NIEHS
National Institute of Environmental Health Sciences

Validation Status of the Isolated Chicken Eye (ICE) Test Method

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), NIEHS/NIH/DHHS, RTP, NC

Abstract

Concerns about animal welfare and interest in higher throughput testing have led researchers to develop alternative *in vitro* test methods for the current rabbit eye test. NICEATM evaluated four *in vitro* ocular test methods for their ability to identify substances that cause irreversible or severe irritation or corrosion. One of these test methods, ICE, is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the eye in an isolated system. The ability of ICE to correctly identify ocular corrosives and severe irritants using available ICE and corresponding *in vivo* eye irritation data was evaluated according to current hazard classification schemes for the U.S. EPA (n=90), the European Union (n=121), and the UN Globally Harmonized System (n=92). Depending on the classification scheme used, ICE had a false positive rate of 8-10% and a false negative rate of 30-40%. In terms of reliability, the assay has acceptable interlaboratory reproducibility; intralaboratory reproducibility could not be assessed. A proposed standardized test method protocol and a proposed recommended list of reference substances have been developed for future validation and testing studies to further assess the accuracy, reliability, and the applicability domain of ICE for the detection of ocular corrosives/severe irritants. Investigators should consider using ICE prior to eye irritation testing in animals. When used in a tiered testing strategy, positive results could be used to classify and label a substance, while substances with negative results would undergo additional eye irritation testing to identify false negative ocular corrosives/severe irritants and to identify those chemicals with reversible ocular effects. This approach would reduce the number of animals used for eye irritation testing and reduce the number of animals experiencing pain and distress. ILS staff supported by NIEHS contract N01-ES 35504.

Introduction

The Draize rabbit eye test (Draize et al. 1944), in one form or another, has been used since 1944 to evaluate the ocular irritation or corrosion potential of substances to which humans might be exposed. In recent years, increased efforts have focused on the development of *in vitro* alternatives to this *in vivo* test. Although progress has been made, there is currently no validated nonanimal alternative test method for ocular irritancy. Recently, the U.S. Environmental Protection Agency (EPA) formally nominated to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) four *in vitro* test methods for evaluation of their ability to identify potential ocular corrosives and severe irritants in a tiered testing strategy. Adequate validation[1] of a test method is a prerequisite for it to be considered for regulatory acceptance (ICCVAM 1997, 2003). One of the four test methods nominated for review was the Isolated Chicken Eye (ICE) test. The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) worked with the ICCVAM Ocular Toxicity Working Group (OTWG) and with the European Centre for the Validation of Alternative Methods (ECVAM) to compile information and data to support the ICCVAM technical evaluation of the test method.

NICEATM, which administers ICCVAM and provides scientific support for ICCVAM activities, prepared a comprehensive background review document (BRD) reviewing the available data and information for the ICE test. The objectives of this BRD were to:

- describe the current validation status of the ICE test method, including what is known about its accuracy and reliability, the scope of the substances tested, and the availability of a standardized protocol for its use in a regulatory tiered testing strategy (e.g., the UN Globally Harmonised System; GHS [UN 2003])[2].
- identify the usefulness and limitations of the ICE test method, based on existing data, for identifying ocular corrosives and severe irritants in a regulatory tiered testing strategy
- propose a standardized ICE test method protocol
- propose additional test method protocol optimization studies that might enhance the accuracy and/or reliability of the ICE test method and validation studies to further characterize its usefulness and limitations
- propose reference substances for future optimization/validation studies of this and other alternative test methods intended to identify ocular corrosives and severe irritants

This BRD was based on published studies using the ICE test method, and other data and information submitted in response to a 2004 *Federal Register* (*FR*) request (*FR* Vol. 69, No. 57, pp. 13859-13861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). The ICE test method was reviewed by an independent Expert Panel on January 11-12, 2005 and their report, including conclusions and recommendations, will be available in mid-March at this website. After considering the Expert Panel's report and any public comments, ICCVAM will prepare final recommendations on the ICE test method, which will be published in a report that is expected to be available in August 2005.

[1]Validation is the process by which the reliability and relevance of a test method are established for a specific purpose (ICCVAM 1997, 2003).

[2]In a regulatory tiered testing strategy, substances positive in a validated *in vitro* test method could be classified and labeled as ocular corrosives or severe irritants, while negative substances would undergo additional testing in the *in vivo* rabbit eye test or a validated alternative test method capable of detecting ocular corrosives and severe irritants that were false negatives in the validated *in vitro* test method.

