EXECUTIVE SUMMARY

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recently completed the technical evaluation of the validation status of four *in vitro* ocular irritation test methods proposed as screening tests[2] for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy[3], as part of a weight-of-evidence approach. The four test methods are the Bovine Corneal Opacity and Permeability (BCOP) assay, the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay, the Isolated Chicken Eye (ICE) assay, and the Isolated Rabbit Eye (IRE) assay. The U.S. Environmental Protection Agency (EPA) formally nominated these test methods for evaluation by ICCVAM in October 2003. In addition to evaluating their current usefulness and limitations as screening tests for identifying ocular corrosives and severe irritants, ICCVAM developed a recommended standardized protocol for each test method; made recommendations, where considered appropriate, for further research and development, optimization, and/or validation efforts; and developed a list of reference substances for such activities.

None of the four *in vitro* test methods evaluated can be considered to be replacements for the *in vivo* rabbit eye test. However, based on the available data, BCOP and ICE can be used, in appropriate circumstances and with certain limitations, as screening tests for the detection of ocular corrosives and severe irritants in a tiered-testing strategy, as part of a weight-of-evidence approach. At the present time, HET-CAM, using the decision criteria of Luepke (1985), and IRE are not recommended as screening tests for the identification of ocular corrosives and severe irritants for regulatory hazard classification purposes. Before HET-CAM and IRE can be recommended for this purpose, the protocol and the decision criteria for the identification of ocular corrosives and severe irritants need to be optimized and undergo further validation.

This evaluation provides validation information that should be helpful to various stakeholders (e.g., applicable U.S. Federal regulatory agencies; the international regulatory community; the pharmaceutical, pesticide, and commercial chemical industries) in determining when these test methods might be useful and which test method might be the most appropriate for a specific testing situation. These *in vitro* test methods, when used appropriately, will reduce and refine animal use for ocular safety testing.

---

[2]According to the *ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods*, a **screen** or **screening test** is "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 generally are 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)" (ICCVAM 2003).

[3]A tiered-testing strategy approach may not be applicable to purposes other than regulatory classification and labeling.

Specific Test Method Recommendations

BCOP Test Method

There are sufficient data to support the use of the BCOP test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, United Nations [UN] Globally Harmonized System of Classification and Labelling of Chemicals [GHS] Category 1, European Union [EU] R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations for this test method are based on the false negative and false positive rates observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols and solids range from 67% (2/3)[4] to 100% (2/2) and 42% (5/12) to 50% (5/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols, ketones, and solids range from 50% (7/14) to 56% (9/16), 40% (4/10), and 10% (2/20 to 2/21), respectively, depending on the hazard classification system. When substances within these chemical and physical classes are excluded from the database, the accuracy of BCOP across the EU, EPA, and GHS classification systems ranges from 87% (72/83) to 92% (78/85) and the false negative and false positive rates range from 0% (0/27) to 12% (3/26) and 12% (7/58) to 16% (9/56), respectively.

Intralaboratory repeatability of *In Vitro* Irritancy Scores was assessed by analyzing two studies. In the first study, the median coefficient of variation (CV) for *In Vitro* Irritancy Scores for replicate corneas (evaluated in three laboratories) ranged from 11.8% to 14.2%. In the second study, the median CV value for *In Vitro* Irritancy Scores for replicate corneas was 35%.

Intralaboratory reproducibility evaluations indicated mean and median CV values for permeability values were 33.4% and 29.0%, respectively, for 25 surfactant-based personal care cleaning formulations in one study. Mean CV values of *In Vitro* Irritancy Scores for 16 substances tested two or more times in three laboratories ranged from 12.6% to 14.8%, while the median CV values ranged from 6.7% to 12.4%.

In a qualitative assessment of interlaboratory reproducibility of hazard classification category, 67% to 94% of the substances were classified the same by the participating laboratories. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as solvents, surfactants, chemical intermediates, and pesticides.

[4]The numbers in parentheses represent the numbers used to calculate the percentages. For the false negative or false positive rates, the numerators represent the total number of substances incorrectly identified as negatives or positives, respectively, by the *in vitro* test method, while the denominators represent the total number of substances identified as negatives or positives, respectively, by the *in vivo* rabbit eye test method.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies by performing a CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. In these studies, the mean and median CV values were (a) 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories, (b) 25% and 22%, respectively, for results obtained in five laboratories, and (c) 32.4% and 22.8%, respectively, for results obtained in three laboratories.

