NIH Publication No: 99-4496

Evaluation of the Validation Status of Toxicological Methods:

General Guidelines for Submissions to ICCVAM

Prepared by the
Interagency Coordinating Committee on the
Validation of Alternative Methods (ICCVAM)
and the
National Toxicology Program (NTP) Interagency Center for the Evaluation
Of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
Nation Institutes of Health
US Public Health Service
Department of Health and Human Services

This document is a “work in progress” that reflects the current status of the guidelines for submitting toxicological methods for consideration by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). ICCVAM encourages comments and suggestions related to these guidelines. Comments should be forwarded to the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) via e-mail (niceatm@niehs.nih.gov) or by mail (NICEATM; NIEHS; MD-EC17; RTP, NC, 27709, USA).

Original Publication Date: March, 1997
Revised Publication Date: October, 1999

TABLE OF CONTENTS

ICCVAM Agency Representatives.....iii

Preface.....v

I. Introduction.....1

II. General Principles.....3

III. Submission Guidelines.....4

1.0 Introduction and Rationale for the Proposed Test Method.....4

2.0 Proposed Test Method Protocol.....4

3.0 Characterization of the Materials Tested.....5

4.0 Reference Data Used for Performance Assessment.....6

5.0 Test Method Data and Results.....7

6.0 Test Method Performance Assessment.....7

7.0 Test Method Reliability (Repeatability/Reproducibility).....8

8.0 Test Method Data Quality.....9

9.0 Other Scientific Reports and Reviews.....9

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement).....9

11.0 Other Considerations.....10

12.0 Supporting Materials.....10

IV. General Summary.....11

V. References.....12

Appendices

A. Outline of the General Guidelines for Submissions to ICCVAM.....A-1

B. Suggested Formats for Presenting Data.....B-1

C. Glossary.....C-1

D. Validation and Regulatory Acceptance Criteria.....D-1

INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS (ICCVAM)

AGENCY REPRESENTATIVES

Agency for Toxic Substances and Disease Registry (ASTDR)
Dr. William Cibulas

Consumer Product Safety Commission (CPSC)
Dr. Susan Aitken
Dr. Kailash Gupta
Dr. Marilyn Wind

Department of Defense (DOD)
Dr. Robert Finch
Dr. John M. Frazier
Dr. Harry Salem

Department of Energy (DOE)
Dr. Marvin Frazier

Department of the Interior (DOI)
Dr. Barnett A. Rattner

Food and Drug Administration (FDA)
Dr. William T. Allaben - NCTR
Dr. June A. Bradlaw – CFSAN
Dr. Joseph F. Contrera - CDER
Dr. Atin Datta – ORA/ORO
Dr. Joseph DeGeorge - CDER
Dr. Suzanne C. Fitzpatrick - ORA
Dr. David G. Hattan - CFSAN
Ms. Abby Jacobs - CDER
Dr. Devaraya Jagannath - CVM
Dr. Raju Kammula - CDRH
Mr. Louis T. Mulligan - CVM
Dr. Anita O'Connor – Office of Science
Dr. Anne M. Pilaro - CBER
Dr. Neil Sass - CFSAN
Dr. Leonard M. Schechtman - CDER
Dr. Melvin Stratmeyer - CDRH
Dr. Neil Wilcox – Office of Science

National Cancer Institute (NCI)
Dr. Victor A. Fung
Dr. David Longfellow

National Institute of Environmental Health Sciences (NIEHS)
Dr. John Bucher
Dr. Rajendra Chhabra
Dr. William S. Stokes (ICCVAM Co-Chair)
Dr. Errol Zeiger

National Institutes of Health (NIH) Office of the Director
Ms. Christina Blakeslee
Dr. Louis Sibal

National Institute for Occupational Safety and Health (NIOSH)
Dr. Doug Sharpnack
Dr. Kenneth Weber

National Library of Medicine (NLM)
Ms. Vera Hudson

Occupational Safety and Health Administration (OSHA)
Dr. Surender Ahir

U.S. Department of Transportation (US DOT)
Dr. George Cushmac
Mr. James K. O'Steen

U.S. Environmental Protection Agency (US EPA)
Dr. Angela Auletta
Dr. Karen Hamernik
Dr. Richard Hill (ICCVAM Co-Chair)
Dr. William R. Mundy
Dr. Hugh Tilson
Dr. Maurice Zeeman

PREFACE

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) was established as a standing collaborative effort by the National Institute of Environmental Health Sciences (NIEHS) and 13 other regulatory and research agencies. ICCVAM coordinates issues within the Federal government that relate to the development, validation, acceptance, and national/international harmonization of toxicological test methods. The Committee's functions include the coordination of interagency scientific reviews of toxicological test methods and communication with outside groups throughout the process. The following Federal regulatory and research agencies participate in this effort:

Consumer Product Safety Commission
Department of Defense
Department of Energy
Department of Health and Human Services
- Agency for Toxic Substances and Disease Registry
- Food and Drug Administration
- National Cancer Institute
- National Institute for Occupational Safety and Health/Centers for Disease Control
- National Institute of Environmental Health Sciences
- National Institutes of Health
- National Library of Medicine

Department of the Interior
Department of Labor
- Occupational Safety and Health Administration

Department of Transportation
- Research and Special Programs Administration

Environmental Protection Agency

Before a new or revised test method is used to generate information to support regulatory decisions, it must be (a) validated to determine its reliability and relevance for its proposed use, and (b) determined to be acceptable by one or more regulatory agencies to fill a specific need. Criteria for validation and regulatory acceptance have been prepared and are described in the report, *Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods* (1). Prior to the initiation of any test method development or validation efforts, sponsors are encouraged to consider the validation and acceptance criteria developed by the Federal government (1).

