Workshop 5.13
Strategies for prioritising and streamlining the validation process

Validation *via* Weight-of-Evidence Approaches

Michael Balls and Robert Combes
FRAME, Nottingham, UK

Summary
It is not always necessary, or even possible, to conduct a practical laboratory study to establish the validity of tests or testing strategies. A weight-of-evidence approach aims to use already-available information in a structured, systematic, independent and transparent assessment. Crucial aspects include: study management and design; selection of experts without vested interests; collection of data and data quality control; differential weighing of various types of evidence, alone and in combination; evaluation of test performance in terms of reliability and relevance in relation to purpose; publication of outcome; and peer review.

Keywords: invalidation, test, validation, weight-of-evidence

Introduction

Early discussions on evaluating the reliability and relevance of non-animal tests for particular purposes, i.e. on their validation, focused on the performance of practical multi-laboratory studies (Balls et al., 1990, 1995). However, it subsequently became clear that such practical studies would not always be necessary or even possible, and that, in some circumstances, a weight-of-evidence approach would be more appropriate, i.e. an evaluation based on the independent collection and assessment of appropriate and sound pre-existing supporting evidence. For example, there could be evidence of sufficient volume and quality to permit an evaluation of a method which had long been in use. In other cases, there might not be *in vivo* benchmark data of sufficient breadth and quality to serve as acceptable reference standards for the practical evaluation of an *in vitro* method.

What is weight-of-evidence validation?

The weighing of evidence is, of course, fundamental to the evaluation of any scientific work, but, in this paper, we distinguish between practical validation studies (those involving a multi-laboratory performance of the test) and weight-of-evidence validation studies (those without a dedicated practical study).

This type of validation has been referred to in different ways; for example, *validation by retrospective analysis* (US Interagency Coordinating Committee on the Validation of Alternative Methods, ICCVAM [NIH, 1997]) and *validation through available data* (OECD, 2003). However, we strongly suggest that it would be preferable to call it weight-of-evidence validation, since this more clearly indicates that the approach could be used either retrospectively or prospectively.

There are four principal types of weight-of-evidence validation:
1. The re-evaluation of a previous multi-laboratory validation study (or series of studies).
2. The analysis of data obtained with the same test in different laboratories, but at different times, in studies that were not intended to be parts of a validation exercise.
3. The analysis of data obtained in one or more laboratories using a variation of the protocol that was used in an earlier multi-laboratory validation study.
4. The validation of a testing strategy comprising the use of several test methods, each of which has been previously validated, either in a multi-laboratory validation study or by weight-of-

evidence validation, or a different approach to testing, such as read-across for chemical hazard and risk analysis.

An example of the first type of weight-of-evidence validation would be when a test method is being proposed for a slightly different purpose than that for which it was originally validated. In the second type, it is likely that the protocols used at different times in the various laboratories involved, as well as other protocol parameters, such as sources and types of test chemicals and other materials, would not have been standardised. The third and fourth types of weight-of-evidence validation require judgements to be made about the performances of tests, either when they were combined or when there were small differences in the ways in which they had been conducted.

Criteria for readiness for evaluation and for the quality of the assessment

As with practical validation studies, it must first be established that a procedure or strategy is ready to be evaluated, and, in particular, that there is:

1. a clear definition of the scientific purpose and proposed practical application of the procedure or strategy;
2. a clear description of its scientific basis;
3. a convincing case for its relevance, including an explanation of the need for it in relation to other procedures or strategies;
4. an optimised protocol for the procedure or a clear indication of how the strategy is applied;
5. a statement about the limitations of the approach; and
6. evidence concerning its performance, intra-laboratory reproducibility and, if available, inter-laboratory transferability.

A decision that a procedure or strategy is ready for validation would be taken by a recognised validation authority, such as ECVAM or ICCVAM, who, as sponsors of the study, would appoint a Management Team, who would prepare an overall design for the study. This would proceed in a series of stages.

Stage 1. Appointment of Management Team by Study Sponsors (responsible for overall design of study and appointment of Data Collection and Evidence Assessment groups).

Stage 2. Independent data collection from all available sources by Data Collection Group (to include experienced and independent information scientists and scientists familiar with the type of method and purpose).

Stage 3. Submission of evidence to Evidence Assessment Group and joint review of adequacy of data coverage and quality.

Stage 4. Weight-of-evidence assessment of data by Evidence Assessment Group (according to previously-agreed criteria).

Stage 5. Reporting of outcome to and by Management Team and to Study Sponsors.

Stage 6. Publishing of report of conduct and outcome of study in the peer-review literature.

Stage 7. Independent peer review of study as a whole.

Stage 8. Publication of peer review report.

Stage 9. Consideration by Regulatory Authorities and other appropriate bodies of the acceptability of a validated method for application.

Since the acceptability of the weight-of-evidence assessment itself will also have to be evaluated at a later stage by peer review, it is important that the following criteria are taken into account when the assessment is being planned:

1. Clarity of the defined goals.
2. Quality of the overall design.
3. Independence of management.
4. Quality criteria for evidence.
5. Independence of collection of evidence.
6. Criteria for weighing of evidence.
7. Independence of weighing of evidence procedure.
8. Quality of reporting of outcome.
9. Publication of outcome in peer-reviewed literature.
10. Transparency of whole process (including the identities and affiliations of all the experts involved).

Collecting the evidence

Clearly, the type of evidence to be collected, how it is to be obtained and selected, how its quality is to be checked, and whether it is relevant and reliable, are crucial issues, but it must also be established that the evidence is truly representative of the performance of the procedure or strategy and that its collection is without bias. In addition, how the data are applied, for example, in a prediction model to classify and label chemicals according to a particular type of toxicity, must be included.

It is particularly important that the collection of the evidence is performed by a group of experts who include information scientists and scientists familiar with the type of method under evaluation and the purpose at which it is aimed. This Data Collection Group should be appointed by the Management Team and should be independent both of the developers or proponents of the test procedure or testing strategy and of those who will weigh the evidence once it has been collected.

Weighing the evidence

The performance criteria to be met by a procedure or strategy should be defined by the Management Team in advance of the assessment, according to its proposed purpose, and should be both reasonable and scientifically-based.

The assessment itself cannot be used to improve the evidence but can be used to optimise the use of a method and/or to identify or confirm its strengths and weaknesses. For example, it may become clear that the method is suitable for certain classes of chemicals, but not for others.

It is vital that the members of the Evidence Assessment Group, appointed by the Management Team and responsible for conducting the assessment, have a sufficient breadth and depth of experience and expertise, and that they are independent of the developers or proponents of the method, and of those who collected and presented the evidence.

