Performance of the Isolated Chicken Eye (ICE) Test Method in Detecting Ocular Corrosives and Severe Irritants

D Allen[1,2], N Choksi[1,2], C Inhof[1,2], J Truax[1,2], R Tice[1,2], W Stokes[2].

[1]ILS, Inc., RTP, NC; [2]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, NC.

Introduction

The ocular irritation or corrosion potential of substances, to which humans may be exposed, has been evaluated since 1944 by the Draize rabbit eye test (Draize et al. 1944). There have been widespread efforts to develop and validate *in vitro* alternatives that might reduce or replace the use of animals for ocular irritancy assessments. The U.S. EPA[1] formally nominated to ICCVAM[1] four *in vitro* test methods, including the ICE[1] assay, for evaluation of their ability to identify ocular corrosives and severe irritants in a tiered testing strategy.

NICEATM[1], in conjunction with the ICCVAM Ocular Toxicity Working Group, prepared a comprehensive BRD[1] reviewing the available data and information on the ICE test method[2]. NICEATM released the draft ICE BRD for public comment on November 1, 2004. On January 11-12, 2005, ICCVAM convened an Expert Panel to independently evaluate the validation status of ICE and three other *in vitro* test methods for identifying ocular corrosives or severe irritants[3]. Public comments at that meeting indicated that additional data was available. The Expert Panel recommended that any additional data that could be obtained should be considered for a reanalysis of the accuracy and reliability of each test method.

In response to the Expert Panel recommendation, an *FR*[1] notice requesting the submission of all available *in vitro* ICE test data and corresponding *in vivo* rabbit eye test data was reissued on February 28, 2005 (FR Vol. 70, No. 38, pp. 9661-9662). In addition to considering any ICE data received in response to the *FR* notice, the reanalysis of the accuracy and reliability of this test method took into account (1) changes that occurred in the ocular irritancy classification of a few substances in response to clarification of the EU[1] (2001) and UN GHS[1] (UN 2003) ocular irritation classification rules, (2) a decision to use classifications based on *in vivo* rabbit eye test data only, and (3) revised chemical class assignments for some substances. Additional information on the reanalysis can be obtained at http://iccvam.niehs.nih.gov/methods/ocudocs/reanalysis.htm.

[1] BRD = Background Review Document; FR = Federal Register; GHS = Globally Harmonized System; ICCVAM = Interagency Coordinating Committee on the Validation of Alternative Methods; ICE = Isolated Chicken Eye; NICEATM = National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods; UN = United Nations; U.S. EPA = U.S. Environmental Protection Agency.
[2] The draft ICE BRD can be obtained at http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm
[3] The Expert Panel Report can be obtained at http://iccvam.niehs.nih.gov/methods/ocudocs/EPreport/ocureport.htm

Test Method Overview

During an ICE study, a test substance is applied to the cornea of eyes isolated from chickens processed for human consumption. Test substances are applied as a single dose (30 µL or 30 mg) for 10 sec followed by rinsing with isotonic saline. A single negative control eye (treated with saline) is used to verify assay conditions. Corneal reactions (swelling and opacity) are measured at 0, 30, 75, 120, 180, and 240 min post-treatment, and mean values (at each time point for all eyes) for each endpoint are determined. Fluorescein retention is evaluated at 0 and 30 min. The maximum mean value for each endpoint is used to categorize the response and then the categories for all the endpoints are used to assign an *in vitro* irritancy classification (**Table 1**). Morphological (e.g., loosening of the epithelium; roughening of the corneal surface) and histopathological assessments can also be included on a case-by-case basis to discriminate borderline cases, although decision criteria to assign an irritancy classification have not been established for histopathological endpoints.

Table 1. ICE Decision Criteria for Classifying Ocular Corrosives and Severe Irritants

Corneal Swelling		Corneal Opacity	
Max. Mean Swelling[1] (%)	Category	Max. Mean Score[1]	Category
0 - 5	I	0 - 0.5	I
>5 – 12	II	0.6 - 1.5	II
>12 – 18 (>75 min post-treatment)	II	1.6 - 2.5	III
>12 – 18 (≤ 75 min post-treatment)	III	2.6 – 4.0	IV
>18 – 26	III	Fluorescein Retention	
>26 – 32 (>75 min post-treatment)	III	Mean Score[2]	Category
>26 - 32 (≤ 75 min post-treatment)	IV	0 – 0.5	I
>32	IV	0.6 - 1.5	II
		1.6 - 2.5	III
		2.6 - 3.0	IV

[1]ICE endpoint measurements for three eyes are averaged at each time point. The greatest mean value at any time point (maximum mean value) is used for categorization.
[2]Recorded at 30 min post-treatment.