ICCVAM subsequently developed the present document to provide additional guidance to developers in organizing information needed to assess the validation status of a new or revised test method at any stage of development through validation. This document, *Evaluation of the Validation Status of Toxicological Methods: General Guidelines for Submissions to ICCVAM,* is available online (http://iccvam.niehs.nih.gov/doc1.htm), and additional copies can be obtained from the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). These guidelines call for the development of an ICCVAM submission on a given test method that addresses the extent to which the validation and acceptance criteria have been addressed, or how they will be addressed in proposed studies. The information provided in the submission should provide a comprehensive review of existing information and data for the method, and provide the basis for decisions on standardized protocols and the design of proposed studies. The submission should be prepared well in advance of any peer review of the validation status of a method.

Test method developers are encouraged to consult with NICEATM and ICCVAM during submission preparation and throughout test method development, prevalidation, and validation. The objective of the submission and these interactions is to maximize the likelihood that information generated will adequately characterize the usefulness and limitations of a test method and serve as a basis for assessing the status of validation through an independent ICCVAM peer review process. This process also enhances the likelihood that agencies can determine the extent that the method can generate information to meet their stated regulatory needs.

The initial submission guidelines, first released in May of 1998, were prepared by ICCVAM and incorporated much of the guidance developed for data submission for the 2nd workshop of the Interagency Regulatory Alternatives Group (2). It has now been updated by ICCVAM to reflect experience gained with the first two test methods, Local Lymph Node Assay (LLNA) and Corrositex®, that were reviewed by ICCVAM in 1998-99; further modifications are anticipated as experience accrues. ICCVAM welcomes suggestions on how to improve the usefulness of these guidelines.

We gratefully acknowledge the contributions of the ICCVAM agency representatives, ICCVAM working group members, and peer review panel members that assisted in the original preparation and subsequent revision of this document. We also appreciate the constructive suggestions received from scientists that used earlier versions of the guidelines to prepare submissions.

William S. Stokes, Co-Chair, ICCVAM
Richard N. Hill, Co-Chair, ICCVAM

I. INTRODUCTION

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) coordinates issues that relate to the development, validation, acceptance, and national/ international harmonization of toxicological test methods throughout the Federal government of the United States of America. The focus is on new and revised test methods that are applicable to multiple Federal agencies. Emphasis is given to test methods that provide for improved prediction of adverse human health or ecological effects, and that reduce[1], refine[2], or replace[3] animal use. In the report of the *ad hoc* ICCVAM, *Validation and Regulatory Acceptance of Toxicological Test Methods* (1), various stages were identified to move a proposed test method from concept to regulatory acceptance. A critical stage is the communication of a proposed test method by the sponsor to ICCVAM for consideration and review. The ICCVAM review process typically involves an assessment by an ICCVAM working group comprised of government scientists, followed by an independent peer review evaluation by an expert scientific panel. Following this review process, ICCVAM forwards recommendations on the usefulness and limitations of the proposed test method for regulatory purposes to regulatory agencies. Based on their specific regulatory mandates, each Federal agency then makes a determination regarding the acceptability of the test method. If the test method is accepted, appropriate actions (e.g., revision of existing regulations, guidelines, and/or guidance documents) are taken to inform the regulated community.

The purpose of this document is to provide guidance to sponsors on the information needed by ICCVAM to evaluate the validation status of a proposed test method. In preparation of their

[1] Reduction alternative: A new or revised test method that reduces the number of animals required.
[2] Refinement alternative: A new or revised test method that refines procedures to lessen or eliminate pain or distress to animals, or that enhances animal well-being.
[3] Replacement alternative: A new or revised test method that replaces animals with non-animal systems or one animal species with a phylogenetically lower one.

submission, sponsors should use the attached outline (**Appendix A**) to discuss the extent to which each of the validation and acceptance criteria have been addressed or will be addressed in proposed validation studies. Sponsors may be asked to provide additional information to augment or complement the information described in these guidelines.

Sponsors are encouraged to communicate with ICCVAM throughout the development and validation process. If requested, ICCVAM will solicit interagency comments on proposed study designs and the protocol. If the proposed test method is submitted for ICCVAM to provide comments on validation study design, the submission should describe the basis for the proposed protocol and the proposed validation studies.

To support the activities of ICCVAM, NIEHS established NICEATM. NICEATM staff and office are located at NIEHS and NICEATM serves as a communication link between test sponsors and Federal agencies during the development and validation process. NICEATM also coordinates independent scientific peer review, where appropriate and recommended by ICCVAM. Sponsors are encouraged to contact NICEATM (http://www.iccvam.niehs.nih.gov) prior to submission of proposed test methods for guidance on the submission and evaluation process.

II. GENERAL PRINCIPLES

Validation is a process designed to establish the operational characteristics of a proposed test method. These characteristics include the test method's reproducibility within and among laboratories, its relevance (i.e., its ability to measure or predict correctly), and its limitations. The submission must contain sufficient information for an independent scientific peer review panel to assess the validation status of the proposed test method and for agencies to assess the acceptability of the proposed test method for providing useful information for hazard or risk assessment. Criteria for validation and regulatory acceptance are published in the ICCVAM report (1), a workshop report of the Organization for Economic Co-operation and Development (OECD) (3), and a report by the European Center for the Validation of Alternative Methods (ECVAM) (4). The submission to ICCVAM should include all relevant data; information on materials, methods, statistical analyses, and conclusions; and appropriate literature citations. This material should be submitted in both electronic and printed format. The preferred software for electronic submission of text is Microsoft Word; databases are preferred in Microsoft Excel format. However, other electronic formats are acceptable.

