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

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

[1]Integrated Laboratory Systems, Inc., RTP, NC; [2]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences, RTP, 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 IRE[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 IRE test method[2]. NICEATM released the draft IRE 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 IRE 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* IRE test data and corresponding *in vivo* rabbit eye test data was reissued on February 28, 2005 (*FR* Vol. 70, No. 38, pp. 9661-9662). Although additional IRE test data was not received in response to the *FR* notice, a reanalysis of the accuracy and reliability of this test method was conducted that took into account (1) data already available but not previously considered, (2) 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, (3) a decision to use classifications based on *in vivo* rabbit eye test data only, and (4) 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; IRE = Isolated Rabbit 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 IRE 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

The IRE test method was initially developed by Burton et al. (1981) as an alternative to the Draize rabbit eye test (Draize et al. 1944) to reduce the pain and suffering of rabbits exposed to ocular corrosives or severe irritants. Enucleated rabbit eyes of good quality are physiologically maintained and exposed to liquid or solid test substances. The corneas of these eyes are examined visually and by slit-lamp at various times before and after exposure. In a modification of the original protocol (Guerriero et al. 2004), four endpoints are used (i.e., corneal opacity, corneal swelling, fluorescein penetration, and epithelial integrity). A substance is identified as an ocular corrosive/severe irritant if any single endpoint value meets or exceeds a defined cutoff value (decision criteria).

Test Method Database

The database for the original IRE accuracy analysis (IRE BRD) consisted of a total of 149 test substances obtained from four studies (CEC 1991; Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). However, only Guerriero et al. (2004) used all four ocular endpoints to identify corrosives or severe irritants. Although no additional data were received in response to the resissued *FR* notice, the Expert Panel recommended that an "expanded" data base be considered, which in addition to the substances tested by Guerriero et al. (2004) included substances from CEC (1991), Balls et al. (1995), and Gettings et al. (1996) that were identified as corrosive or severe irritants using the recommended IRE protocol decision criteria. Nonsevere irritants from theses additional studies could not be included in the reanalysis, since only one to three ocular endpoints were used in these studies and the use of any missing endpoint might have resulted in the substance being classified as a corrosive/severe irritant.

References

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

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

CEC. 1991. Doc. XI/632/91/V/E/1/131/91 Part I and II.

Draize J, et al. 1944. J Pharmacol Exper Therapeut 82:377-390.

EPA. 1996. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001. Official Journal of the European Communities L255:1-333.

Gettings S, et al. 1996. Food Chem Toxic 34:79-117.

Guerriero FJ, et al. 2004. [Abstract No. 1282]. Toxicol Sci (The Toxicologist Supplement) 78(1-S).

ICCVAM. 1997. NIH Publication No.: 97-3981. Research Triangle Park, NC: National Toxicology Program.

ICCVAM. 2003. NIH Publication No: 03-4508. Research Triangle Park, NC: National Toxicology Program.

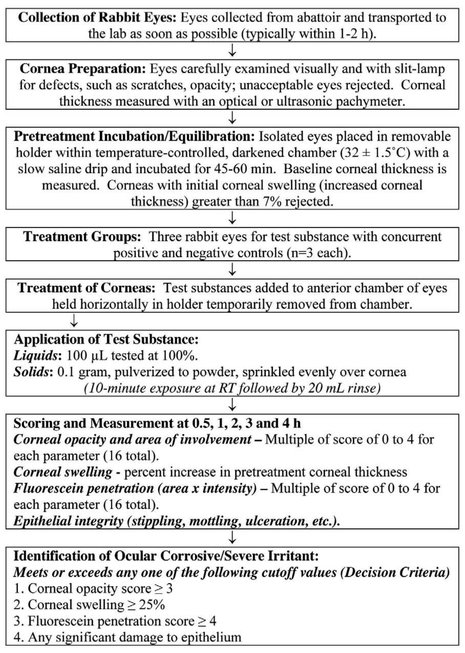
McDonald TO, et al. 1977. Eye irritation. In: Advances in Modern Toxicology: Dermatotoxicology. 1st ed. (Marzulli FN, Maibach HI. eds.). Washington: Hemisphere Publishing Corporation. 135-189.

UN. 2003. New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/officialtext.html [accessed 15 July 2005].