Test Method Overview

The ICE test method is an *in vitro* organotypic model that provides short-term maintenance of the chicken eye in an isolated system. The protocol used for the ICE test method is based that of the Isolated Rabbit Eye test (Burton et al. 1981) and has remained virtually unchanged since its initial publication (**Table 1**). The primary difference among various ICE studies was the number of treated eyes per test substance (3 to 5). In the ICE test method, damage by a test substance is assessed by qualitative measurements of corneal opacity and fluorescein retention, and quantitative measurements of corneal swelling. 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 seconds followed by rinsing with isotonic saline. Historically, a single negative control eye (treated with saline) has been used to verify assay conditions; positive controls have not been included. The corneal reactions are measured at baseline and at 30, 75, 120, 180, and 240 minutes post-treatment, and mean values (at each time point for all eyes) for each endpoint are determined (fluorescein retention is evaluated only at 30 minutes). Each measurement can be converted into a qualitative categorization using the maximum mean values for each endpoint that is used to assign an *in vitro* irritancy classification (**Table 2**). 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. Protocol Variations Among ICE Test Method Studies

Study	n	# Eyes: Neg	# Eyes: Treat	# Eyes: Pos	Exposure Duration	CO	CS	FR	Morph. Eval.	Histopath.
Prinsen and Koëter (1993)	21	1	5	-	10 sec	X	X	X	X	Case-by-case basis
Balls et al. (1995)	59	1	3	-	10 sec	X	X	X	X	Not specified
Prinsen (1996)	44	1	3	-	10 sec	X	X	X	X	Not specified

CO: Corneal opacity; CS: Corneal swelling; FR: Fluorescein retention; Histopath.: Histopathology; Morph. Eval.: Morphological evaluation; n: Number of substances tested; Neg: negative control; Pos: positive control; Treat: test substance treated

Table 2. ICE Decision Criteria for Classifying Ocular Corrosives and Severe Irritants

Corneal Swelling		Corneal Opacity	
Max. Mean Swelling* (%)	Category	Max. Mean Score*	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**	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

*ICE endpoint measurements are averaged at each time point across the three test eyes. The mean value for each endpoint that is the greatest at any time point (maximum mean value) is used for categorization.
**Recorded at 30 minutes 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 I**
- **CO ≥ 3 at 30 min**
- **CO = 4 at any time**
- **Severe loosening of the epithelium**

Test Method Database

A total of 121 substances in three studies (Prinsen and Koëter 1993; Balls et al. 1995; Prinsen 1996) were used to evaluate the accuracy of the ICE test method; data for 59 of these substances could also be used to evaluate interlaboratory reproducibility (**Table 3**). Twenty-two different chemical classes were represented among these substances (alcohols, acids and surfactants were the most commonly represented), while 20 different product classes were represented (chemical/ pharmaceutical intermediates, herbicides/pesticides, industrial chemicals, and soaps/surfactants/detergents were the most commonly represented). Much of the published *in vivo* rabbit eye test data on the substances used to evaluate the accuracy of ICE for detecting ocular corrosives and severe irritants was limited to average score data or the reported irritancy classification. However, detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal at 24, 48, and 72 hours and/or assessment of the presence or absence of lesions at 7, 14, and 21 days was necessary to calculate the appropriate EPA (1996), European Union (EU 2001), and GHS (UN 2003) ocular irritancy hazard classification. Thus, many of the test substances for which there was only limited *in vivo* data could not be used for evaluating test method accuracy and reliability for all three ocular irritancy classification systems.

Table 3. Primary ICE Data Sources

Study		Accuracy: GHS	Accuracy: EPA	Accuracy: EU	Accuracy: Total	Intralab: CVs	Intralab: GHS classific.	Interlab: CVs	Interlab: GHS classific.
Prinsen and Koëter (1993)	S	3	8	8	8	-	-	-	-
	NS	7	13	13	13	-	-	-	-
	Total	10	21	21	21	-	-	-	-
Balls et al. (1995)	S	22	21	21	21	-	-	59	22
	NS	34	38	38	38	-	-	59	34
	Total	56	59	59	59	-	-	59	56
Prinsen (1996)	S	0	6	6	6	-	-	-	-
	NS	29	38	38	38	-	-	-	-
	Total	29	44	44	44	-	-	-	-

classific. = classification; CV = coefficient of variation; EPA = Environmental Protection Agency; EU = European Union; GHS = Globally Harmonized System; NS = nonsevere irritants or nonirritants; S = severe irritant or corrosive

More information on ICCVAM and NICEATM can be accessed at: http://iccvam.niehs.nih.gov/.