The possible need for confidentiality of proprietary information is recognized by ICCVAM and such material must be appropriately designated. Submission of adequate and complete information will facilitate the review process. It should be emphasized that the amount and type of information needed to substantiate the usefulness of a test method for a specified purpose will vary depending on the test method and its proposed use. An outline to organize the information covered in this document is given in **Appendix A**.

III. SUBMISSION GUIDELINES

1.0 Introduction and Rationale for the Proposed Test Method

In this section, the sponsor must summarize and document the mechanistic basis of the proposed test method and the context in which the method will be used to measure or predict the toxicological activity of a test material or substance. When possible, what is known about the similarities and differences of modes and mechanisms of action in the test system compared to the species of interest (e.g., humans for human health-related toxicity testing) should be described. The sponsor should also discuss the use of the proposed test method in the context of current or anticipated regulatory applications (e.g., as a screen in a tiered testing strategy, as an adjunct test to provide mechanistic information, or as a substitute or replacement for an existing test method). The sponsor should indicate the relevant classes of chemicals and products that can and cannot be evaluated using the proposed test method. Finally, the sponsor should indicate where and how the proposed test method would be included in the overall safety or hazard assessment process. In particular, if the proposed test method is part of a tiered approach, the relative weight given to the new method relative to other tests in the tier should be addressed.

2.0 Proposed Test Method Protocol

This section should include a complete, detailed protocol for the proposed test method. The basis for any protocol modifications made during the validation of the test method should be provided. The technical parameters of the proposed test method (e.g., vehicles, exposure time), the nature of the response(s) evaluated, and the concurrent controls must be described clearly. Concurrent vehicle and, where appropriate, positive and negative controls provide a basis for experiment-to-experiment comparisons and are usually part of the acceptance criteria for a given experiment. The acceptable range for the control responses should be included.

The nature of the data to be collected, the methods used for data collection, the type of media in which data are stored, measures of variability, the statistical or non-statistical method(s) used to analyze/evaluate the data (including methods used to analyze for a dose-response relationship), and the decision criteria (and their rationale) used to classify the response as positive or negative must be described. The procedure for dose selection and the number of animals required, if any, for dose-selection and the actual test must be stated. Both the statistical and non-statistical methods used for data evaluation should be justified. Any confidential information associated with the test method should be clearly indicated. However, such designation is discouraged.

The number of replicate and/or repeat experiments needed to ensure an adequate study must be indicated and the basis for the design should be described. If replicate and/or repeat experiments are not part of the proposed test method protocol, a rationale for their exclusion must be provided.

The basis for the test system must be provided. If an animal model is used, the rationale for selecting the strain or stock, sex, acceptable age range, diet, frequency of dosing, the number of doses, and other applicable aspects should be included.

3.0 Characterization of the Materials Tested

The specific chemical or formulation names and relevant chemical and product classes for the materials tested must be specified. A test method may be more effective for the evaluation of certain classes of test chemicals. In addition, not all data sets will be homogeneous for a given chemical characteristic (e.g., water solubility). In such cases, it may be useful to separate the data set into smaller, more uniform subsets for data analysis. To the extent possible, information concerning physical and chemical characteristics; concentrations tested; purity and source of the test chemical; stability of the test material in the test medium; and the Chemical Abstracts Service (CAS) Registry number of the test chemical should be provided. Any

characteristics thought to have direct relevance to test performance should be clearly indicated. Information concerning coding of compounds during validation studies should be included. In the case of mixtures, the constituents and their relative concentrations should be stated, whenever possible. A suggested spreadsheet format for listing this information is provided in **Appendix B**. The rationale for selecting numbers and types of chemicals tested during the validation process should be described. Information regarding the use of coded chemicals and blind testing should be included.

4.0 Reference Data Used for Performance Assessment

If the proposed test method is intended to replace or substitute for an existing test method, then a comparison of data between the proposed test method and the reference test method is necessary. Such comparisons may provide insight into underlying toxicological mechanisms or processes. The submission should include a description of the protocol(s) employed to generate the reference data used to assess the performance characteristics of the proposed test method. Any modifications to the reference protocol(s) should be clearly stated for each data set, along with a discussion of the potential impact of these modifications. The submission should also include comparative test method data (reference and human data, where available) and data evaluation criteria. If possible, individual animal and human reference data should be provided. The source of the reference data, including the literature citation for published information, and the laboratory study director and year generated for unpublished data should be identified. A description of the quality of the reference data, including the extent of Good Laboratory Practices (GLP) compliance (5) and the use of coded test chemicals, should be described. In addition, the availability of original datasheets for the reference data should be indicated, and a summary of the availability and use of relevant toxicity information from the species of interest (e.g., data from human studies or accidental exposures for human health-related toxicity test methods) should be included.

5.0 Test Method Data and Results

This section should include the complete, detailed protocol(s) used to generate the proposed test method data provided in the submission. Any deviations should be indicated, including the rationale for the deviation. Any protocol modifications made during the development process and their impact should be clearly stated for each data set. All data, both original and derived, should be submitted along with each laboratory's summary judgment as to the outcome of each test. The submission should include data (and explanations) from unsuccessful, as well as successful, experiments. The statistical approach used to evaluate the data should be described and justified.

It is also important to note the lot-to-lot consistency of the test chemicals, the time frame of the various studies, and the laboratory in which the study or studies were conducted. A coded designation for each laboratory is acceptable. Any original data not submitted should be available for review, if requested. The importance of presenting all available data, including appropriate data from published sources, cannot be over-emphasized. Such information is essential for an adequate scientific assessment of the method.

Results may be presented in graph or tabular form for easy comparison of results from the reference tests with those from the proposed test method. A suggested tabular format for presenting the results for use in performance assessment (see Section 6.0) is provided in **Appendix B**.