When studies are conducted using the BCOP test method, the study protocol should be based on the recommended standardized test method protocol provided in **Appendix D**. Exceptions and/or changes to the standardized test method protocol should be accompanied by a scientific rationale.

Users should be aware that BCOP's performance characteristics and the standardized test method protocol could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other chemical and product classes. Therefore, prior to initiation of BCOP studies, investigators are encouraged to consult the ICCVAM/National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To further characterize and potentially improve the usefulness of the BCOP test method for identifying ocular corrosives and severe irritants, and to evaluate its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations are recommended:

1. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be conducted. Such data will allow for the development of standardized decision criteria and a more comprehensive evaluation of the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.
2. Studies should be conducted to evaluate the impact of using a corneal holder that maintains normal corneal curvature (e.g., the corneal mounting system designed by Ubels et al. 2002) on accuracy and/or reliability of the BCOP test method.
3. The effect of modifying various test method protocol components (e.g., changing the duration of exposure) on the accuracy and/or reliability of the BCOP test method should be evaluated.

ICE Test Method

There are sufficient data to support the use of the ICE test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations for this method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols, surfactants and solids range from 33% (1/3) to 50% (1/2), 44% (4/9) to 57% (4/7), and 46% (6/13) to 70% (7/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols range from 27% (3/11) to 50% (5/10), depending on the hazard classification system evaluated. When substances within these chemical and physical classes are excluded from the database, the accuracy of ICE across the EU, EPA, and GHS classification systems ranges from 91% (72/79 to 75/82) to 92% (69/75) and the false negative and false positive rates range from 29% (2/7) to 33% (3/9) and 5% (4/73) to 6% (4/68 to 4/70), respectively.

The range of CV values for the corneal thickness measurement, when results were compared within experiments, was from 0.9% to 6.1%. The other endpoints evaluated produced ranges of CV values that were larger, with variability most prominent with the nonirritating substance.

The range of CV values for the corneal thickness measurement, when results were compared across experiments, was from 1.8% to 6.3%. The CV values for the remaining endpoints had a larger range (e.g., corneal swelling CV = 13.9% to 138.7%). However, if the nonirritating substance was removed, the range of CV values was reduced (e.g., corneal swelling CV = 13.9% to 22.4%).

One interlaboratory comparative study involving four laboratories contained test data on 59 substances for an assessment of interlaboratory reproducibility. Based on a qualitative analysis, 60% to 70% of the substances classified as ocular corrosives or severe irritants, depending on the regulatory classification system employed (i.e., EPA 1996, EU 2001, GHS [UN 2003]), were correctly identified by all four participating laboratories. A CV analysis of these same data indicated that the mean and median CV for severe substances tested was less than 35% for all test method endpoints, with the exception of corneal swelling.

When studies are conducted using this test method, the study protocol should be based on the recommended standardized ICE test method protocol provided in **Appendix E**. Exceptions and/or changes to the standardized test method protocol should be accompanied by a scientific rationale.

Users should be aware that ICE's performance characteristics and the standardized test method protocol could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of ICE studies,

investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To further characterize and potentially improve the usefulness of the ICE test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations are recommended:

1. Additional optimization studies/evaluations should be conducted in an attempt to decrease the 29% to 33% false negative rate of the ICE test method. After optimization, additional studies to further assess the reliability and accuracy of the test method are recommended.
2. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be included when the ICE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed on the optical pachymeter, which is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

IRE Test Method

Based on the accuracy (64% [68/107] to 69% [79/114]), false negative (24% [12/49] to 31% [14/45]), and false positive (35% [23/65] to 40% [25/62]) rates across the EU, EPA, and GHS classification systems, the use of the IRE test method for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended. There also are insufficient data using all four recommended IRE endpoints (corneal opacity, fluorescein penetration, corneal swelling, and observations of significant effect on corneal epithelium) to assess test method accuracy and reliability when all these endpoints are evaluated in a single study.