A case-by-case approach will be essential, and different kinds of evidence will have different levels of value in contributing to

the overall assessment. This will involve evaluations of the plausibility, relevance, consistency, volume and overall strength of the evidence.

Conclusions from a weight-of-evidence assessment

The assessment should lead to a clearly-stated outcome, supported by reasoned argument. There are likely to be three main types of conclusions, depending on the degree to which the weighing of the evidence resolves uncertainty about the relevance and reliability of the test or strategy for its purpose:

1. There is sufficient evidence that a test procedure/testing strategy is reliable and relevant for its stated purpose, and it should be accepted for use for that purpose.
2. There is conflicting evidence about the relevance and reliability of a test procedure/testing strategy for its stated purpose, and a) either the balance is in favour of provisional acceptance or b) further evidence should be obtained.
3. There is sufficient evidence that the test procedure/testing strategy is not reliable and relevant for its stated purpose, and it should not be accepted for use for that purpose.

The outcome of the assessment should be published in a peer-reviewed journal, by or on behalf of the Management Team, as well as being submitted to the sponsors of the exercise and other relevant bodies for an independent and transparent peer review of the study as a whole (design, data collection, weight-of-evidence assessment and reporting).

The application of weight-of-evidence validation

Up to now the European Centre for the Validation of Alternative Methods (ECVAM) has tended to favour a practical approach to validation, whereas the ICCVAM has favoured the weight-of-evidence approach. Thus, of the first ten methods endorsed as valid by the ECVAM Scientific Advisory Committee (ESAC), eight involved practical studies sponsored by ECVAM (tab. 1). The other two endorsements involved methods previously reviewed by ICCVAM.

There is likely to be a trend toward weight-of-evidence assessments, especially as it is increasingly likely that the non-animal tests of the future will contribute evidence that will be used along with other evidence as parts of test batteries and decision-tree testing strategies.

In addition, retrospective validation assessments, based on the ability of tests to give the same predictions as previously obtained, for example with animal tests, will progressively be replaced by prospective assessments, especially where testing strategies are based on more-modern toxicological methods, themselves based on a greater understanding of mechanisms of toxicity and the application of emerging biotechnologies such as toxicogenomics and toxicoproteomics (Bhogal et al., 2005).

Pitfalls to be avoided

It is already clear from experience gained so far that a number of serious pitfalls may be encountered when planning and conducting a weight-of-evidence evaluation, some of which also apply to practical validation studies. These include:

1. implausibility of the test system;
2. inadequate development of the test or testing strategy;
3. lack of evidence and/or poor quality of evidence;
4. bias in selection/presentation of evidence;
5. failure to establish relevance of evidence;
6. lack of a prediction model for applying outcome;
7. lack of clarity/precision in the weighing procedure;
8. pseudosophistication of the weighing procedure;
9. bias in derivation/application of weighing procedure;
10. unreasonably demanding or unreasonably undemanding test performance criteria;
11. injudicious application of the precautionary principle;
12. bias in Data Selection Group or Evidence Assessment Group; and
13. politicisation of the whole process.

The importance and consequences of these pitfalls were discussed at an ECVAM workshop on weight-of-evidence approaches to validation, held in 2004, for which a report is in preparation.

Tab. 1: Endorsement by the ECVAM Scientific Advisory Committee of validated tests for chemical toxicity, 1997-2001.

		Practical study
1. The 3T3 NRU test for phototoxic potential	November 1997	+
2. The EPISKIN™ skin corrosivity test	April 1998	+
3. The rat TER skin corrrosivity test	April 1998	+
4. The application of the 3T3 NRU phototoxicity test to UV filter chemicals	May 1998	+
5. The local lymph node assay for skin sensitisation	March 2000	-
6. The EpiDerm™ skin corrosivity test	March 2000	+
7. The CORROSITEX™ skin corrosivity test	December 2000	-
8. The embryonic stem cell test for embryotoxicity	May 2001	+
9. The whole-embryo culture test for embryotoxicity	May 2001	+
10. The micromass test for embryotoxicity	May 2001	+

Conclusions

There is no doubt that, when permitted to operate effectively and without bias, the ECVAM/ICCVAM/OECD validation process can be used to independently establish that new animal and non-animal test procedures are sufficiently relevant and reliable for their stated purposes and should be considered for regulatory use. However, we have come to the conclusion that the validation process is under threat because of vested interests of various kinds. In addition, it is clear that many currently-accepted animal tests and candidate animal and non-animal tests do not, and could never, meet the agreed criteria for necessity, test development, prevalidation, validation and acceptance.

We have therefore proposed that there is an urgent need for an *invalidation* process to parallel and protect the validation process, so that such methods can be independently reviewed and, where necessary, declared irrelevant and/or unreliable for their claimed purposes (Balls and Combes, 2005). An additional advantage of such a process would be that valuable resources would no longer be wasted in attempts to secure the acceptance of inherently inadequate tests or strategies. This proposal was discussed at an ECVAM/FRAME workshop in September 2005.

Our aim is to ensure that methods will only be put into practice if they have been satisfactorily and independently shown to be reliable and relevant for their stated purposes. We want to protect the reputation of the validation process and, more importantly, prevent the adoption of methods which will ultimately fail, leading to damage to human health.

References

Balls, M., Blaauboer, B., Brusick, D. et al. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA 18*, 313-337.

Balls, M., Blaauboer, B. J., Fentem, J. H. et al. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshop 5. *ATLA 23*, 129-147.

Balls, M. and Combes, R. (2005). The need for a formal invalidation process for animal and non-animal tests. *ATLA 33*, 299-308.

Bhogal, N., Grindon, G., Combes, R. and Balls, M. (2005). Toxicity testing: creating a revolution based on new technologies. *Trends Biotechnol. 23*, 299-307.

NIH (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. Research Triangle Park, NC: NIEHS.

OECD (2003). Draft Guidance Document No 34 on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment, Environment Directorate. Paris: OECD.

