4.0 THE IRE TEST METHOD

4.1 IRE Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the IRE BRD, which reviewed the available data and information for the test method.[13] The BRD describes the current validation status of the IRE 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/).

4.1.1 Test Method Description

The IRE test is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the entire rabbit eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, corneal opacity, fluorescein retention, and effects on the corneal epithelium. Identification of severe ocular irritants and corrosives is based on reaching or exceeding predetermined cut-off values in any one of the four endpoints (e.g., product of the corneal opacity and area scores ≥3; product of area and intensity scores for fluorescein penetration ≥4; corneal swelling ≥25%; or any significant effect on corneal epithelium (pitting, mottling, stippling, ulceration) (See **Appendix F** for details).

The IRE test method protocols used in the various studies are similar, but not identical.[14] Examples of some of the test method components that differed among the IRE protocols used to generate data include:

- temperature of solution used to rinse solids from the eyes ranged from room temperature to 32 °C,
- amount of substance applied as a solid ranged from 25 mg to 100 mg, and
- decision criteria used for classification of substances was based on scores from two to four endpoints.

4.1.2 Validation Database

A total of 149 substances were evaluated in three studies, of which 25 were commercial products or formulations (ICCVAM 2006c). The chemical classes tested included, but were not limited to, alcohols, amides, amines, carboxylic acids, esters, ethers, formulations, heterocyclic, ketones, onium compounds, and sulfur compounds. The commercial products or formulations tested were skin cleansers, soaps, shampoos, conditioners, surfactants, and solvents.

---

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

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

4.1.3 Test Method Accuracy

The overall accuracy (based on the pooled data set[15]) for the IRE test method ranged from 64% (68/107) to 69% (79/114) when compared to the *in vivo* test method data classified according to the GHS (UN 2003), EPA (1996), and EU (2001) regulatory classification systems. The overall false positive rates, when compared to these regulatory classification systems, ranged from 35% (23/65) to 40% (25/62). The overall false negative rates, when compared to the three regulatory classification systems, ranged from 24% (12/49) to 31% (14/45).

There were some trends in the performance of the IRE test method among substances grouped according to chemical class and/or physicochemical properties (**Table 4-1**). The chemical classes that were consistently overpredicted (i.e., false positives), when compared to classifications based on the GHS classification system, were alcohols (55%, 6/11), amines (50%, 3/6), and ketones (67%, 4/6). The chemical classes that were underpredicted (i.e., false negatives), when compared to classifications based on the GHS classification system, were carboxylic acids (67%, 4/6) and organic compounds (50%, 3/6).

With regard to physical form, liquids have a higher false positive rate (49%, 18/37) when compared to solids (22%, 5/23) for the IRE test method. The false negative rates for liquids and solids were relatively similar (29%, 8/28 vs. 32%, 6/19; respectively).

A subset of the substances evaluated had pH information available. For these substances, the overall false positive rate was 24% (4/17) and the overall false negative rate was 0% (0/10).

Of the surfactant-based formulations evaluated by this test method, the false positive rate was 25% (2/8) and the false negative rate was 38% (6/16). Comparatively, for substances identified as surfactants in the database, the false positive rate was 40% (2/5) and the false negative rate was 12% (1/8).

Finally, the underpredicted substances were more likely to be classified *in vivo* (according to the GHS classification) system based on persistent lesions, rather than severe lesions. However, three substances that caused severe lesion *in vivo* (corneal opacity=4) were false negatives.

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 IRE test method for the EPA and EU classification systems can be obtained from **Section 6.0**, **Appendix B**, and the IRE BRD.

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

Due to the lack of available quantitative IRE test method data for replicate eyes within individual experiments or for replicate experiments within an individual laboratory, an

[15]The pooled dataset represents the results from all the available studies combined, regardless of the number of endpoints evaluated by each of the individual studies. Additional information about this dataset can be obtained from the IRE BRD.