Based on a qualitative analysis of interlaboratory reproducibility in one study, 100% of the 12 to 18 tested substances were correctly identified as ocular corrosives or severe irritants by the IRE test method by all participating laboratories, depending on the regulatory classification system employed (i.e., EPA 1996, EU 2001, GHS [UN 2003]). Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds; and such product classes as organic solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for two studies by performing a CV analysis. The CV analysis of the first study indicated that the median CV for all substances tested was 43.4% for the 4-hour corneal opacity endpoint and 49.7% for the 4-hour swelling endpoint. When only ocular corrosives or severe irritants were considered, the CV values were 33.6% for the 4-hour corneal opacity endpoint and 35.5% for the 4-hour corneal swelling endpoint. In the second study, the median CV values for the endpoints evaluated (corneal opacity, corneal swelling, and fluorescein penetration) ranged from 24.0% to 40.0% when all substances were considered and from 15.4% to 35.5% when only ocular corrosives or severe irritants were considered.

When non-regulatory, validation, or optimization studies are conducted using the IRE test method, the protocol should be based on the standardized protocol provided in **Appendix F**. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that IRE's performance characteristics and the standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To potentially improve the usefulness of the IRE test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations should be conducted:

1. The IRE test method decision criteria should be optimized. Once optimized, additional validation studies should be conducted to further evaluate the relevance and reliability of the IRE test method.
2. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, of the corneal tissue should be included when the IRE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed when an optical pachymeter is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

HET-CAM Test Method

ICCVAM evaluated several HET-CAM analysis methods proposed for identifying substances that are ocular corrosives or severe irritants. These included one analysis method termed Irritation Score (IS)(B)-10 and another analysis method termed IS(B)-100. The range of hazard classification accuracy rates across the EU, EPA, and GHS classification systems for these two analysis methods ranged from 65% (64/98) to 68% (69/101) for IS(B)-10 and 52% (69/133) to 57% (94/164) for IS(B)-100, when the decision criteria of Luepke (1985) were used. The overall false negative and false positive rates of the IS(B)-10 analysis method range from 30% (10/33 to 12/40) to 32% (10/31) and 33% (20/61) to 36% (24/67), respectively, depending on the classification system. The overall false negative and false positive rates for the IS(B)-100 analysis method range from 6% (2/33) to 13% (5/39) and 52% (68/131) to 59% (58/99), respectively, depending on the classification system. Based on these rates, the use of these analyses methods and decision criteria for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended.

The analysis of intralaboratory repeatability was evaluated using data from two different studies for the IS(B) analysis method. In both studies, the hemorrhage endpoint had the highest CV value (109.10%-117.56%). Similar results were obtained for an analysis of intralaboratory reproducibility for the same two studies.

A qualitative analysis of interlaboratory reliability also was conducted for the IS(B) analysis method. For the IS(B)-10 analysis method, the participating laboratories were in 100% agreement for 84 to 85 (79% to 81%) of 104 to 107 substances evaluated, when compared to all three hazard classification systems. For the IS(B)-100 analysis method, the participating laboratories in a study were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated, when compared to all three hazard classification systems.

A quantitative evaluation of interlaboratory reproducibility for 14 substances, evaluated at 100% concentration (IS(B)-100), indicated that the mean and median CV values were 31.86% and 33.04%, respectively. For 12 substances evaluated at 10% concentration (IS(B)-10) in the same study, the mean and median CV values were 66.29% and 60.75%, respectively. For the substances evaluated in another study, which used the IS(B) analysis method, the mean and median CV values for substances tested at 10% concentration were 60.17% and 42.65%, respectively. For substances tested at 100% concentration in the same study, the mean and median CV values were lower: 35.21% and 26.22%, respectively.

When non-regulatory, validation, or optimization studies are conducted using the HET-CAM test method, the protocol should be based on the standardized protocol provided in **Appendix G**. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that HET-CAM's performance characteristics and the standardized test method protocol could be revised as additional data becomes available. Therefore, prior

to initiation of HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To potentially improve the usefulness of the HET-CAM test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), additional studies should be conducted to further optimize the HET-CAM prediction models and the decision criteria (e.g., mtc10) that would be used to identify ocular corrosives and severe irritants for the EPA, GHS, or EU classification systems.