Correspondence to

Prof. Michael Balls
FRAME
Russell & Burch House
96-98 North Sherwood Street
Nottingham NG1 4EE
UK
e-mail: Michael.Balls@btopenworld.com

Streamlining the Validation Process: The ICCVAM Nomination and Submission Process and Guidelines for New, Revised and Alternative Test Methods*

Leonard M. Schechtman[1], Marilyn L. Wind[2] and William S. Stokes[3]

[1]U.S. FDA, National Center for Toxicological Research, Rockville, Maryland, USA; [2]U.S. Consumer Product Safety Commission, Bethesda, Maryland, USA; [3]NIEHS, Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, North Carolina, USA

Summary

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) has developed and implemented a process for the nomination and submission of test methods and for their prioritisation for review and evaluation. Prioritisation of proposed test methods is a function of their regulatory applicability, anticipated multi-agency interest and use, responsiveness to the replacement, reduction, and refinement of animal use, potential for improved predictivity of adverse effects relative to currently employed methods, and efficiency and economic savings. The newly revised ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods (http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm) were developed to assist test method sponsors/nominators in organising the information needed to assess the validation status of test methods at any stage of the validation process and the extent to which the ICCVAM validation and acceptance criteria have been or will be addressed. The original guidelines, in use since 1998 to evaluate the scientific validity of test methods that have since achieved regulatory acceptance, have been updated to reflect experience gained and to help to facilitate a more efficient process. Adherence to these revised guidelines will help ensure the sufficiency of data and information for independent peer review and for regulatory authorities to determine the scientific validity and regulatory acceptability of test methods. The elements comprising these guidelines have now been incorporated into international guidance for the evaluation of methods proposed for new test guidelines. The ICCVAM nomination, submission and prioritisation process and the content and organisation of submissions or nominations are described.

Keywords: ICCVAM guidelines, nominations, submissions, prioritisations

Introduction

The ICCVAM Authorization Act of 2000 stipulates that *Each Federal agency … shall ensure that any new or revised … test method, including animal test methods and alternatives, is determined to be valid for its proposed use prior to requiring, recommending, or encouraging the application of such test method*. In order to facilitate the validation process and provide Federal regulatory agencies with the needed relevant and reliable test methods for regulatory decision-making purposes, ICCVAM was established (initially as an ad hoc committee in 1994, as a standing committee in 1997, and as a permanent body in 2000) and was comprised of 15 Federal regulatory and research agencies. ICCVAM, together with its scientific and operational support centre, NICEATM (the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods), were created to ensure that new, revised and alternative test methods (a) are scientifically validated to meet agency needs, (b) are more predictive of human health and ecological effects than current methods, and (c) contribute to improved public health. ICCVAM was also tasked with increasing the efficiency and effectiveness of agency test method review, thus eliminating unnecessary duplicative review efforts among different agencies and creating an environment in which agencies could share relevant experiences, information and expertise. Such review efforts would also be aided by optimising the use of scientific expertise recruited from outside the government. ICCVAM was also designated as the principal national body that would promote the replacement, reduction and refinement of animal use for testing and research purposes, the 3Rs (Russel and Burch, 1959).

In order to accomplish these objectives, ICCVAM and NICEATM (a) consider test method nominations and submissions from agencies and the public for review and evaluation of validation status of test methods, (b) review and evaluate new, revised and alternative test methods that may be potentially acceptable for specific regulatory uses, (c) coordinate technical reviews of proposed new or revised or alternative test methods of interagency interest, (d) submit test recommendations to Federal agencies, solicit agency responses, and make these publicly available, and (e) facilitate and/or provide guidance on test method development, validation criteria, validation studies, validation processes, acceptance of scientifically validated test

*The views expressed are solely those of the authors and do not represent any official positions of any Federal agency or organisation.

methods, implementation of ICCVAM-recommended validated methods, awareness of accepted test methods by Federal agencies and other stakeholders, and interagency and international harmonisation of test methods.

Until the advent of ICCVAM, test method validation and acceptance among regulatory agencies had largely been an *ad hoc* procedure, and there was limited communication between agencies regarding validation efforts and processes. ICCVAM responded by devising procedures designed to streamline the validation process and make it more efficient as well as provide a forum for cross-agency communication and consensus-building regarding the evaluation and use of those methods. Although ICCVAM was initially established with a focus on test methods that would address the concept of the 3Rs, it became evident that the validation process and the validation criteria that were to be promulgated by ICCVAM would be equally applicable to all test methods developed for regulatory use, irrespective of whether they were new, revised or alternative in nature.

ICCVAM Guidelines

In 1997, ICCVAM published its Guideline on the *Validation and Regulatory Acceptance of Toxicological Test Methods* (1997). This report provides information on (a) the test method validation process, (b) regulatory acceptance processes, (c) criteria for validation of a test method, (d) criteria for regulatory acceptance of a method, and (e) implementation of a validated test method. The guideline also contains a glossary of terms that would be useful in understanding the terminology used for test method validation purposes and negotiating the validation and regulatory acceptance processes. In 2003, the *ICCVAM Guidelines for Nomination and Submission of New, Revised, and Alternative Test Methods* (2003) were published and were intended to complement the 1997 ICCVAM report. As its title suggests, the document was designed to serve as a tool that would provide guidance to both submitters and regulators on the ICCVAM test method nomination and submission process in order to ensure that nominations/submissions contain adequate information and that the proposed test methods have regulatory applicability. The guideline presents details regarding the ICCVAM criteria for test method prioritisation, validation and regulatory acceptance and it conveys the concept of Performance Standards (Stokes et al., 2006) for test methods. It was developed to guide test method development and to serve as a resource for standardisation and validation efforts. It provides a framework for test method submissions, recommending a skeleton structure outlining the essentials for assisting test method sponsors/nominators in organising the information supporting the validity of a new or modified test method. It describes the information (data, supporting records) needed in test method nominations or submissions to evaluate the validation status of a test method at any stage from development to completion, i.e. the extent to which the validation and acceptance criteria have been or will be addressed.

These two guidelines, together with the encouragement they offer for establishing an ongoing dialogue between sponsors/submitters and ICCVAM/NICEATM, are viewed as a mechanism for facilitating efficient and effective review of submissions and facilitating regulatory acceptance decisions. It is worth noting that the standard outline derived from the ICCVAM nomination and submission guideline is now included in international guidance for submission of test methods proposed for regulatory consideration (OECD, 2005).

ICCVAM nomination, submission and prioritisation process

A test method *nomination* and a test method *submission* differ from each other primarily by the amount of information furnished in support of it. A test method nomination is a test method proposed to ICCVAM for review and evaluation for which a complete test method submission is not available. Examples of nominations are: (a) test methods for which adequate validation studies have apparently been completed but lack a complete submission package; (b) test methods that appear promising based on limited pre-validation or validation data and are proposed for additional validation studies; (c) test methods that have been developed and are proposed for pre-validation or validation studies; and (d) test methods that are recommended for consideration via a workshop or other activity. A test method submission is a test method proposed to ICCVAM for review and evaluation for which adequate validation studies have been completed to characterise the usefulness and limitations of the test method for a specific proposed regulatory testing requirement or application, and adequate documentation of the scientific validity of the test method has been prepared in accordance with ICCVAM test method submission guidelines.

