3.0 THE ICE TEST METHOD

3.1 ICE Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the ICE BRD, which reviewed the available data and information for the test method.[11] The BRD describes the current validation status of the ICE test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/).

3.1.1 Test Method Description

The ICE test method is an organotypic model that provides short-term maintenance of the chicken eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, opacity, and fluorescein retention. While the latter two parameters involve a subjective assessment, analysis of corneal swelling provides an objective measurement. This objective measure potentially provides improved precision and reduced interlaboratory variability compared to the traditional *in vivo* rabbit eye test, which relies only on subjective measurements. Each measurement is either converted into a quantitative score used to calculate an overall Irritation Index, or assigned a qualitative categorization that is used to assign an *in vitro* irritancy classification. Either of these outcomes can then be used to predict the *in vivo* ocular irritation potential of a test substance. A histopathological assessment also can be included on a case-by-case basis to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category).

The ICE test method protocols used in the various studies are similar, but not identical.[12] The primary difference among these protocols was the number of treated eyes per test substance. Acceptable ranges for negative control responses, historical data used to establish these ranges, and procedures to determine the optimum quantity of test substance to be applied have not been published.

3.1.2 Validation Database

A total of 154 substances in five studies were used to evaluate ICE test method accuracy. These substances represent a variety of chemical and product classes (ICCVAM 2006b). The chemical classes tested included, but were not limited to, acyl halides, alcohols, alkalis, amines/amidines, carboxylic acids, esters, heterocyclic, hydrocarbons, inorganic salts, ketones, onium compounds, and organophosphates. Commercial products or formulations tested included, but were not limited to, detergents, pesticides, silicone powder, ink, solvents, surfactants, toilet cleaners, and thermal paper coatings.

---

[11]Comparison of the performance analysis for ICE to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

[12]For additional information on this evaluation, please see the ICE BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice).

3.1.3 Test Method Accuracy

Based on all available data, the ICE test method has an overall accuracy of 83% (120/144) to 87% (134/154), an overall false positive rate of 6% (7/122) to 8% (9/114 to 9/116), and an overall false negative rate of 41% (13/32) to 50% (15/30), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), EU (2001), or GHS (UN 2003) classification systems.

There were some notable trends in the performance of the ICE test method among substances grouped according to chemical class and/or physicochemical properties (**Table 3-1**). The chemical class of substances that was most consistently overpredicted (i.e., were false positives) by the ICE test method, according to the GHS classification system, is alcohols (50%, 5/10). With regard to physical form, liquids (10%, 9/90) appear more likely than solids (0%, 0/24) to be overpredicted by the ICE test method.

No single chemical class was prominently represented among 15 substances that were underpredicted. Five of the 15 underpredicted substances were unclassified coded substances and three were carboxylic acids. No other chemical class was represented more than twice. However, these studies do suggest that surfactants or formulations containing surfactants (e.g., detergents) (56%, 5/9) may be underpredicted by the ICE test method. They also suggest that pesticides (60%, 3/5) may be underpredicted.

With regard to physical form, eight of the 15 underpredicted substances were liquids while seven were solids. However, considering that the total number of solids (36) in the database is much smaller than the number of liquids (108), solids, with a false negative rate of 58% (7/12), appear more likely to be underpredicted than liquids, with a false negative rate of 44% (8/18).

ICE test method performance statistics also were evaluated when substances from the classes that gave the most discordant results were excluded (i.e., alcohols, surfactants, solids). When using the GHS classification system, exclusion of surfactants and solids individually resulted in small changes in the performance statistics. However, exclusion of alcohols from the data set caused a two-fold decrease in the false positive rate from 8% (9/114) to 4% (4/104). When both alcohols and surfactants were excluded, the false positive rate decreased from 8% (9/114) to 4% (4/92). The largest changes were observed when all three discordant classes were excluded from the data set; accuracy increased from 83% (120/144) to 92% (69/75), the false negative rate decreased from 50% (15/30) to 29% (2/7), and the false positive rate decreased from 8% (9/114) to 6% (4/68).

Among the eight underpredicted substances for which pH information was available, four were acidic (pH <7.0) and four were basic (pH >7.0). Basic substances (8) occupy a smaller proportion of the total database than acidic substances (12), and were more often underpredicted (50% vs. 33%). However, pH information was obtained for only 20 of the 30 total Category 1 substances.

Finally, the underpredicted substances were more likely to be classified *in vivo* based on persistent lesions (according to the GHS classification system) than on severe lesions.