Table 4-1 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System (Analysis Based on the Pooled Data Set)

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall	107	38	23/60	30	14/47
Chemical Class[5]					
Alcohol	13	55	6/11	50	1/2
Amide	5	0	0/3	0	0/2
Amine	11	50	3/6	20	1/5
Carboxylic acid	12	33	2/6	67	4/6
Ester	10	30	3/10	-	0/0
Ether	9	33	2/6	0	0/3
Formulation	24	25	2/8	38	6/16
Heterocycle	18	44	4/9	11	1/9
Ketone	6	67	4/6	-	0/0
Onium compound	10	33	1/3	0	0/7
Organic	12	17	1/6	50	3/6
Sulfur compound	8	20	1/5	33	1/3
Properties of Interest					
Liquid/Solution	65	49	18/37	29	8/28
Solids	42	22	5/23	32	6/19
Surfactant-based formulation	24	25	2/8	38	6/16
Surfactants	13	40	2/5	12	1/8
-nonionic	4	33	1/3	0	0/1
-anionic	2	0	0/1	100	1/1
-cationic	7	100	1/1	0	0/6
pH – Total[6]	27	24	4/17	0	0/10
-acidic	18	20	2/10	0	0/8
-basic	7	33	2/6	0	0/1
-equals 7	2	0	0/1	0	0/1
Category 1 Subgroup[7] **- Total**	37[9]	-	-	32	12/37
- 4 (CO=4 at any time)	11	-	-	27	3/11
- 3 (severity/persistence)	4	-	-	25	1/4
- 2 (severity)	3	-	-	33	1/3
- 2-4 combined[8]	18	-	-	28	5/18
- 1 (persistence)	19	-	-	37	7/19

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); IRE = Isolated Rabbit Eye.
[1]N = number of substances.
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[3]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *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 IRE 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 number of *in vivo* Category 1 substances evaluated, since some substances could not be classified into the subgroups used in the evaluation.

evaluation of the intralaboratory repeatability and reproducibility of the IRE test method could not be conducted. However, two studies contained sufficient IRE test data (n=59 and 21 substances, respectively) for a qualitative and quantitative assessment of interlaboratory reproducibility based on data reported for three or four different laboratories.

For the qualitative analysis of interlaboratory reproducibility, 100% of the 12 to 18 tested substances were correctly identified as ocular corrosives or severe irritants by the IRE test method by all four 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.

The quantitative evaluation of interlaboratory reproducibility was conducted for these two studies by performing a CV analysis. For the first study (n=59 substances), corneal opacity and corneal swelling were evaluated. For the second study (n=21 substances), corneal opacity, corneal swelling, and fluorescein penetration were evaluated. The CV analysis of the first study indicated that the median CV for all 59 substances tested was 43.4% for the 4-hour corneal opacity endpoint and 49.7% for the 4-hour swelling endpoint. The CV values were 33.6% for the 4-hour corneal opacity endpoint and 35.5% for the 4-hour corneal swelling endpoint when only ocular corrosives or severe irritants were considered. 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% (the largest variability was for corneal swelling) when all substances were considered. When only ocular corrosives or severe irritants were considered, the CV values ranged from 15.4% to 35.5%.

4.2 ICCVAM Recommendations for the IRE Test Method

4.2.1 Use of the 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.

Users should be aware that IRE's performance characteristics could be revised as additional data become available. Therefore, prior to initiation of non-regulatory, validation, or optimization 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, 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.

4.2.2 IRE Test Method Protocol

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**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that IRE's 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 recommended standardized test method protocol.

ICCVAM recommends that, for all studies, raw data be collected and maintained. The availability of such data will allow for further retrospective evaluation of test method accuracy and/or reliability.

4.2.3 Optimization of the Current IRE Test Method Protocol

ICCVAM recommends that additional evaluation studies be conducted to increase the current IRE database and optimize the IRE test method decision criteria. Once these studies are conducted, ICCVAM recommends that additional validation studies be conducted to further evaluate the relevance and reliability of the IRE test method.

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