The process followed when ICCVAM receives a nomination to consider the evaluation of a test method or receives a submission of a data package containing complete or partial information needed to assess the validation status of a test method is depicted in figure 1. Upon receipt of a nomination or submission, NICEATM initiates its preliminary evaluation to consider the information received and the extent of the supporting data furnished that would lend credence to the regulatory applicability of the test method. That preliminary evaluation summarises the extent to which proposed test method submissions or nominations address the following ICCVAM prioritisation criteria:

- The extent to which the proposed test method is applicable to regulatory testing needs and applicable to multiple agencies/programmes.
- The degree to which the proposed test method is warranted, based on the extent of expected use or application and impact on human, animal or ecological health.
- The potential for the proposed test method, compared to current test methods accepted by regulatory agencies, to refine animal use (decrease or eliminate pain and distress), reduce animal use, or replace animal use.
- The potential for the proposed test method to provide improved prediction of adverse health or environmental effects, compared to current test methods accepted by regulatory agencies.

- The extent to which the test method provides other advantages (e.g. reduced cost and time to perform) compared to current methods.

NICEATM then presents its recommendations to ICCVAM regarding the next steps that should be considered in the evaluation process. Based upon those recommendations and comments received from its Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) as well as input from the public, ICCVAM determines whether the test method is of sufficient interest and applicability to one or more agencies to warrant further evaluation and renders its recommendations as to the prioritisation that should be assigned to the test method nomination/submission under consideration. NICEATM determines the resources necessary to accomplish ICCVAM's recommendations and management decides whether those resources can be made available for the proposed purpose(s). Upon notification of the availability of the necessary resources, ICCVAM establishes an interagency test method Working Group of knowledgeable scientists to work with NICEATM in organising the necessary evaluation or validation study. Together, ICCVAM, the interagency Working Group, and NICEATM organise workshops, expert panel meetings, independent peer reviews, validation studies, or expedited reviews, as appropriate, to evaluate the validation status of the proposed test method.

Content of a test method submission

The recommended content of a dossier submitted for an ICCVAM assessment of the validation status of a test method is outlined in figure 2. Following is a description of the preferred information that should be included in a test method nomination or submission to ICCVAM. For further, more comprehensive details, the reader is referred to Appendix A of the ICCVAM nomination and submission guideline (ICCVAM, 2003), which provides a detailed outline of the content sought for nominations and submissions to ICCVAM. Although there is no mandatory minimum requirement regarding the information to be provided, ICCVAM's consideration of the proposed test method will be expedited by providing as much information as possible. Furthermore, experience has demonstrated that the more closely a submission adheres to reporting the suggested information in a format approximating that of the suggested outline, the more efficient is the evaluation process and the fewer the data gaps that would be encountered and that would need filling after the fact. Areas where the requested information is not available or is incomplete should be indicated and the studies that would be conducted to generate those necessary data should be specified.

Section 1 provides the introduction and rationale for the proposed test method. It introduces the proposed test method and describes its regulatory and scientific rationale. This would include a description of how the proposed test method can be used in a regulatory setting (e.g. as a screen in a tiered testing strategy, as an adjunct test to provide mechanistic information, as a surrogate for an existing test method) and how the proposed test method might be included in the overall safety or hazard assessment process. This information would also include a discussion of the mechanistic basis of the proposed test method and the context in which it will be used to measure or predict the toxicological activity of a test material or substance.

Section 2 reports on the test method protocol components. This would include an explanation and description of the basis for decisions on critical functional, structural, and procedural ele-

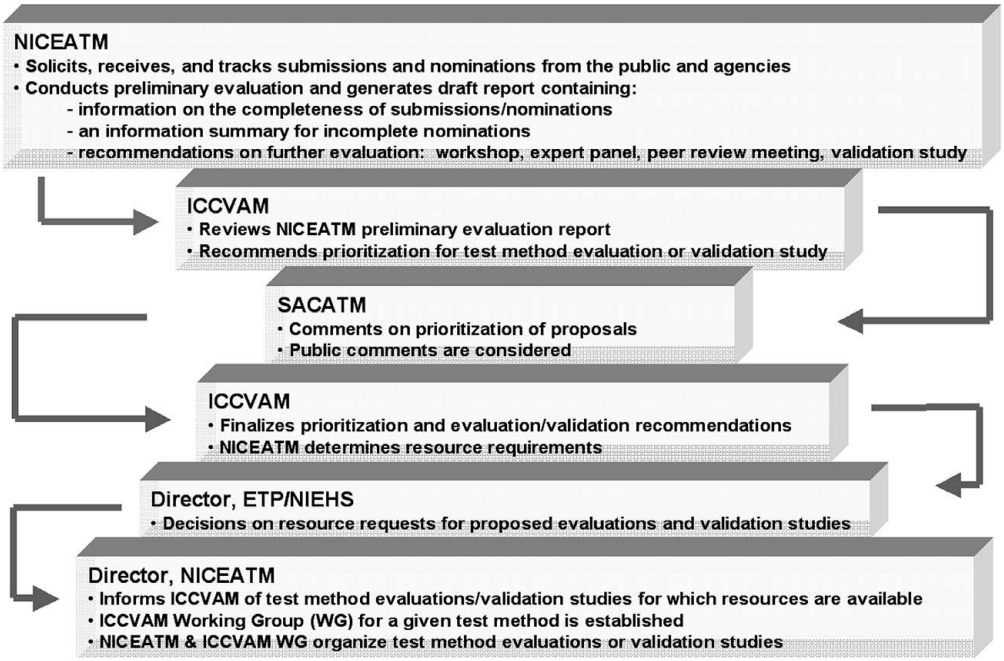

Fig. 1: ICCVAM test method nomination, submission, and prioritisation process

ments of the test method protocol. Critical protocol elements would include such factors as (a) dose/concentration selection procedures, (b) duration of exposure, (c) number of replicate and/or repeat experiments, (d) endpoint(s) measured, (e) concurrent (positive, negative, vehicle) controls and the basis for their selection, (f) acceptable response range, (g) statistical or other methods of data analysis (including dose-response relationships if applicable), (h) decision criteria, i.e. those criteria used to designate a response as positive, negative or equivocal, (i) test system specifications, e.g. cell systems and properties, animal models (species, strain, sex, age, genetic background, diet), (j) details regarding any proprietary components of the test method and the maintenance of their integrity (reliability and accuracy). In addition, a complete, detailed protocol for the proposed test method should be appended to the submission.