Possible combinations of the three ICE endpoint categories yielding a severe irritant/corrosive classification:

- **3 x IV**
- **2 x IV, 1 x III or II or 1**
- **CO ≥ 3 at 30 min**
- **CO = 4 at any time**
- **Severe loosening of the epithelium**

Test Method Database

A total of 175 substances from five different studies (Prinsen and Koëter 1993; Balls et al. 1995; Prinsen 1996, 2000, 2005) were used to evaluate the accuracy of the ICE test method; data for 59 substances were appropriate for evaluation of interlaboratory reproducibility, while data for four substances were appropriate for analysis of intralaboratory reproducibility. The primary difference among various ICE studies was the number of treated eyes per test substance (3 to 5).

ICE Test Method Accuracy Analysis

The overall ICE test method accuracy with regard to each of the three ocular irritation classification systems (EPA 1996, EU 2001, UN 2003) ranged from 83% to 87%, while the false positive and false negative rates ranged from 6% to 8% and 41% to 50%, respectively (**Table 2**). The small number of substances representing most chemical classes allows for only limited conclusions with respect to the accuracy of ICE by chemical class or property of interest (e.g., solids vs. liquids, basic vs. acidic pH, surfactants) (**Table 3**). However, among classes with at least six substances for analysis, alcohols tend to be overpredicted, while surfactants and solids tend to be underpredicted. Substances that are classified *in vivo* as ocular corrosives/severe irritants are most likely to be underpredicted when this classification is based solely on persistent lesions (i.e., those lasting 21 days) (data not shown).

Table 2. ICE Test Method Accuracy[1]

Statistic	Ocular Hazard Classification System		
	GHS (n=144)	EPA (n=145)	EU (n=154)[2]
Accuracy	83% (120/144)	84% (122/145)	87% (134/154)
Sensitivity	50% (15/30)	52% (15/29)	59% (19/32)
Specificity	92% (105/114)	92% (107/116)	94% (115/122)
False Positive Rate	8% (9/114)	8% (9/116)	6% (7/122)
False Negative Rate	50% (15/30)	48% (14/29)	41% (13/32)

Abbreviations: EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = Globally Harmonized System; n = number of substances in the database. The numbers in parenthesis indicate the data on which the % value is based.
[1]Data from Prinsen and Koëter (1993), Balls et al. (1995), Prinsen (1996), Prinsen (2000), and Prinsen (2005).
[2]Additional chemicals available for EU analysis only (individual animal data not available for GHS or EPA classification).

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

Category	N[2]	False Positive Rate[3]	False Negative Rate[3]
Overall	144	8% (9/114)	50% (15/30)
Chemical Class			
Alcohols	12	50% (5/10)	50% (1/2)
Carboxylic acids	10	0% (0/3)	43% (3/7)
Esters	9	13% (1/8)	0% (0/1)
Heterocycles	9	0% (0/3)	33% (2/6)
Onium compounds	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)
Pesticides	11	0% (0/6)	60% (3/5)
Surfactants	21	0% (0/12)	56% (5/9)

[1]GHS = Globally Harmonized System (UN [2003]).
[2]N = number of substances.
[3]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*; False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*. The data used to calculate the percentage are provided in parenthesis.
Highlighted cells indicate those chemical classes and properties of interest where the rate of misclassification is (a) greater than the overall rate, (b) is based on a sufficient number of substances (N ≥ 6 substances), and (c) would be expected to have an appreciable effect on the overall rate, if excluded from the database when conducting an accuracy analysis.

ICE Test Method Reliability Analysis

***Intralaboratory Repeatability*[4]**
Coefficient of variation (CV) analysis performed on within-experiment ICE test method data indicated that the corneal thickness measurement was generally repeatable (data not shown). The other endpoints evaluated produced somewhat more variable responses, most prominent with the nonirritating substance (SP-1). 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. A similar discussion 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).

***Interlaboratory Reproducibility*[4]**
To evaluate intralaboratory reproducibility, CV values were calculated for within laboratory values for each test method endpoint along with the ICE Irritation Index for each test substance (**Table 4**). Similar to the repeatability assessment, the corneal thickness measurement was generally reproducible, but the %CV values for the remaining endpoints had a much larger range. However, if the nonirritating substance is removed, the range of %CV values is reduced.

***Interlaboratory Reproducibility*[4]**
Table 5 summarizes the results of a qualitative analysis for the three regulatory classification systems. The four participating laboratories were in 100% agreement in regard to the ocular irritancy classification of ~75% of the substances tested, and were in at least 75% agreement for 90% of the substances tested. When only severe irritants (based on *in vivo* rabbit eye test results) were considered, the participating laboratories were in 100% agreement for ~70% of the substances tested, and were in at least 75% agreement for at least 95% of the substances tested.

Table 6 summarizes a quantitative analysis using %CV values. A wide range of %CV values was evident for all endpoints evaluated in the ICE test method. When all available data were considered, median CV values for both corneal opacity and fluorescein retention were ~35%, while the median CV for corneal swelling was 75%. When only severe irritants (based on *in vivo* rabbit eye test results) were considered, median CV values for corneal opacity, fluorescein retention, and corneal swelling were reduced to 25%, 23%, and 70%, respectively.