6.0 Test Method Performance Assessment

The data submission must include a statement of the performance of the proposed test method with respect to its ability to measure or predict the effect of interest. The performance (i.e., accuracy, sensitivity, specificity, positive and negative predictivity, and false positive and

negative rates) of the proposed test method should be compared to the reference test method currently accepted by regulatory agencies and to data or recognized toxicity information from the species of interest (e.g., humans for human health-related toxicity testing). The bases, where known, for any discordance between the results in the proposed test method and the reference test method and/or data from the species of interest should be presented and interpreted. 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 information from the species of interest. When the proposed test method is discordant from the reference test method, the frequency of correct predictions of each test method compared to recognized toxicity information from the species of interest should be described. The submission should include a discussion of the strengths and limitations of the proposed test method and should describe salient issues of data interpretation.

7.0 Test Method Reliability (Repeatability/Reproducibility)

An assessment of test method reliability must be provided, including discussion of the selection rationale for the chemicals used to evaluate the intra- and inter-laboratory reproducibility, and the extent to which they represent the range of possible test outcomes. Outlying values should be identified and discussed. A quantitative statistical analysis of the extent of intra- and inter-laboratory variability, such as that described in ASTM E691-92 (6) or coefficient of variation analysis, is essential. The number of trials, as well as the measures of central tendency and variation, should be summarized for the historical vehicle, positive, and negative control data.

8.0 Test Method Data Quality

The extent of adherence to national/international GLP guidelines (5) for the data presented in the submission, as well as the results of any data quality audits, must be stated. Deviations from GLPs and the impact of any non-compliance detected in audits should be described.

9.0 Other Scientific Reports and Reviews

The submission should include and discuss all data from other published or unpublished studies conducted using the proposed test method. Comment should be provided on any conclusions presented in independent peer-reviewed reports or other scientific reviews of the test method. The conclusions of such scientific reports and/or reviews should be compared to the conclusions reached in the submission. Any ongoing evaluations of the proposed method should be described.

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement)

A description should be provided as to how the proposed test method will refine, reduce, or replace animal use as compared to current methods used for the endpoint of interest. If the test method requires the use of animals, the rationale should be provided. A description of the sources used to determine the availability of alternative methods that would refine, reduce, or replace animal use for the endpoint of interest should be provided (7, 8). The description should include, at a minimum, the databases searched, the search strategy, the search date(s), the database search results, and the rationale for not utilizing available alternative methods. The basis for determining the appropriate number of animals for the proposed test method should be described. If the test involves potential animal pain and distress, the procedures and approaches that have been incorporated to minimize, and whenever possible, to eliminate the occurrence of such pain and distress should be discussed.

11.0 Other Considerations

Test method transferability and the cost and time involved in conducting the proposed test should be described and compared to the reference test method(s). With regard to test method transferability, the submission should include a discussion of the facilities and major fixed equipment needed to conduct the test, the general availability of other necessary equipment and supplies, and the required level of training and expertise needed by the study personnel. Both the cost and time involved in conducting the proposed test method by trained personnel should be clearly specified and compared to the cost and time involved in conducting the reference test method, when possible.

12.0 Supporting Materials

Copies of all relevant publications, including those containing data from the proposed test method or the reference test method must be included, along with all available non-transformed original data used to evaluate the validity of the proposed test method. The results of any peer reviews should be provided and summarized. Any ongoing or planned reviews should also be described. The availability of laboratory notebooks and other data retained by the sponsor(s) for external audits by ICCVAM must be stated. Unpublished data should be supported by laboratory notebooks.

IV. GENERAL SUMMARY

This document is intended to provide guidance to sponsors of test methods proposed to ICCVAM for use as new or revised tests in toxicological studies. A general outline for organizing the submitted information is provided in **Appendix A**. It is strongly recommended that the sponsor of a proposed test method contact NICEATM for guidance before submitting the test method for review, and that the points in **Appendix A** be specifically addressed in the submission to ICCVAM. Sponsors are encouraged to consider and describe how these guidelines will be adequately addressed.

Inquiries should be directed to Dr. W.S. Stokes, NICEATM, National Institutes of Health, NIEHS, MD EC-17, P.O. Box 12233, Research Triangle Park, NC 27709, 919-541-7997 (phone); 919-541-0947 (FAX); iccvam@niehs.nih.gov (e-mail).

V. REFERENCES

· National Institute of Environmental Health Sciences (NIEHS). 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. Research Triangle Park, North Carolina: NIEHS, March, 1997, 105 pp. Available on the Internet at http://ntp-server.niehs.nih.gov/htdocs/ICCVAM/iccvam.html.

· Scala, R.A., and Springer, J. IRAG Working Group 6: Guidelines for the evaluation of eye irritation alternative tests: Criteria for data submission. Food. Chem. Toxicol. 35(1), 13-22, 1997.

· Organization for Economic Co-operation and Development (OECD). Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Methods. Paris: OECD, 1996, 60 pp.

· Balls, M., B.J. Blaauboer, J.H. Fentem, L. Bruner, R.D. Combes, B. Ekwall, R.J. Fielder, A. Guillouzo, R.W. Lewis, D.P. Lovell, C.A. Reinhardt, G. Repetto, D. Sladowski, H. Spielmann, and F. Zucco. Practical aspects of the validation of toxicity test procedures. The report and recommendation of ECVAM workshop 5. ATLA 23, 884-886, 1995.