Section 3 addresses the Substances Used for Validation of the Proposed Test Method. Here, the rationale for the numbers and types of substances tested during the validation process and the specific chemical or formulation names and relevant chemical and product classes are reported. It is important to recognise that the limitations associated with a test method are often indicative of its effectiveness in evaluating certain classes of chemicals and therefore any characteristics thought to have direct impact on test method accuracy and/or reliability should be described. To the extent possible, the information provided for each test substance should include (a) Chemical Abstracts Service Registry Number (CASRN), (b) physical and chemical characteristics, (c) concentrations tested, (d) purity, (e) source, (f) stability of the test substance in the test medium, and (g) in the case of mixtures, the constituents and their relative concentrations. Information regarding the use of coded substances and blind testing during the validation process should also be reported. In cases where the proposed test method is mechanistically and functionally similar to a previously validated ("pioneer") test method, a discussion should be presented regarding the performance of the proposed ("me-too") test method using reference chemicals recommended in the Performance Standards established for the pioneer test method.

1.0 Introduction and scientific and regulatory rationale for the proposed test method
2.0 Proposed test method protocol
3.0 Substances used for validation
4.0 Reference data used for performance assessment
5.0 Test method data and results
6.0 Test method accuracy assessment
7.0 Test method reliability (repeatability/reproducibility) assessment
8.0 Test method data quality
9.0 Other scientific reports, reviews pertinent to the proposed test method
10.0 Animal welfare considerations: refinement, reduction, and replacement
11.0 Practical considerations (cost, time, training, equipment, transferability)
12.0 References
13.0 Supporting materials (e.g., detailed test method protocol) in appendices

Fig. 2: Content of a test method submission

Section 4 discusses the *in vivo* reference data used for the performance assessment of the proposed test method. If the proposed test method is intended to replace or substitute for an existing *in vivo* reference test method, then a comparison of data generated by each is essential. Information sought would also include such factors as (a) the protocol(s) used to generate the *in vivo* reference data, (b) the *in vivo* reference data (human, animal) used to assess the accuracy of the proposed (new, modified) test method, (c) the quality of the *in vivo* reference test method data and the extent of Good Laboratory Practice (GLP) compliance (U.S. EPA, 2003a; U.S. EPA, 2003b; U.S. FDA, 2003; OECD, 1998; Japanese Good Laboratory Practice Standards, 1997; Cooper-Hannan et al., 1999), (d) the criteria used to select the *in vivo* reference test method (or human) data, (e) the availability of original study data for the *in vivo* reference test method studies reported.

Section 5 conveys the test method data and results. Here, the data generated by the testing of chemicals and substances using the proposed test method protocol are reported, including all data (accompanied by "raw" data) from all studies (successes and failures), summary data presented in tabular or graphic form as appropriate, the statistical approach(es) used, together with their justification, in the analysis of the data. Accompanying those data would be discussions of protocol modifications, if any, that might have been introduced during method development and their impact.

Section 6 presents the assessment of the performance of the proposed test method relative to the currently accepted reference method. The accuracy of the test method (e.g. its concordance, sensitivity, specificity, false positive and false negative rates, positive and negative predictivity) is calculated and discussed with respect to its ability to measure or predict the effect of interest and any results that may be discordant with those generated by the reference test method would be discussed. In instances where the proposed test method is measuring or predicting an endpoint for which there is no pre-existing test method, the frequency of correct predictions should be compared to relevant information from the species of interest. In cases where the proposed test method is a me-too method, i.e. is mechanistically and functionally similar to a validated (pioneer) test method with established performance standards, the accuracy of both test methods should be compared. Also addressed in this section would be a discussion of the usefulness and limitations of the test method based on its performance characteristics.

Section 7 provides an assessment of the reliability (repeatability and reproducibility) of the test method. Here, analyses and conclusions would be presented for intra-laboratory repeatability and intra- and inter-laboratory reproducibility together with a quantitative statistical analysis of the extent of any variability encountered. This assessment would include a discussion of the rationale for the selection of the substances used to evaluate the intra- and inter-laboratory reproducibility and the extent to which those substances represent the range of possible test outcomes. Historical negative, vehicle and positive control data and

the responses of the proposed test method relative to these control values would also be presented and discussed. In cases where the proposed test method is a me-too method similar to a validated pioneer test method with established performance standards, the reliability of both test methods should be compared and the potential impact of any differences discussed.

Section 8, which discusses test method data quality, communicates the extent of adherence to national and international GLP guidelines (OECD, 2005; U.S. EPA, 2003; U.S. EPA, 2003b; U.S. FDA, 2003; OECD, 1998; Japanese Good Laboratory Practice Standards, 1997) as well as the results of any data quality audits. Deviations from GLP guidelines and their impact, unpublished data (supported by laboratory notebook entries), and the availability of laboratory notebooks and other retained data would also be addressed.

Section 9, which deals with other relevant scientific reports and reviews, discusses data and reports from other studies (published and unpublished) conducted using the proposed test method. Conclusions regarding the proposed test method available from independent peer-reviewed reports or other scientific reviews would be presented and those conclusions should be compared to those reached in the submission. In cases where the proposed test method is a me-too method similar to a validated pioneer test method with established performance standards, the results of any studies conducted subsequent to the ICCVAM evaluation should be reported and any impact on the reliability and accuracy of the proposed test method would be discussed.

Section 10 addresses animal welfare considerations. In this section a description of how the method will refine, reduce or replace animal use compared to currently accepted methods used for the endpoint of interest is presented. If animals are used in the test method, the discussion addresses (a) the rationale for the use of animals, (b) the basis for the number of animals required, (c) approaches used to prevent or minimise pain and distress as appropriate and (d) sources used to determine availability of 3Rs alternatives (USDA, 1966; Public Health Service (1996).

Section 11, which addresses practical considerations, includes discussions of the transferability aspects of the proposed test method (e.g. facilities, equipment, supplies, expertise, training and proficiency necessary), as well as time and cost considerations.

Section 12, references, cites all publications referenced in the submission.

Section 13, supporting materials, is comprised of all necessary appendices containing information relevant to the proposed test method that would supplement the central content of the submission. Such material would include, but would not necessarily be limited to, (a) the complete detailed protocol and relevant standard operating procedures (SOPs) as appropriate, (b) copies of all relevant publications, (c) all non-transformed original data, (d) suggested performance standards developed for the proposed test method, and (e) any other supporting documentation considered informative and helpful.