Test Method Accuracy Analysis

The ability of the ICE test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS, EPA, and EU classification systems (EPA 1996; EU 2001; UN 2003)[3], was evaluated. Accuracy statistics were calculated for each ICE test method protocol by report and, where appropriate:

- classifications were pooled into one classification per substance (i.e., majority call among studies used)
- using individual studies, where a balanced design existed (multiple substances in multiple labs as in Balls et al. 1995)

The overall ICE test method accuracy with regard to each of the three classification systems ranged from 82% to 85%, while the false positive and false negative rates ranged from 8% to 10% and 30% to 40%, respectively (**Table 4**). **Table 5** lists the false negative and false positive rates of the ICE test method for identifying ocular corrosives/severe irritants based on chemical class and physicochemical properties. This table highlights the fact that, among this limited dataset, alcohols appear to be often overpredicted (50% [5/10] false positive rate) and surfactants appear to be often underpredicted (57% [4/7] false negative rate). In regard to physical properties solids were commonly underpredicted (55% [6/11] false negative rate).

There were notable limitations of the ICE test method accuracy analysis. A lack of individual rabbit test data prevented an accuracy evaluation for all 121 substances using the GHS and EPA classification schemes. In addition, the small number of substances representing each chemical class allows for only limited conclusions with respect to the accuracy of ICE by chemical class or physicochemical property. However, it appears that alcohols tend to be overpredicted, while surfactants and solids tend to be underpredicted in the ICE test method.

[3]For the purposes of this analysis, an ocular corrosive or severe irritant was defined as a substance that would be classified as Category 1 according to the GHS classification system (UN 2003), as Category I according to the EPA classification system (EPA 1996), or as R41 according to the EU classification system (EU 2001).

Table 4. ICE Test Method Accuracy

Statistic	GHS (n=92)*		EPA (n=90)*		EU (n=121)*	
	%	n	%	n	%	n
Accuracy	82	75/92	82	74/90	85	103/121
Sensitivity	60	15/25	61	14/23	70	26/37
Specificity	90	60/67	90	60/67	92	77/84
False Positive Rate	10	7/67	10	7/67	8	7/84
False Negative Rate	40	10/25	39	9/23	30	11/37

***from Prinsen and Koëter (1993), Balls et al. (1995), and Prinsen (1996); Additional chemicals available for EU analysis only (individual animal data not available for GHS or EU classification)**

Table 5. ICE Test Method Accuracy According to Chemical Class and Physical Property

Class	# of Substances: Total	Cat I	Cat 2A, 2B, NI	False Positive Rate: %	n	False Negative Rate: %	n
OVERALL	92	25	67	10	7/67	40	10/25
Surfactant	13	7	6	0	0/6	57	4/7
Alcohol	10	0	10	50	5/10	-	-
Acid	7	5	2	0	0/2	20	1/5
Acetate	6	0	6	17	1/6	-	-
Heterocyclic	6	5	1	0	0/1	40	2/5
Hydrocarbon	6	2	4	0	0/4	50	1/2
Inorganic	3	1	2	0	0/2	100	1/1
Ketone	3	0	3	33	1/3	-	-
Amine	2	1	1	0	0/1	0	0/1
Acyl halide; Lactone; Aldehyde; Amide; Organometallic; Organophosphate	1	0	1	0	0/1	-	-
Alkali	1	1	0	-	-	0	0/1
Diol	1	1	0	-	-	100	1/1
Solids	23	11	12	0	0/12	55	6/11
Liquids	69	14	55	13	7/55	29	4/14

***Prinsen and Koëter (1993), Balls et al. (1995), and Prinsen (1996)**
NI = nonirritant

Test Method Reliability Analysis

Due to the lack of available ICE test data, analyses of intralaboratory repeatability and reproducibility were not conducted. However, an assessment of interlaboratory reproducibility was feasible using data from a single study (Balls et al. 1995) in which 59 substances were tested among four different laboratories. Two separate analyses were conducted:

- A qualitative analysis, in which the extent of agreement between testing laboratories when identifying ocular corrosives and severe irritants was compared
- A quantitative analysis using a coefficient of variation calculation to compare variability in measurements of the three primary ICE test method endpoints

Table 6 summarizes the results of the qualitative analysis for each of the three regulatory classification systems. The four participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 to 45 (75% to 76%) of the 59 substances tested, depending on the classification system used. The four participating laboratories were in at least 75% agreement for 53 (90%) of the 59 substances tested, regardless of the classification system used. When only severe irritants (based on *in vivo* classification) were considered, the four participating laboratories were in 100% agreement for 15 to 16 (71% to 75%) of the 20 to 22 substances tested, depending on the classification system used. The four participating laboratories were in at least 75% agreement for 20 to 21 (95% to 100%) of the 20 to 22 substances tested.

Table 7 summarizes the quantitative analysis using a coefficient of variation (CV) calculation. Mean and median %CV values were calculated to provide an assessment of overall variability. A wide range of %CV values was evident for all endpoints evaluated in the ICE test method. When data from all 59 test substances was considered, median CV values for both corneal opacity and fluorescein retention were approximately 35%, while the median CV for corneal swelling was 75%. When only severe irritants (based on *in vivo* classification) were considered, median CV values for corneal opacity, fluorescein retention, and corneal swelling were reduced to 25%, 23%, and 70%, respectively.