· Good Laboratory Practices (GLP):

a. EPA Federal, Insecticide, and Rodenticide Act (FIFRA); Good Laboratory Practice Standards; Final Rule. 40 CFR Part 160. Washington, D.C.: U.S. Government Printing Office, 1998.

b. EPA Toxic Substances Control Act (TSCA); Good Laboratory Practice Standards; Final Rule. 40 CFR Part 792. Washington, D.C.: U.S. Government Printing Office, 1998. http://www.epa.gov/fedrgstr/EPA-TOX/1997/December/Day-24/t33451.htm

c. FDA Good Laboratory Practice for Nonclinical Laboratory Studies. 21 CFR Part 58. Washington, D.C.: U.S. Government Printing Office, 1999. http://iccvam.niehs.nih.gov/fda-glp.htm

d. OECD Good Laboratory Practice Guidelines, including the following titles: Quality Assurance and GLP, Compliance of Laboratory Suppliers with GLP Principles, the Application of the GLP Principles to Field Studies, the Application of the GLP Principles to Short-Term Studies, the Role and Responsibilities of the Study Director in GLP Studies, and the Application of the GLP Principles to Computerised Systems. Available on the Internet at http://www.oecd.org/ehs/ehsmono/#GLP.

e. Japanese Good Laboratory Practice Standards including: (1) Pharmaceutical GLP Standard, Pharmaceutical Affairs Bureau, Ministry of Health and Welfare, and (2) GLP Standard, Japanese Ministry of Agriculture, Forestry, and Fisheries. Both are available on the Internet at the Japan Society of Quality Assurance website, http://www.jsqa.com/english/home-e.htm.

6. American Society for Testing and Materials (ASTM). Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. ASTM E691-92. In: Annual Book of ASTM Standards. Philadelphia: ASTM, 1992, 19 pp.

7. USDA Animal Welfare Act; 7 USC 2131-2156. Washington, D.C.: U.S. Government Printing Office, 1966. http://warp.nal.usda.gov:80/awic/legislat/awa.htm

8. Public Health Service. Public Health Service Policy on Humane Care and Use of Laboratory Animals. Washington, D.C.: U.S. Department of Health and Human Services, 1996. http://grants.nih.gov/grants/oprr/phspol.htm

APPENDIX A: OUTLINE OF GENERAL GUIDELINES FOR SUBMISSION TO ICCVAM

1.0 Introduction and Rationale for the Proposed Test Method

1.1 Scientific basis for the proposed test method.

1.1.1 Describe the purpose, including the mechanistic basis, of the proposed test method.

1.1.2 When possible, describe what is known about the similarities and differences of modes and mechanisms of action in the test system as compared to the species of interest (e.g., humans for human health-related toxicity testing).

1.2 Intended uses of the proposed test method:

1.2.1 Describe the intended regulatory use(s) (e.g., screen, substitute, replacement, or adjunct) and provide a rationale.

1.2.2 Where appropriate, indicate how the proposed method can substitute, replace, or complement an existing test method(s).

1.2.3 Describe how the method fits into the overall strategy of hazard or safety assessment. If a component of a tiered assessment process, indicate the weight that should be applied relative to other measures.

1.2.4 Describe the intended range of materials amenable to the test and/or the limits of the proposed test method according to chemical class or physico-chemical factors.

2.0 Proposed Test Method Protocol

2.1 Provide the detailed protocol for the proposed test method, including the:

2.1.1 Materials, equipment, and supplies needed;

2.1.2 Detailed procedures for conducting the test;

2.1.3 Dose-selection procedures, including the need for any dose range-finding studies or acute toxicity data prior to conducting the test, if applicable;

2.1.4 Endpoint(s) measured;

2.1.5 Duration of exposure;

2.1.6 Known limits of use;

2.1.7 Nature of the response assessed;

2.1.8 Appropriate vehicle, positive, and negative controls and the basis for their selection;

2.1.9 Acceptable range of vehicle, positive and negative control responses;

2.1.10 Nature of the data to be collected and the methods used for data collection;

2.1.11 Type of media in which data are stored;

2.1.12 Measures of variability;

2.1.13 Statistical or non-statistical method(s) used to analyze the resulting data (including methods to analyze for a dose-response relationship). The method(s) employed should be justified and described;

2.1.14 Decision criteria or the prediction model used to classify a test chemical (e.g., positive, negative, or equivocal), as appropriate;

2.1.15 Information that will be included in the test report.

2.2 Explain the basis for selection of the test system. If an animal model is being used, this should include the rationale for selecting the strain or stock, sex, acceptable age range, diet, and other applicable aspects.

2.3 Any confidential information associated with the test method should be clearly indicated. However, such designation is discouraged.

2.4 Explain the basis for the decision criteria established for the test.

2.5 Describe the basis for the number of replicate and repeat experiments; provide the rationale if studies are not replicated or repeated.

2.6 Discuss the basis for any modifications to the proposed protocol that were made based on results from validation studies.

3.0 Characterization of Materials Tested (See Appendix B)

3.1 Describe the rationale for the chemicals/products selected for evaluation. Include information on the suitability of chemicals selected for testing, indicating any chemicals that were found to be unsuitable.

3.2 Discuss the rationale for the number of chemicals that were tested.

3.3 Describe the chemicals/products evaluated, including the:

3.3.1 Chemical or product name; if a mixture, provide information on all components;

3.3.2 CAS Registry number(s);

3.3.3 Chemical and product classes;

3.3.4 Physical/chemical characteristics (e.g., water and lipid solubility, pH, pKa, etc.). Any characteristics thought to have direct relevance to test performance should be clearly indicated;

3.3.5 Stability of the test material in the test medium;

3.3.6 Concentrations tested;

3.3.7 Purity, including the presence and identity of contaminants and stabilizing additives;

3.3.8 Supplier/source.

3.4 If mixtures were tested, constituents and relative concentrations should be provided whenever possible.

3.5 Describe coding used in the validation studies.

4.0 Reference Data Used for Performance Assessment

4.1 Provide a clear description of the protocol used for the reference test method. If a specific guideline has been followed, it should be provided. Any deviations should be indicated, including the rationale for the deviation.