Test Method Protocol

A standardized protocol for the IRE test method based on approach used by Guerriero et al. (2004) is provided in **Figure 1**. Essentially, enucleated rabbit eyes with healthy corneas evaluated before and after harvest are placed in a removable holder held in a temperature-controlled chamber (31 ± 1.5ºC) with a saline drip providing constant moisture. The eyes are exposed to 0.1 mL or 0.1 gram of a test substance for 10 seconds followed by a 20 mL rinse. Corneas are then evaluated for opacity and swelling (measured as a change in thickness), fluorescein penetration, and epithelial damage at 0.5, 1, 2, 3, and 4 hours. Substances that induce a response that exceeds a cutoff score in any one of four ocular endpoints (corneal opacity score [opacity x area] > 3, corneal swelling > 25%, fluorescein penetration score [intensity x area] > 4, or any sign of epithelial damage [stippling, mottling, ulceration, etc.]) are identified as corrosive or severe ocular irritants. Test substances that do not meet these criteria are identified as nonsevere irritants.

Figure 1. Protocol for the IRE Test Method



Test Method Accuracy Analysis

The accuracy of the IRE test method, based on the substances tested by Guerriero et al. (2004), compared to *in vivo* irritancy using the EPA (1996), GHS (UN 2003), and the EU (2001) regulatory classification systems are provided in **Table 1**. The overall IRE test method accuracy was identical between the three classification systems at 79% (30/38), with a false positive rate of 30% (8/27), and a false negative rate of 0% (0/11). When substances classified as corrosives/severe irritants by CEC (1991), Balls et al. (1995), and Gettings et al. (1996) based on IRE test data were combined with the substances tested by Guerriero et al. (2004), accuracy ranged from 66 to 70% with a false positive rate of 56 to 58%, and a false negative rate of 0% (n=76 to 80) for the three regulatory classification systems (**Table 1**). The accuracy analysis of the expanded data set was influenced by the exclusion of substances classified as nonsevere irritants/nonirritants correctly identified by the IRE test method in the CEC (1991), Balls et al. (1995), and Gettings et al. (1996) studies. The small number of substances representing most chemical classes allows for only limited conclusions with respect to the accuracy of IRE by chemical class or property of interest (e.g., solids vs. liquids, basic vs. acidic pH, surfactants) (**Table 2**). However, among classes with at least six substances for analysis, ketones and liquids tend to be overpredicted compared to the overall false positive rate of 56%.

Table 1. IRE Test Method Accuracy

Statistic	Study	Ocular Hazard Classification System		
		GHS (n=38/76)	EPA (n=38/76)	EU (n=38/80)
Accuracy	Guerriero et al. (2004)	79% (30/38)	79% (30/38)	79% (30/38)
	Expanded Data Set[1]	68% (52/76)	66% (50/76)	70% (56/80)
Sensitivity	Guerriero et al. (2004)	100% (11/11)	100% (11/11)	100% (11/11)
	Expanded Data Set	100% (33/33)	100% (31/31)	100% (37/37)
Specificity	Guerriero et al. (2004)	70% (19/27)	70% (19/27)	70% (19/27)
	Expanded Data Set	44% (19/43)	42% (19/45)	44% (19/45)
False Positive Rate	Guerriero et al. (2004)	30% (8/27)	30% (8/27)	30% (8/27)
	Expanded Data Set	56% (24/43)	58% (26/45)	56% (24/43)
False Negative Rate	Guerriero et al. (2004)	0% (0/11)	0% (0/11)	0% (0/11)
	Expanded Data Set	0% (0/33)	0% (0/31)	0% (0/37

Abbreviations: EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals; n = number of substances in the database. The numbers in parenthesis indicate the data on which the % value is based.
[1]Substances in CEC (1991), Balls et al. (1995), and Gettings et al. (1996) identified as corrosives/severe irritants using the Guerriero et al. (2004) decision criteria were added to the substances tested by Guerriero et al (2004). Substances identified as nonsevere irritants in these studies could not be used in the accuracy analysis, since only one to three of the four ocular endpoints were used and any missing endpoint might have resulted in a severe irritant classification.

Table 2. False Negative and False Positive Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS[1] Classification System (Analysis Based on the Expanded Data Set)

Category	N[2]	False Positive Rate[3]	False Negative Rate[3]
Overall	76	56% (24/43)	0% (0/33)
Chemical Class			
Alcohols	11	60% (6/10)	0% (0/1)
Amines	9	60% (3/5)	0% (0/4)
Esters	6	67% (4/6)	-
Ethers	8	40% (2/5)	0% (0/3)
Formulations	12	100% (2/2)	0% (0/10)
Heterocycles	16	50% (4/8)	0% (0/8)
Ketones	6	67% (4/6)	-
Onium compounds	9	33% (1/3)	0% (0/6)
Sulfur compounds	7	20% (1/5)	0% (0/2)
Properties of Interest			
Liquids/Solutions	43	83% (19/23)	0% (0/20)
Solids	33	25% (5/20)	0% (0/13)
Surfactants	10	50% (2/4)	0% (0/6)

[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.