[4]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; in this case, this refers to the variability among replicate corneas. Intralaboratory Reproducibility=A determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times. Interlaboratory Reproducibility=A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results.

Table 4. Intralaboratory Reproducibility of ICE Test Method Endpoints – Prinsen (2000)

Substance (Experimental Replicates)	EU Class[1]	CT (mean[2])	CT (%CV)	CS (mean)	CS (%CV)	CO (mean)	CO (%CV)	FR (mean)	FR (%CV)	Index (mean)	Index (%CV)
SP-1 (5)	NI	62.1	2.2	1.1	138.7	0.2	95.8	0.2	141.4	9.3	91.8
SP-4 (5)	R36	70.7	4.0	15.1	22.4	2.5	18.1	2	0	104.4	10.3
SU-4 (5)	R36	70.5	6.3	12.3	15.2	0.7	10.6	1	0	46.3	4.1
SU-5 (4)	R41	76.1	1.8	20.2	13.9	1.9	8.7	2	0	98.6	6.1

Abbreviations: Class = Classification; CO = Corneal opacity; CS = Corneal swelling; CT = Corneal thickness; CV = coefficient of variation; EU = European Union; FR = fluorescein retention; Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20])
[1]*In vivo* animal data were not provided. EU classification was provided by the *in vitro* testing laboratory.
[2]Mean of three eyes.

Table 5. Qualitative Analysis of ICE Test Method Interlaboratory Reproducibility: Agreement Among Four Laboratories[1]

% Interlaboratory Agreement	Classification System		
	GHS (59 substances)	EPA (59 substances)	EU (59 substances)
100% (all substances)	75% (44/59)	75% (44/59)	76% (45/59)
≥75% (all substances)	90% (53/59)	90% (53/59)	90% (53/59)
100% (severe *in vivo* and *in vitro* substances)[2]	72% (16/22)	74% (14/19)	68% (13/19)
≥75% (severe *in vivo* and *in vitro* substances)[2]	95% (21/22)	100% (19/19)	95% (18/19)

Abbreviations: EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals
The numbers in parenthesis indicate the data on which the % value is based.
[1]Balls et al. (1995).
[2]Scores for fluorescein retention and corneal swelling were not provided for one severe irritant/corrosive (30% trichloroacetic acid). Classification based on results from only 3 laboratories.

Table 6. Quantitative Analysis of ICE Test Method Interlaboratory Reproducibility: % CV Values

		%CV Values[1]		
		FR	CO	CS
Total (59 Substances)	Mean	38.8%	46.8%	77.2%
	Median	35.6%	37.1%	74.5%
	Range	0% - 158.7%	0% - 158.7%	30.8% - 159.4%
GHS Category 1 (22 Substances)	Mean	29.9%	34.2%	72.4%
	Median	23.0%	25.0%	69.5%
	Range	0% - 158.7%	0% - 118.6%	32.2% - 132.2%

Abbreviations: CO = Corneal opacity; CS = Corneal swelling; %CV = CV (standard deviation/mean), expressed as a percentage; FR = Fluorescein retention.
[1]Balls et al. (1995). Interlaboratory %CV values based on results from four laboratories.

References

Balls M, et al. 1995. Toxicol In Vitro 9(6):871-929.

Burton ABG, et al. 1981. Food Cosmet Toxicol 19:417-480.

Draize J, et al. 1944. J Pharmacol Exp Ther 82:377-390.

EPA. 1996. EPA737-B-96-001. Washington, DC: U.S. E.P.A.

EU. 2001. Official Journal of the European Communities L255:1-333.

ICCVAM. 1997. NIH Publication No: 97-3981. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate/pdf.

ICCVAM. 2003. NIH Publication No: 03-4508. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm.

Prinsen MK. 2005. Unpublished data (n = 94 substances) provided directly to NICEATM by MK Prinsen, TNO Nutrition and Food Research Institute.

Prinsen MK. 2000. Unpublished report (TNO Report V99.521b) provided directly to NICEATM by M Prinsen, TNO Nutrition and Food Research Institute.

Prinsen MK. 1996. Food Chem Toxicol 34:291- 296.

Prinsen MK, Koëter BWM. 1993. Food Chem Toxicol 31:69-76.

UN. 2003. New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html.

U.S. Department of Health and Human Services
National Institutes of Health
National Institute of Environmental Health Sciences

Acknowledgments

We would like to especially thank Menk Prinsen (TNO Nutrition and Food Research, The Netherlands) for his many contributions to the ICE BRD. NICEATM and ICCVAM also gratefully acknowledge the Procter and Gamble Company, which also provided data for this review of the ICE test method.

This poster was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. ILS staff supported by NIEHS contract N01-ES 35504. The views expressed above do not necessarily represent the official positions of any federal agency.