4.2 Provide the reference test method data used to assess the performance of the proposed test method. Individual human and/or animal reference test data, if available, should be provided. Provide the source of the reference data, including the literature citation for published data, or the laboratory study director and year generated for unpublished data.

4.3 Indicate if original datasheets are available for the reference data.

4.4 Indicate the quality of the reference test data, including the extent of GLP compliance and any use of coded chemicals.

4.5 Discuss the availability and use of relevant toxicity information from the species of interest (e.g., human studies and or reported toxicity from accidental/ occupational exposure for human health-related toxicity testing).

5.0 Test Method Data and Results

5.1 Provide the complete, detailed protocol used to generate each set of data for the proposed test method. Any deviations should be described, including the rationale or explanation for the deviation. Any protocol modifications made during the development process and their impact should be clearly stated for each data set.

5.2 Provide all data obtained using the proposed test method. This should include copies of original data from individual animals and/or individual samples, as well as derived data. The laboratory's summary judgement as to the outcome of each test should be indicated. The submission should include data (and explanations) from unsuccessful, as well as successful, experiments.

5.3 Describe the statistical approach used to evaluate the data resulting from the proposed test method.

5.4 Provide a summary, in graphic or tabular form, of the results. The suggested tabular format for providing data for use in performance assessment is provided in **Appendix B**.

5.5 For each set of data, indicate whether coded chemicals were tested, whether experiments were conducted blind, and the extent to which experiments followed GLP procedures.

5.6 Indicate the lot-to-lot consistency of the test materials, the time frame of the various studies, and the laboratory in which the study or studies were done. A coded designation for each laboratory is acceptable.

5.7 Any data not submitted should be available for external audit, if requested.

6.0 Test Method Performance Assessment

6.1 Describe performance characteristics (e.g., accuracy, sensitivity, specificity, positive and negative predictivity, and false positive and negative rates) of the proposed test method compared with the reference test method currently accepted by regulatory agencies for the endpoint of interest. Explain how discordant results in the same or multiple laboratories from the proposed test were considered when calculating performance values.

6.2 Discuss results that are discordant with results from the reference method.

6.3 Discuss the performance characteristics of the proposed test method compared to data or recognized toxicity from the species of interest (e.g., humans for human health-related toxicity testing), where such data or toxicity classification are available. This is essential when the method is measuring or predicting an endpoint for which there is no pre-existing method. In instances where the proposed test method was discordant from the reference test method (6.2 above), describe the frequency of correct predictions of each test method compared to recognized toxicity information from the species of interest.

6.4 State the strengths and limitations of the method, including those applicable to specific chemical classes or physical-chemical properties.

6.5 Describe the salient issues of data interpretation, including why specific parameters were selected for inclusion.

7.0 Test Method Reliability (Repeatability/Reproducibility)

7.1 Discuss the selection rationale for the chemicals used to evaluate intra- and inter-laboratory reproducibility, and the extent to which the chosen set of chemicals represents the range of possible test outcomes.

7.2 Provide analyses and conclusions reached regarding inter- and intra-laboratory repeatability and reproducibility. Acceptable methods of analyses include those described in ASTM E691-92 (6) or by coefficient of variation analysis.

7.3 Summarize historical positive and negative control data, including number of trials, measures of central tendency, and variability.

8.0 Test Method Data Quality

8.1 State the extent of adherence to national/international Good Laboratory Practice (GLP) guidelines (5) for all submitted data, including that for the proposed test method and the reference test method. Information regarding the use of coded chemicals and blind testing should be included.

8.2 Summarize the results of any data quality audits, if conducted.

8.3 Discuss the impact of deviations from GLPs or any non-compliance detected in the data quality audits.

9.0 Other Scientific Reports and Reviews

9.1 Include all data from other published or unpublished studies conducted using the proposed test method.

9.2 Comment on and compare the conclusions published in independent peer-reviewed reports or other independent scientific reviews of the test method. The conclusions of such scientific reports and/or reviews should be compared to the conclusions reached in this submission. Any ongoing evaluations of the method should be described.

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement)

10.1 Describe how the proposed test method will refine (reduce pain or distress), reduce, and/or replace animal use compared to the current methods used.

10.2 If this test method requires the use of animals, address the following:

10.2.1 Describe the rationale for the need to use animals and describe why the information provided by this test requires the use of animals (i.e., cannot be obtained using non-animal methods).

10.2.2 Include a description of the sources used to determine the availability of alternative methods that might further refine, reduce, or replace animal use for this testing. This should, at a minimum, include the databases searched, the search strategy used, the search date(s), a discussion of the results of the search, and the rationale for not incorporating available alternative methods.

10.2.3 Describe the basis for determining that the numbers of animals being used are appropriate.

10.2.4 If the test method involves potential animal pain and distress, discuss the methods and approaches that have been incorporated to minimize, and whenever possible, eliminate the occurrence of such pain and distress.

11.0 Other Considerations

11.1 Discuss the following aspects of test method transferability. Include an explanation of how this compares to the transferability of the reference test method.

11.1.1 Discuss the facilities and major fixed equipment needed to conduct the test.

11.1.2 Discuss the required level of training and expertise needed for personnel to conduct the test.

11.1.3 Discuss the general availability of other necessary equipment and supplies.

11.2 Discuss the cost involved in conducting the test. Discuss how this compares to the cost of the reference test method.

11.3 Indicate the amount of time needed to conduct the test and discuss how this compares with the reference test method.

12.0 Supporting Materials

12.1 Provide copies of all relevant publications, including those containing data from the proposed test method or the reference test method.

12.2 Include all available non-transformed original data for both the proposed test method and the reference test method.

12.3 Summarize and provide the results of any peer reviews conducted to date, and summarize any ongoing or planned reviews.

12.4 Address the availability of laboratory notebooks or other records for an independent audit. Unpublished data should be supported by laboratory notebooks.