Conclusion

The *ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods* (ICCVAM, 2003), were developed to assist test method sponsors and nominators in organising the information needed by ICCVAM to assess the validation status of a new or modified test method at any stage of the validation process. These guidelines include guidance on the process for submitting nominations to ICCVAM for test methods that are proposed for further consideration, but which may require further compilation of data or even additional validation studies. The guidelines convey the ICCVAM test method nomination and submission process. They also describe the information that should be provided in test method nominations or submissions and provide a framework for that information so that ICCVAM can evaluate appropriately the extent to which the validation and acceptance criteria have been addressed or will be addressed in proposed studies. In addition, the guidelines put into practice the use of performance standards, which communicate the basis on which a validated and accepted proprietary (i.e. copyrighted, trademarked, registered) or non-proprietary test method has been determined to have sufficient accuracy and reliability for a specific testing purpose. Such performance standards are designed to be met by proposed test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect.

Test method sponsors are encouraged to employ these guidelines together with the ICCVAM guideline on the *Validation and Regulatory Acceptance of Toxicological Test Methods* (ICCVAM, 1997) and to seek consultative advice from ICCVAM and NICEATM throughout the test method development, prevalidation, and validation process, as well as during preparation of submissions. This interactive process enhances the likelihood that agencies will have sufficient data and information to determine the extent that a test method can generate information that will meet their regulatory needs. Proper usage of these guidelines and capitalisation on interactions with ICCVAM and NICEATM will help to maximise the likelihood that validation studies and submissions will adequately characterise the usefulness and limitations of the proposed test method. Complete submissions are essential and serve as a basis for assessing the validation status of a proposed test method through an independent ICCVAM peer review process.

References

Cooper-Hannan R, Harbell J, Coecke S, et al. (1999). The principles of good laboratory practice: application to in vitro toxicology studies. *ATLA 27*, 539-577.

ICCVAM Authorization Act of 2000. (2000). Public Law 106-545, December 19, 2000. Available: http://iccvam.niehs.nih.gov/about/PL106545.pdf [accessed 1 November 2005].

ICCVAM. (1997). Validation and Regulatory Acceptance of

Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/docs/docs.htm#general [accessed 1 November 2005].

ICCVAM. (2003). ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. National Institute of Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm [accessed 1 November 2005].

Japanese Good Laboratory Practice Standards, including: (1) Pharmaceutical GLP Standard. (1997). Pharmaceutical Affairs Bureau, Ministry of Health and Welfare, Available: http://www.jsqa.com/en/glp/glp21e.html [accessed 1 November 2005]; and (2) Good Laboratory Practice Standards. 1984. Japanese Ministry of Agriculture, Forestry, and Fisheries. Available: http://www.ovpr.uga.edu/qau/maffintr.html [accessed 1 November 2005].

OECD (1998). Principles of Good Laboratory Practice and Compliance Monitoring. ENV/MC/CHEM (98)17. Paris, France: Organisation for Economic Co-operation and Development. Available: http://www.olis.oecd.org/olis/1998doc.nsf/LinkTo/env-mc-chem(98)17 [accessed 1 November 2005]. Also available at: http://www.oecd.org/document/63/0,2340,en_2649_34381_2346175_1_1_1_1,00.html [accessed 1 November 2005].

OECD. (2005). OECD Series on Testing and Assessment, Number 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, Paris, France. Available: http://appli1.oecd.org/olis/2005doc.nsf/linkto/env-jm-mono(2005)14 [accessed 1 November 2005].

Public Health Service. (1996). Public Health Service Policy on Humane Care and use of Laboratory Animals. Washington, DC, U.S. Department of Health and Human Services. Available: http://grants1.nih.gov/grants/olaw/references/phspol.htm [accessed 1 November 2005].

Russell, W. M. S. and Burch, R. L. (1959). The Principles of Humane Experimental Technique. London: Methuen.

Stokes, W. S., Schechtman, L. M., Wind, M. et al. (2006). The Use of Test Method Performance Standards to Streamline the Validation Process, *ALTEX*, this issue.

U.S. EPA. (2003a). Federal Insecticide, Fungicide, and Rodenticide Act: Good Laboratory Practice Standards; Final Rule. 40 CFR Part 160. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr160_03.html [accessed 1 November 2005].

U.S. EPA. (2003b). Toxic Substances Control Act (TSCA): Good Laboratory Practice Standards; Final Rule. 40 CFR Part 792. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr792_03.html [accessed 1 November 2005].

U.S. FDA. (2003). Good Laboratory Practice for Nonclinical Laboratory Studies. 21 CFR Part 58. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/21cfr58_03.html [accessed 1 November 2005].

USDA. (1966). Animal Welfare Act. 7 USC: 2131-2156. Available: http://warp.nal.usda.gov:80/awic/legislat/usdaleg1.htm [accessed 1 November 2005].

Correspondence to
Leonard M. Schechtman, Ph.D.
U.S. FDA
National Center for Toxicological Research
Rockville
Maryland, USA
e-mail: leonard.schechtman@fda.hhs.gov

The Use of Test Method Performance Standards to Streamline the Validation Process

William S. Stokes[1], Leonard M. Schechtman[2], Amy Rispin[3], Kailash Gupta[4], Karen Hamernik[3], Raymond Tice[1] and Marilyn Wind[4]

[1]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, NC, USA; [2]US Food and Drug Administration, National Center for Toxicological Research, Rockville, MD, USA; [3]U. S. Environmental Protection Agency, Washington, DC, USA; [4]Consumer Product Safety Commission, Bethesda, MD, USA

Summary

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recently defined and established a process for the development and use of test method performance standards. Performance standards communicate the basis by which proprietary and non-proprietary test methods have been determined to be acceptable for proposed specific testing uses. They also provide the basis for evaluating the acceptability of proposed test methods that are structurally and functionally similar to an accepted test method. The development and use of test method performance standards is expected to streamline the validation and regulatory acceptance of new, revised and alternative test methods.

Keywords: performance standards, validation, regulatory acceptance, proprietary test methods

Introduction

Alternative test methods proposed for regulatory use must undergo validation studies to assess their reliability and relevance for specific applications (ICCVAM, 1997, 2003; OECD, 2005a). Regulatory agencies use this information to determine the applicability and acceptability of the test method for addressing specific regulatory safety testing requirements. Many new and alternative test methods are proprietary in nature and are protected by intellectual property laws such as patents, trademarks, and copyrights. Such intellectual protections stimulate innovation by providing financial incentives for companies to develop and market new products, such as *in vitro* testing methods that may reduce, refine, or replace animal use. However, US laws require that government regulatory authorities cannot simply endorse or approve proprietary methods until they first convey the basis by which the proprietary methods have been determined to be acceptable for use (CFR, 2005; EPA, 1995).