Table 3-1 False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall	144	8	9/114	50	15/30
Chemical Class[5]					
Alcohol	12	50	5/10	50	1/2
Amine/Amidine	5	0	0/2	33	1/3
Carboxylic acid	10	0	0/3	43	3/7
Ester	9	13	1/8	0	0/1
Heterocyclic	9	0	0/3	33	2/6
Onium compound	8	0	0/2	33	2/6
Properties of Interest					
Liquids	108	10	9/90	44	8/18
Solids	36	0	0/24	58	7/12
Pesticide	11	0	0/6	60	3/5
Surfactant – Total	21	0	0/12	56	5/9
-nonionic	4	0	0/3	100	1/1
-anionic	2	0	0/1	100	1/1
-cationic	7	0	0/1	33	2/6
pH – Total[6]	20	-	-	40	8/20
- acidic (pH < 7.0)	12	-	-	33	4/12
- basic (pH > 7.0)	8	-	-	50	4/8
Category 1 Subgroup[7]					
- Total	23[9]	-	-	35	8/23
- 4 (CO=4 at any time)	12	-	-	33	4/12
- 3 (severity/persistence)	2	-	-	50	1/2
- 2 (severity)	4	-	-	0	0/4
- 2-4 combined[8]	18	-	-	28	5/18
- 1 (persistence)	5	-	-	60	3/5

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); ICE = Isolated Chicken Eye.

[1]N = number of substances.

[2]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

[4]Data used to calculate the percentage.

[5]Chemical classes included in this table are represented by at least five substances tested in the ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

[6]Total number of GHS Category 1 substances for which pH information was obtained.

[7]NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO=4 at any time.

[8]Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

[9]The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

However, four substances that caused severe lesions *in vivo* (corneal opacity=4) were false negatives in ICE.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the ICE test method for the EPA and EU classification systems can be obtained from **Section 6.0**, **Appendix B**, and the ICE BRD.

3.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Data were received that allowed for a quantitative analysis of intralaboratory repeatability and reproducibility of ICE test method endpoints. 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. However, this could be an exaggeration of variability given the relatively small values that were produced from the nonirritating substance relative to the irritating and corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher CV than values of 11, 14, and 18). A similar discussion also can be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively). 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 is removed, the range of CV values is reduced (e.g., corneal swelling CV = 13.9% to 22.4%).

One interlaboratory comparative study involving four laboratories contained sufficient ICE 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.

3.2 ICCVAM Recommendations for the ICE Test Method

3.2.1 Use of the ICE Test Method

ICCVAM recognizes that the ICE test method is not proposed as a stand alone replacement for the *in vivo* rabbit eye test method currently used for regulatory classification and labeling. ICCVAM concludes that 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.

A tiered-testing strategy for ocular irritation/corrosion (e.g., as described in the Globally Harmonized System of Classification and Labelling of Chemicals; UN 2003) allows for the use of validated and accepted *in vitro* methods prior to the use of animals for ocular safety testing. In a tiered-testing strategy, 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 suggested based on a weight-of-evidence evaluation of supplemental information (e.g., pH, structure-activity relationships, other testing data). Using *in vitro* data in a tiered-testing strategy with a weight-of-evidence decision process to classify substances as ocular corrosives or severe irritants will avoid the potential pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances. A tiered-testing strategy may not be applicable to purposes other than regulatory classification and labeling.

Users should be aware that ICE's performance characteristics 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, and chemical and physical class performance characteristics. 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.

3.2.2 ICE Test Method Protocol

ICCVAM recommends that when testing is conducted, the ICE test method protocol should be based on the ICE standardized test method protocol provided in **Appendix E**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the proposed standardized test method protocol should be accompanied by a scientific rationale. Users should be aware that the test method protocol could be revised based on future optimization and/or validation studies. ICCVAM, therefore, recommends that test method users consult the ICCVAM/NICEATM website to ensure use of the most current recommended test method protocol (http://iccvam.niehs.nih.gov/methods/eyeirrit.htm).

3.2.3 Optimization of the Current ICE Test Method Protocol

The current ICCVAM recommendations are focused on the use of the ICE test method as a screening test for ocular corrosives and severe irritants (see **Section 3.2.1**). For that use, the current test method protocol should be sufficient. To further the use of this test method and to evaluate the use of the ICE test method as a potential replacement for the *in vivo* rabbit eye test method or for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), ICCVAM recommends additional studies be considered and undertaken.

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.

ICCVAM recommends that a histopathological evaluation of the corneal tissue, using a standardized scoring scheme, 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.