APPENDIX B: SUGGESTED FORMATS FOR PRESENTING DATA

Characterization of Materials Tested

In addition to a written description of the materials tested, presentation in the following table format is recommended. This information should be provided in written and electronic format (e.g., Microsoft Word or Excel are preferred, but other programs are acceptable).

Chemical/ Product Name	CAS No.	Chemical Class	Product Class	Concentration(s) Tested	Purity	Supplier/ Source of Compound	Physical/ Chemical Characteristics

Test Method Performance Assessment

The following format is suggested for presenting the information used in the performance assessment. Additional detailed information should also be provided in tabular or written format as described in Sections D and E of these guidelines. This information should be provided in written and electronic format (e.g., Microsoft Word or Excel are preferred, but other programs are acceptable).

Chemical/ Product Name	CASRN	Chemical Class	Product Class	Result Using Proposed Test (quanti-tative)	Call Using Proposed Test (+/-)	Reference Test Result (quanti-tative)[1]	Call Using Reference Test (+/-)	References/ Data Sources	Comments

[1] Where possible, data from the reference test should be separated into single columns for each species with available information. Human data should always be presented independently of nonhuman data.

APPENDIX C: GLOSSARY[1]

Accuracy: (a) The closeness of agreement between a test result and an accepted reference value. (b) The proportion of correct outcomes of a method. Often used interchangeably with concordance (see two-by-two table).

Adjunct test: A test that provides information that adds to or helps interpret the results of other tests, and provides information useful for the risk assessment process.

Assay: The experimental system used. Often used interchangeably with test.

Coded chemicals: Chemicals labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded chemicals are used to avoid intentional or unintentional bias when evaluating laboratory performance or performance of test methods.

Concordance: The proportion of all chemicals tested that are correctly classified as positive or negative. Often used interchangeably with accuracy (see two-by-two table). A measure of test performance. The concordance is highly dependent on the prevalence of positives in the population being examined.

Dose-response assessment: That part of risk assessment associated with evaluating the relationship between the dose of an agent administered or received and the incidence and/or severity of an adverse health or ecological effect.

Endpoint: The biological or chemical process, response, or effect assessed by a test method.

False positive: A nonactive substance incorrectly identified as positive by a test.

[1] From: 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 97-3981.

False positive rate: The proportion of all negative (inactive) substances that are falsely identified as positive (see two-by-two table). An indication of test performance.

False negative: An active substance incorrectly identified as negative by a test.

False negative rate: The proportion of all positive (active) substances falsely identified as negative (see two-by-two table). An indication of test performance.

Good Laboratory Practices (GLPs): Regulations promulgated by the FDA, EPA, OECD, and Japanese authorities that describe recordkeeping and quality assurance procedures for laboratory records that will be the basis for data submissions to the agencies. .

Hazard: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Interlaboratory reproducibility: A measure of whether different qualified laboratories using the same protocol and test chemicals can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test can be successfully transferred among laboratories.

Intralaboratory repeatability: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Intralaboratory reproducibility: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

Mechanistically based methods: Methods that provide a direct relationship between the biological effects observed with the biological effects of interest.

Prediction model: A formula or algorithm used to convert the results from a test method into a prediction of the toxic effect of interest. A prediction model contains four elements: a definition of the specific purpose(s) for which the test is to be used, a definition of all possible results that may be obtained, an algorithm that converts each test result into a prediction of the toxic effect of interest, and an indication of the accuracy of the prediction.

Predictivity (negative): The proportion of correct negative responses among materials testing negative (see two-by-two table). A measure of test performance. Negative predictivity is a function of the sensitivity of the test and the prevalence of negatives among the chemicals tested.

Predictivity (positive): The proportion of correct positive responses among materials testing positive (see two-by-two table). A measure of test performance. Positive predictivity is a function of the sensitivity of the test and the prevalence of positives among the chemicals tested.

Prevalence: The proportion of positives in the population of agents tested (see two-by-two table).

Prevalidation: The process during which a standardized test protocol is evaluated for use in validation studies. Based on the outcome of those studies, the test protocol may be modified or optimized for use in further validation studies.

Protocol: The precise step-by-step description of a test, including the listing of all necessary reagents and all criteria and procedures for generating and evaluating test data.

Quality assurance: A management process by which adherence to laboratory testing standards, requirements, and recordkeeping procedures is assessed independently by individuals other than those performing the testing.

Reduction alternative: A new or revised test method that reduces the number of animals required.

Reference chemicals: Chemicals selected for use in the validation process for which data exist from the reference test method or species of interest. These chemicals should be representative of the classes of chemicals for which the test is expected to be used and should represent different levels of expected responses. Different sets of reference chemicals may be required for the different stages of the validation process, and for different types of tests.

Reference species: The species used in the traditional test method to which a new or revised test is being compared. This may be the target species when it is also the species of interest, or it may be a surrogate species when it is not possible to perform testing on the target species.

Refinement alternative: A new or revised test method that refines procedures to lessen or eliminate pain or distress to animals, or enhances animal well-being.

Relevance: The extent to which the proposed test is related to the effect of interest and whether a test is meaningful and useful for a particular purpose. The extent to which a test method will correctly predict or measure the biological effect of interest.

Reliability: A measure of the degree to which a test can be performed reproducibly within and among laboratories over time.

Replacement alternative: A new or revised test method that replaces animals with non-animal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility: The variability between single test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test samples (see intra- and interlaboratory reproducibility).

Risk: The probability or degree of concern that an agent will cause an adverse effect given some exposure.

Screen/screening test: A rapid, simple test conducted for the purposes of a general classification of substances according to general categories of hazard. The results of a screen are generally used for preliminary decision making and to set priorities for more definitive tests. A screening test may have a truncated response range (e.g., be able to reliably identify active chemicals but not inactive chemicals).

