediction model, the predictive capacity of an alternative test demonstrates how well the test can predict the reference standard (for example, an *in vivo* test result, an ecotoxicological endpoint or a human health effect). Predictive capacity is influenced by the number and range of test substances selected and the quality of the reference standard. In the current ECVAM validation process, the predictive capacity of a test is assessed by testing a defined number of substances in at least three laboratories. The new approach suggests that, once an acceptable level of between-laboratory variability had been established, the predictive capacity of a test might be assessed in one laboratory only, or could involve a larger number of laboratories, each of them testing only a different subset or an overlapping subset of the full set of substances. The data quality would then have to be assured by the reference and control materials, according to the SOP.

In addition to predictive capacity as a measure of relevance, mechanistic relevance, which refers to the scientific basis of the test, should also be considered. Some tests are associated with prediction models that are based purely on the statistical correlation with the endpoint being predicted, whereas other tests are also based on established mechanisms of biological action. Such mechanistically based models are more likely to be relevant to the purpose of the test. Since no objective measure of this quality is possible, it is suggested that current understanding of mechanism should be described within the test definition module.

### Applicability domain

The particular purpose for which a test can be applied should be clearly described — for example, for which toxicological endpoints, chemical classes, test materials, physicochemical properties and/or products. A subsequent change or extension of the applicability domain might require a new peer review.

### Performance standards

At the end of the validation exercise, performance standards should be established for each test. Tests based on a method similar to a validated test could

then be validated much faster (10), as has been shown in an ECVAM catch-up validation study (11) or outlined in the (US) Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) criteria (12). Performance standards can be applied to “me-too” developments or similar tests, but also to any change of a validated test. In applying this concept, care has to be taken to avoid test development tailored only to meet performance standards. While there is an obvious advantage in restricting the number of chemicals evaluated, development efforts should not focus on this limited set of chemicals.

### Catch-up Validation

The information obtained by performing a comprehensive validation exercise on a test can be used to expedite the future validation of a similar test or an improved version of the same test, thus broadening what has been called “catch-up validation” (10, 11). The application of the modular approach to already validated tests can then be carried out by re-evaluating the test with respect to one or more modules, according to technological progress.

It is suggested that a distinction should be made between minor and major changes to a test or differences between tests (criteria to be defined in general, and for each test when defining the performance standards). Minor changes will imply the need to prove equivalence to the validated test by using the performance standards only, while major changes imply the need to define a new set of reference materials and to perform a reassessment of each module affected by the change.

It is foreseeable that an increasing number of validity statements will have to be amended to reflect scientific progress, for example, for “me-too” developments, and for improvements in cell and tissue culture technology, as well as in endpoint measurement or data analysis procedures (including sensor and array technologies). This clearly applies when considering technologies such as QSAR or genomic approaches, which are still evolving technologies, and where validation, implying standardisation, should not be at the expense of innovation and progress. Furthermore, validated tests might often expand their applicability domains after the assessment of further types of products/chemicals. It is therefore suggested that an index for the validity statements of a given method should be created, similar to the release numbers of computer programs, indicating full validations and reassessments employing the performance standards. The need for such periodic reassessments points to the importance of having international validation bodies which track the history of specific tests and their validation status.

## Conclusions

The modular approach to validation being proposed by ECVAM represents a shift in emphasis rather than a fundamentally new concept. The approach is intended to be flexible and efficient, and should be applicable to all kinds of validation exercises. However, it needs to be further elaborated, so ECVAM has scheduled a number of workshops and task force meetings, so that these proposals can be discussed with groups of international experts, with the aim of defining the information required for the validation of pattern-based systems (for example, genomics, proteomics), computer-based models and testing strategies. Detailed guidance will need to be developed on how this information can be obtained by using prospective and/or retrospective validation, depending on the information already available. The new modular approach for applying the ECVAM principles on test validity promises to cover the various types of (alternative) tests and to be effective in responding to changing technologies.

## Acknowledgement

The critical discussion of the manuscript with Dr Leonard Schechtman (Food and Drug Administration and ICCVAM, USA) is gratefully appreciated.

## References

1. Balls, M., Blaauboer, B., Brusick, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stamatii, A.-L. & Walum, E. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA* **18**, 313–337.
2. Balls, M., Blaauboer, B.J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H. & Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshop 5. *ATLA* **23**, 129–147.
3. Curren, R.D., Southee, J.A., Spielmann, H., Liebsch, M., Fentem, J.H. & Balls, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods. *ATLA* **23**, 211–217.
4. Worth, A.P. & Balls, M. (2001). The role of ECVAM in promoting the regulatory acceptance of alternative methods in the European Union. *ATLA* **29**, 525–535.
5. Balls, M. & Karcher, W. (1995). The validation of alternative methods. *ATLA* **23**, 884–886.
6. Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Janusch, R., Prieto, P., Sabbioni, E., Worth, A. & Zuang, V. (2003). ECVAM's response to the changing political environment for alternatives: consequences of the European Union chemicals and cosmetics policies. *ATLA* **31**, 473–481.
7. Hartung, T., Balls, M., Bardouille, C., Blanck, O., Coecke, S., Gstraunthaler, G. & Lewis, G. (2002). Good cell culture practice. ECVAM Good Cell Culture Practice Task Force report 1. *ATLA* **30**, 407–414.
8. Anon. (1995). *CPMP/ICH/281/95: Note for Guidance on Validation of Analytical Procedures: Methodology* (adopted by the CPMP in December 1996), 9pp. London, UK: European Medicines Agency.
9. Anon. (1995). *CPMP/ICH/381/95: Note for Guidance on Validation of Analytical Methods: Definitions and Terminology* (adopted by the CPMP in November 1994), 5pp. London, UK: European Medicines Agency.
10. Balls, M. (1997). Defined structural and performance criteria would facilitate the validation and acceptance of alternative test procedures. *ATLA* **25**, 483–484.
11. Liebsch, M., Traue, D., Barrabas, C., Spielmann, H., Uphill, P., Wilkins, S., McPherson, J.P., Wiemann, C., Kaufmann, T., Remmele, M. & Holzhütter, H.-G. (2000). The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. *ATLA* **28**, 371–401.
12. Anon. (2003). *ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods*. NIH Publication No. 03-4508, 50pp. Research Triangle Park, NC, USA: NIEHS. Website <http://iccvam.niehs.nih.gov/docs/guidelines/subgde034508.pdf> (Accessed 31.3.04).