The requirement for regulatory agencies to provide the basis for acceptability of a proprietary test method was first brought to the attention of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) during USEPA (US Environmental Protection Agency) consideration of four *in vitro* test methods for skin corrosion, of which three were proprietary. All four *in vitro* test methods had been reviewed and recommended by ICCVAM for identifying chemicals that may cause skin corrosivity (ICCVAM, 1999, 2002). The USEPA requested that ICCVAM develop performance standards that could be used by the USEPA to convey the basis by which each of these four *in vitro* methods were found to be useful as a screening test to identify dermal corrosives in a tiered testing strategy.

At about the same time, ICCVAM proposed a generic version of Corrositex® as an OECD test guideline. This generic version was necessary since the OECD Test Guidelines Programme does not allow inclusion of any proprietary test methods or test method components in OECD test guidelines. However, it was not clear how OECD member countries that received data from a generic version of Corrositex® would be able to determine that the generic version was sufficiently valid such that it performed as well as or better than the proprietary test method on which the generic test guideline was based.

In order to address the issues relevant to both the regulatory agency needs and the OECD test guideline issues, an ICCVAM initiative was undertaken to develop and define the concept of test method performance standards that could be used for both situations. This paper will discuss the critical elements of performance standards and present the process by which ICCVAM develops and validates performance standards. An example of successful performance standards will also be presented.

Defining test method performance standards

ICCVAM defined performance standards as the basis on which a proprietary or non-proprietary test method has been determined to have sufficient accuracy and reliability for a specific testing purpose (ICCVAM, 2003). This process involves determination that a test method has sufficient accuracy and reliability for a defined specific testing purpose, and then using this information to develop performance standards that can be used

as the basis for evaluating the acceptability of proposed test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. These performance standards can then be used by regulatory authorities to communicate the basis by which they find test methods to be acceptable for specific regulatory testing purposes. ICCVAM now routinely develops and proposes performance standards during test method evaluations for both proprietary and non-proprietary methods that have undergone adequate validation (ICCVAM, 2003).

Components of performance standards

Performance standards have three critical components: essential test method components, a minimum list of reference chemicals, and defined accuracy and reliability values (ICCVAM, 2003). If a similar test method adequately addresses and meets these standards, then it would be considered to be comparable, in terms of performance, to the test method used to establish the performance standards.

Essential test method components

are the requisite structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed mechanistically and functionally similar test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. If there are deviations from the recommended essential test method components, then a scientific rationale must be provided and any potential impact of the deviations must be discussed. Incorporation of and adherence to essential test method components will help assure that a proposed test method is based on the same concepts as the corresponding validated test method.

The minimum list of reference chemicals

is used to assess the accuracy and reliability of a mechanistically and functionally similar test method that incorporates all of the essential test method components. These chemicals are a representative subset of those used to demonstrate the reliability and accuracy of the validated test method. To the extent possible, these reference chemicals should:

- Represent the range of responses that the validated test method is capable of measuring or predicting (e.g. negative and weak to moderate to strong positives)
- Produce consistent results in the validated test method and in the *in vivo* reference test method and/or target species of interest
- Reflect the accuracy of the validated test method
- Have well-defined chemical structures
- Be readily available, i.e. can be purchased from commercial sources
- Not be associated with excessive hazard or prohibitive disposal costs
- Represent the range of known or suspected mechanisms and/or modes of action for the toxicity measured or predicted by the test method
- Represent the range of physical and chemical properties which the test method is proposed to be capable of testing (e.g. solubility, pH, volatility, etc.)

These reference chemicals are the minimum number that should be used to evaluate the performance of a proposed mechanistically and functionally similar test method. These chemicals should not be used to develop the decision criteria/prediction model for the proposed test method. If any of the recommended chemicals are unavailable, other chemicals for which adequate reference data are available could be substituted with adequate scientific justification. To the extent possible, any substituted chemical(s) should be of the same chemical class and potency as the original chemical(s). If desired, additional chemicals representing other chemical or product classes and for which adequate reference data are available can be used to more comprehensively evaluate the accuracy of the proposed test method. However, these additional chemicals should not include those used to develop the proposed test method.

Accuracy and reliability values

are the comparable performance that should be achieved by the proposed test method when evaluated using the minimum list of reference chemicals. Reference chemicals should be designated for performance standards that will result in accuracy and reliability values similar to the overall values determined from the entire validation database for the reference test method.

Process for developing performance standards

ICCVAM has developed a process for establishing performance standards during the evaluation of proposed new test methods (ICCVAM, 2003). The process is designed to ensure rigorous scientific review and to provide the opportunity for broad stakeholder and public comment. The ICCVAM process for developing performance standards for new test methods is as follows:

- The National Toxicology Program Interagency Center for the Evaluation of Alternative Methods (NICEATM) and the appropriate ICCVAM working group develop proposed performance standards for consideration during the ICCVAM evaluation process. If a sponsor proposes performance standards, these are considered by ICCVAM at this stage. Generally, the proposed performance standards will be based on the information and data provided in the test method submission or on other available applicable data.
- The ICCVAM/NICEATM Peer Review Panel evaluates the proposed performance standards for completeness and appropriateness during its evaluation of the validation status of the proposed test method. The proposed performance standards are made available with the test method submission to the public for comment prior to and during the Peer Review Panel meeting.
- The appropriate ICCVAM working group, with the assistance of NICEATM, prepares the final performance standards for ICCVAM approval, taking into consideration the recommendations of the Peer Review Panel and public comments.
- Performance standards recommended by ICCVAM are incor-

porated into ICCVAM test method evaluation reports, which are published, provided to Federal agencies, and made available to the public. Availability of ICCVAM test method evaluation reports are announced routinely in the *Federal Register*, NTP Newsletters, and ICCVAM/NICEATM e-mail list serve groups.

- Regulatory authorities can then reference the performance standards in the ICCVAM report when they communicate their acceptance of a new test method. In addition, performance standards adopted by regulatory authorities can be provided in guidelines issued for new test methods.