Table 6. Qualitative Analysis of ICE Test Method Interlaboratory Reproducibility: Extent of Agreement Between Four Laboratories

% Interlaboratory Agreement	GHS (59 substances)*		EPA (59 substances)*		EU (59 substances)*	
	%	n	%	n	%	n
100% (all)	75	44/59	75	44/59	76	45/59
≥75% (all)	90	53/59	90	53/59	90	53/59
100% (severes)*+[1]	72	16/22	75	15/20	71	15/21
≥75% (severes)*+[1]	95	21/22	100	20/20	95	20/21

***Balls et al. (1995)**
+Scores for fluorescein retention and corneal swelling were not provided for one severe irritant/corrosive (30% trichloroacetic acid), which was therefore classified based on results from only 3 laboratories.
[1]Severe irritant/corrosive based on *in vivo* classification

Table 7. Quantitative Analysis of ICE Test Method Interlaboratory Reproducibility: Coefficient of Variation (CV)

		%CV: FR	%CV: CO	%CV: 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

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

Draft BRD Proposals

Based on this evaluation of ICE test method performance, a proposed ICE test method protocol, which evaluates corneal opacity, corneal swelling, fluorescein retention, and morphological effects, is proposed. This protocol is based on the method of TNO Nutrition and Food Institute (INVITTOX 2004). The only difference between the two protocols is that the proposed protocol includes additional eyes for negative controls, a concurrent positive control, and, when appropriate, benchmark controls.

Additional optimization studies are suggested that may enhance the performance of the ICE test method for identifying ocular corrosives and severe irritants. These studies include:

- A retrospective analysis of decision criteria used to classify substances as corrosives and severe irritants
- An evaluation of the potential causes of the interlaboratory variability for the corneal swelling endpoint, and procedural modifications that may reduce this variation.
- Additional evaluation of possible increased interlaboratory variability for specific chemical classes appearing more variable, based on the small numbers of representative substances in this evaluation (i.e., alcohols, acetates/esters, cationic surfactants)
- Determining the feasibility of introducing quantitative measurement for corneal opacity
- Determining the utility of histopathology and when it should be included.

Once these studies have been completed, additional validation studies will be necessary using substances from the proposed list of reference substances to further characterize the accuracy and reliability of the optimized method

Proposed Reference Substances For Optimization and Validation Studies

A common list of reference substances proposed for future optimization and validation studies of ICE and other alternative test methods intended to detect ocular corrosives/severe irritants has been developed[4]. Following completion of the proposed validation studies, reference substances from this list can be selected for inclusion in performance standards and for proficiency testing. Substances included in this list are intended to:

- represent the range of responses (i.e., corrosive/severe irritant; nonsevere irritant/noncorrosive) that the test method is expected to be capable of measuring or predicting
- represent the range of chemical/product classes and physicochemical properties that the test method might be expected to be capable of testing
- represent the range of known or anticipated mechanisms or modes of action for severe/irreversible ocular irritation or corrosion
- have been generated by high-quality *in vivo* studies following Organisation for Economic Co-operation and Development Test Guideline 405 and preferably conducted in compliance with Good laboratory Practice guidelines
- have well-defined chemical composition, with defined purity, be readily available, and not associated with excessive hazard or prohibitive disposal costs

This list of substances is intended to represent the minimum number of substances that should be used to evaluate the accuracy and reliability of an *in vitro* ocular test method proposed for the detection of ocular corrosives and severe irritants.

[4]See ICE BRD: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice

References

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The ECHO international validation study on alternatives to the Draize eye irritation test. Toxicol In Vitro 9:871-929.

Burton ABG, York M, Lawrence RS. 1981. The *in vitro* assessment of severe irritants. Food Cosmet Toxicol 19:417-480.

Draize J, Woodard G, Calvery H. 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J Pharmacol Exper Therapeut 82:377-390.

EPA. 1996. Label Review Manual: 2nd Ed. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.

EU. 2001. Commision Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L255:1-333.

ICCVAM. 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. Research Triangle Park, NC: National Toxicology Program.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park, NC: National Toxicology Program.

INVITTOX. 2004. Protocol No. 80: Chicken enucleated eye test (CEET). Available: http://ecvam-sis.jrc.it/invittox/static/index.html.

NICEATM. 2004. Draft Background Review Document: The Isolated Chicken Eye (ICE) Test Method. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice.

Prinsen MK. 1996. The chicken enucleated eye test (CEET): A practical (pre)screen for the assessment of eye irritation/corrosion potential of test materials. Food Chem Toxicol 34:291- 296.

Prinsen MK, Koëter BWM. 1993. Justification of the enucleated eye test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. Food Chem Toxicol 31:69-76.

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/officialtext.html.

Acknowledgments

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

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