Comparison of Performance Characteristics and General Recommendations for Four *In Vitro* Test Methods

Results from appropriately validated *in vitro* ocular toxicity test methods are recommended for use in a weight-of-evidence decision making process in accordance with the EPA and EU ocular testing regulations (EPA 1996, EU 2004) and the GHS tiered-testing strategy (UN 2003). In these testing schemes, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are indicated based on a weight-of-evidence evaluation of supplemental information (e.g., structure-activity relationships, other testing data). Use of a weight-of-evidence decision making process and a tiered-testing strategy for classification of substances as ocular corrosives or severe irritants will eliminate the pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances.

The comparative accuracy and false positive/false negative rates of these four *in vitro* ocular toxicity test methods in identifying ocular corrosives and severe irritants using the EU, EPA, and GHS classification systems are summarized in **Table 6-1**. Exclusion of specific chemical and physical classes increases the accuracy and decreases the false positive and false negative rates for BCOP and ICE. ICCVAM recommends that users consider, to the extent possible, the chemical and physical structures of the substances to be tested to determine whether either of these test methods would be appropriate to use as a screening test for ocular corrosion or severe irritation. Additional studies with each test method are recommended to determine if modification of the test method standardized protocol and/or the decision criteria for classification of a test substance as a corrosive/severe irritant or as a nonsevere irritant/nonirritant can improve test method sensitivity and specificity.

Additional research and development, optimization, and/or validation efforts should use reference substances with existing rabbit data. Additional rabbit studies should be conducted only if important data gaps are identified. If such studies are conducted, they should be designed to minimize the number of rabbits tested, to minimize or avoid pain and distress, and to maximize the information collected. Design and conduct of such studies should be in accordance with the recommendations from the Scientific Symposium on Mechanisms of Chemically-Induced Ocular Injury and the Scientific Symposium on Minimizing Pain and Distress in Ocular Safety Testing (see http://iccvam.niehs.nih.gov/methods/ocudocs/ocumeet/sympinfo.htm). These symposia were organized by ICCVAM, NICEATM, and the European Centre for the Validation of Alternative Methods.

All raw data generated using any of the recommended standardized *in vitro* ocular testing protocols and the *in vivo* rabbit eye test on the same substance should be submitted to NICEATM to expand the available validation database for these four test methods. The availability of such data will allow for additional retrospective evaluations of test method accuracy and/or reliability. Ideally, all substances should be completely identified (e.g., chemical name, chemical class, physicochemical properties). However, if this is not possible for proprietary reasons, data may be submitted using coded labels for each substance tested. If such coding is used, as much information as possible on physical and chemical properties should be provided to NICEATM.

Although the IRE and HET-CAM test methods cannot currently be recommended for meeting regulatory testing requirements, there may be non-regulatory uses for these two test methods. Accordingly, the four *in vitro* test methods should be considered prior to conducting *in vivo* ocular testing and an alternative test method should be used where determined appropriate for the specific testing situation. Since ocular irritancy testing frequently involves more than slight or momentary pain or distress, consideration of alternative test methods prior to the use of animals is necessary to comply with provisions of U.S. Animal Welfare Act regulations (9 CFR, Part 2, Section 2.31 and 9 CFR, Part 2, Section 2.32), the Public Health Service Policy on the Humane Care and Use of Laboratory Animals (PHS 2002), and the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (National Research Council 1996).

The potential usefulness of combining two or more *in vitro* test methods in a battery to identify ocular corrosives and severe irritants should be evaluated. Currently, there is insufficient guidance on the utility of a battery approach for such determinations.

Interested stakeholders are encouraged to support research and development of alternative test methods and technologies that may provide for a more accurate assessment of ocular toxicity and/or advantages in terms of time and cost.

ICCVAM Recommended Substances for Validation of *In Vitro* Ocular Toxicity Test Methods for the Evaluation of Ocular Corrosives and Severe Irritants

ICCVAM developed a list of reference substances recommended for the development of alternative ocular toxicity test methods and for evaluating the performance of any optimized test method protocol (**Appendix H**). Use of this standardized list of reference substances will aid in evaluating the comparative performance of different alternative test methods and, thus, in the selection of the most appropriate test method(s) to be used for a particular testing purpose. In accordance with ICCVAM procedures, once an adequate validation database is available for any of these test methods, performance standards will be developed that can be used to evaluate the performance of other test methods that are structurally and functionally similar. These performance standards will include essential test method components, a minimum list of reference chemicals (i.e., a subset of the recommended list in this report), and comparable performance that should be achieved.