Performance standards for dermal corrosivity test methods

Following the concept development and definition of performance standards, ICCVAM then used this framework to address the request by the USEPA to develop performance standards for the three proprietary dermal corrosivity test methods previously recommended by ICCVAM: Corrositex®, Episkin™, and Epiderm™ (ICCVAM 1999, 2002). In addition, and in accordance with the adopted process for developing performance standards, ICCVAM also developed performance standards for the one recommended non-proprietary dermal corrosivity test method, the rat skin transcutaneous electrical resistance (TER) method.

Due to the structural and functional differences of the four methods, three different sets of performance standards were developed (ICCVAM, 2004). Since Episkin™, and Epiderm™ are structurally and functionally similar, one set of performance standards was developed for these two methods. The standards were based on Episkin™, since this method had a larger validation database than Epiderm™ (60 vs. 24). In addition to the essential test method components, a minimum list of 24 reference chemicals were selected from the 60 chemicals used for the validation of Episkin™. This included 12 corrosives and 12 non-corrosives. All of the selected reference chemicals are commercially available. Accuracy and reliability values for the 24 minimum reference chemicals closely matched the overall performance for the 60 chemicals in the validation database. For the rat skin TER, a minimum list of 24 reference chemicals was selected, which also provided similar accuracy and reliability values as those for the total validation database of 60 chemicals.

Performance standards based on Corrositex® were developed for a generic *in vitro* membrane barrier test system for skin corrosion (ICCVAM, 2004). The essential test method components were included in the test method description for an OECD test guideline based on the validated Corrositex® (OECD, 2005b). This test method is capable of identifying the three subcategories of corrosivity described by the United Nations Packing Group (PG) classification system. Accordingly, the validation database contained a larger number of substances (129). The selected minimum list of reference chemicals contained a total of 40 chemicals, including 12 non-corrosive methods and 28 corrosive chemicals (9 UN PGI, 9 UN PGII and 10 UN PG III). As with the other methods, the accuracy and reliability values for the minimum list of reference chemicals were similar to those for the total validation database. These performance standards were subsequently included in the proposed OECD Test Guideline 435 for an *in vitro* membrane barrier test system for skin corrosion (OECD, 2005b). This is the first OECD test guideline to incorporate all three critical performance standard components.

Using performance standards for validation studies

The availability of performance standards is expected to significantly expedite the validation and acceptance of improved test methods that are similar to previously accepted methods.

Since validation can be a lengthy and expensive process, it is clearly advantageous to reduce the time and expense required to determine the usefulness and limitations of new test methods, especially those that may be structurally and functionally similar to test methods for which there has already been adequate validation. In fact, the idea that performance criteria could be used for this purpose has been previously proposed (Balls, 1997). Obviously the complete and full validation of structurally and functionally similar test methods could be expensive and potentially unnecessarily duplicative. For example, without performance standards, evaluation studies on a generic version of Corrositex®, which underwent validation with over 129 chemicals (ICCVAM, 1999), might be expected to use a similar number of chemicals. The establishment of performance standards that can be used to determine if similar test methods have comparable or better performance than one that has undergone extensive validation and been determined to be acceptable for a specific regulatory testing purpose will save considerable time and expense. The availability of performance standards should also facilitate the validation of improved versions of tests that may provide for better predictivity and efficiency. The concept and definition of test method performance standards has also recently been included in international guidance on validation (OECD, 2005a).

Conclusions

The establishment of test method performance standards now permits the regulatory acceptance of proprietary test methods and test methods with proprietary components. Test method performance standards have now been successfully developed for three different types of *in vitro* dermal corrosivity test methods. Performance standards have also now been incorporated into international test guidelines and international guidance on validation and regulatory acceptance. The continued development and availability of test method performance standards can be expected to streamline both the validation and acceptance processes for improved test methods that are mechanistically and functionally similar to previously accepted methods.

References

Balls, M. (1997). Editorial: Defined Structural and Performance Criteria would Facilitate the Validation and Acceptance of Alternative Test Procedures. *ATLA 25*, 483-484.

Environmental Protection Agency (1995). Ethics Advisory 95-18, Use of Public Office for Private Gain – Endorsements. Based on provisions of the "Standards of Ethical Conduct for Employees of the Executive Branch" at 5 CFR 2635.702 and on provisions of 5 CFR 2635.703. Washington, DC, USA: US Environmental Protection Agency.

ICCVAM (1997). Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), NIH Publication No. 97-3981, NIEHS, Research Triangle Park, NC, USA. (URL: http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf).

ICCVAM (1999). Corrositex®: An In Vitro Test Method for Assessing Dermal Corrosivity Potential of Chemicals, NIH Publication No. 99-4495, NIEHS, Research Triangle Park, NC, USA. (URL: http://iccvam.niehs.nih.gov/docs/reports/corprrep.pdf).

ICCVAM (2002). ICCVAM Evaluation of EPISKIN™, EpiDerm™, (EPI-200), and Rat Skin Transcutaneous Electrical Resistance (TER): In Vitro Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 02-4502. NIEHS, Research Triangle Park, NC, USA. (URL: http://iccvam.niehs.nih.gov/methods/epiderm.htm).

ICCVAM (2003). ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. NIEHS, Research Triangle Park, NC USA. (URL: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm).

ICCVAM (2004). ICCVAM Recommended Performance Standards for In Vitro Test Methods for Skin Corrosion. NIH Publication No. 04-4510. NIEHS, Research Triangle Park, NC USA May 2004. (URL: http://iccvam.niehs.nih.gov/docs/docs.htm#invitro).

OECD (2005a). Guidance Document No. 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Paris, France: Organization for Economic Cooperation and Development.

OECD (2005b). In Vitro Membrane Barrier Test Method for Skin Corrosion, Test Guideline 435. Organization for Economic Cooperation and Development, Paris, France. (URL: http://www.oecd.org).

U.S. Code of Federal Regulations (2005). "Standards of Ethical Conduct for Employees of the Executive Branch" at 5 CFR 2635.702. Washington, DC, USA: US Government Printing Office.

Acknowledgements

The authors wish to acknowledge the many designated agency representatives from the 15 member agencies of ICCVAM who contributed to the development and review of the ICCVAM guidance on performance standards, and members of the ICCVAM Dermal Corrosivity and Irritation Working Group who contributed to development of performance standards for *in vitro* methods for the *in vitro* corrosivity methods. Finally, the authors wish to acknowledge the expert assistance of Ms. Debbie McCarley in the preparation of this manuscript.

Correspondence to

Dr. William S. Stokes
National Institute of Environmental Health Sciences
Mail Drop EC-17, Post Office Box 12233
79 T. W. Alexander Drive
Research Triangle Park, North Carolina
USA 27709
e-mail: stokes@niehs.nih.gov