Sensitivity: The proportion of all positive chemicals that are correctly classified as positive in a test. A measure of test performance (see two-by-two table).

Specificity: The proportion of all negative chemicals that are correctly classified as negative in a test. A measure of test performance (see two-by-two table).

Standard operating procedures (SOPs): Formal, written procedures that describe how specific laboratory operations are to be performed. Required by GLPs.

Substitute method: A new or revised test method proposed for use in lieu of a currently used method, regardless of whether that method is for a definitive, screening or adjunct test.

Test: The experimental system used; used interchangeably with assay.

Test method: A process or procedure used to obtain information on the characteristics of a chemical or agent. Toxicological test methods generate information regarding the ability of a chemical or agent to produce a specified biological effect under specified conditions. Used interchangeably with test and assay.

Transferability: The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two (2x2) table: The two-by-two table can be used for calculating accuracy (concordance) ([a+d]/[a+b+c+d]), negative predictivity (d/[c+d]), positive predictivity (a/[a+b]), prevalence ([a+c]/[a+b+c+d]), sensitivity (a/[a+c]), specificity (d/[b+d]), false positive rate (b/[b+d]), and false negative rate (c/[a+c]).

		New Test Outcome		
		Positive	Negative	Total
Reference Test Classification	Positive	a	c	a+c
	Negative	b	d	b+d
	Total	a+b	c+d	a+b+c+d

Validation: The process by which the reliability and relevance of a procedure are established for a specific purpose.

APPENDIX D: VALIDATION AND REGULATORY ACCEPTANCE CRITERIA

Validation Criteria[1]

For a new or revised test method to be considered validated for regulatory risk assessment purposes, it should generally meet the following criteria (the extent to which these criteria are met will vary with the method and its proposed use). However, there needs to be flexibility in assessing a method given its purpose and the supporting database. Because tests can be designed and used for different purposes by different organizations and for different categories of substances, the determination of whether a specific test method is considered by an agency to be useful for a specific purpose must be made on a case-by-case basis. Validation of a test method is a prerequisite for it to be considered for regulatory acceptance.

- The scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.

- The relationship of the test method's endpoint(s) to the biologic effect of interest must be described. Although the relationship may be mechanistic or correlative, tests with biologic relevance to the toxic process being evaluated are preferred.

- A detailed protocol for the test method must be available and should include a description of the materials needed, a description of what is measured and how it is measured, acceptable test performance criteria (e.g., positive and negative control responses), a description of how data will be analyzed, a list of the species for which the test results are applicable, and a description of the known limitations of the test including a description of the classes of materials that the test can and cannot accurately assess.

- The extent of within-test variability, and the reproducibility of the test within and among laboratories must have been demonstrated. Data must be provided describing the level of intra- and interlaboratory reproducibility and how it varies over time. The degree to which biological variability affects this test reproducibility should be addressed.

- The test method's performance must have been demonstrated using reference chemicals or test agents representative of the types of substances to which the test method will be applied, and should include both known positive and known negative agents. Unless it is hazardous to do so, chemicals or test agents should be tested under code to exclude bias.

[1] NIEHS (National Institute of Environmental Health Sciences). 1997. Validation and regulatory acceptance of toxicological methods: A report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. NIEHS, Research Triangle Park, NC.

- Sufficient data should be provided to permit a comparison of the performance of a proposed substitute test with that of the test it is designed to replace. Performance should be evaluated in relation to existing relevant toxicity testing data, and relevant toxicity information from the species of concern. Reference data from the comparable traditional test method should be available and of acceptable quality.

- The limitations of the method must be described; for example, *in vitro* or other non-animal test methods may not replicate all of the metabolic processes relevant to chemical toxicity that occur *in vivo*.

- Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with Good Laboratory Practices (GLPs). Aspects of data collection not performed according to GLPs must be fully described, along with their potential impact.

- All data supporting the assessment of the validity of the test method must be available for review.

- Detailed protocols should be readily available and in the public domain.

- The method(s) and results should be published or submitted for publication in an independent, peer-reviewed publication.

- The methodology and results should have been subjected to independent scientific review.

Regulatory Acceptance Criteria[2]

Validated methods are not automatically accepted by regulatory agencies; they need to fit into the regulatory structure. Flexibility is essential in determining the acceptability of methods to ensure that appropriate scientific information is considered in regulatory risk assessment. A test method proposed for regulatory acceptance generally should be supported by the following attributes:

- The method should have undergone independent scientific peer review by disinterested persons who are experts in the field, knowledgeable in the method, and financially unencumbered by the outcome of the evaluation.

- There should be a detailed protocol with standard operating procedures (SOPs), a list of operating characteristics, and criteria for judging test performance and results.

[2] NIEHS (National Institute of Environmental Health Sciences). 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. NIEHS, Research Triangle Park, NC.

- Data generated by the method should adequately measure or predict the endpoint of interest and demonstrate a linkage between either the new test and an existing test, or the new test and effects in the target species.

- There should be adequate test data for chemicals and products representative of those administered by the regulatory program or agency and for which the test is proposed.

- The method should generate data useful for risk assessment purposes, (i.e., for hazard identification, dose-response assessment, and/or exposure assessment). Such methods may be useful alone or as part of a battery or tiered approach.

- The specific strengths and limitations of the test must be clearly identified and described.

- The test method must be robust (relatively insensitive to minor changes in protocol) and transferable among properly equipped and staffed laboratories.

- The method should be time and cost effective.

- The method should be one that can be harmonized with similar testing requirements of other agencies and international groups.

- The method should be suitable for international acceptance.

- The method must provide adequate consideration for the reduction, refinement, and replacement of animal use.