Test Method Reliability Analysis

Intralaboratory Repeatability[4] and Reproducibility[4]

Data from replicate eyes within a study or replicate tests within a laboratory were not provided, precluding an analysis of intralaboratory repeatability and reproducibility.

Interlaboratory Reproducibility[4]

Two interlaboratory reproducibility analyses were conducted:

- ***Qualitative analysis***: This analysis evaluated the extent of agreement among testing laboratories for prediction of ocular irritancy outcomes (i.e., positive, negative, false positive, and false negative)
- ***Quantitative analysis***: This analysis used a percent coefficient of variation (%CV) calculation to compare variability in ocular irritancy classification

Qualitative Analysis

The four testing laboratories in the Balls et al. (1995) study, which examined corneal opacity and corneal swelling in their IRE test method evaluation, were in 100% agreement with respect to the *in vivo/in vitro* outcomes (severe/nonsevere) 59 to 63% (35-37/59) of the time, depending on the ocular hazard classification system used. When agreement between three of the four laboratories was considered, the agreement increased to 85% (50/59) across all regulatory classifications. For the CEC (1991) study, which examined corneal opacity, corneal swelling, and fluorescein penetration in their IRE test method evaluation, the three testing laboratories agreed 81% (17/21) of the time when the EU classification system was used (*in vivo* rabbit eye test data, which would have enabled an ocular hazard determination using the GHS or EPA classification systems, was not provided). When agreement between two of three laboratories was considered, the agreement increased to 95% (20/21).

Quantitative Analysis

The evaluation of interlaboratory reproducibility is shown in **Table 4**. In the Balls et al. (1995) study, mean %CV values of 63.8 and 53.5 for corneal opacity (CO) and corneal swelling (CS), respectively, were obtained for all 59 substances. When only GHS Category 1 substances are considered, the mean %CV values were reduced to 40.5 and 36.9 for CO and CS, respectively, but the ranges were only marginally impacted. In the CEC (1991) study, mean %CV values of 37.7, 57.3, and 58.9 were obtained for CO, CS, and fluorescein penetration (FP), respectively. When only GHS Category 1 substances are considered, the mean %CV values were reduced to 15.5, 35.4, and 22.1 for CO, CS, and FP, respectively, and the range of values was reduced for all endpoints.

[4] Intralaboratory Repeatability = 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. 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 3. IRE Interlaboratory Reproducibility –Overall Classification Agreement Among Laboratories

Agreement[1]	Balls et al. (1995) 4 Laboratories 2 IRE Endpoints[2]			CEC (1991)[3] 3 Laboratories 3 IRE Endpoints[2]
	GHS	EPA	EU	EU
100%	59% (35/59)	61% (36/59)	63% (37/59)	81% (17/21)
75%	85% (50/59)	85% (50/59)	85% (50/59)	-
67%	-	-	-	95% (20/21)
50%	100% (59/59)	100% (59/59)	100% (59/59)	-
33%	-	-	-	100% (21/21)

Abbreviations: EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals; n = number of substances in the database. The numbers in parenthesis indicate the data on which the % value is based.
[1]Percent of agreement with all outcomes combined (i.e., +/+, +/-, -/+, ?/-, ?/+) based on *in vivo* and *in vitro* irritancy classifications, respectively. A "+" indicates corrosive/severe irritant classification, "-" indicates nonsevere irritant, "?" indicates substance could not be classified *in vivo*.
[2]Corneal opacity and corneal swelling were measured in Balls et al. (1995); Corneal opacity, corneal swelling, and fluorescein penetration were measured in CEC (1991).
[3]The EU classification was provided in the CEC (1991) study; GHS and EPA classification could not be determined for this study, since *in vivo* rabbit eye test data was not available.

Table 4. Quantitative Analysis of IRE Test Method Interlaboratory Reproducibility: Coefficient of Variation (CV) Values

		%CV Values[3]		
		CO	CS	FP
Total (59 Substances)[1]	Mean	63.8%	53.5%	-
	Median	43.4%	49.7%	
	Range	0% - 200%	10% - 118%	-
GHS Category 1 (22 Substances)[1]	Mean	40.5%	36.9%	-
	Median	40.6%	36.0%	-
	Range	0% - 200%	1% - 118%	-
Total (21 Substances)[2]	Mean	37.7%	57.3%	58.9%
	Median	24.0%	40.0%	28.0%
	Range	0% - 141%	7.2% - 173%	0% - 175%
GHS Category 1 (8 Substances)[2]	Mean	15.5%	35.4%	22.1%
	Median	15.4%	35.5%	21.0%
	Range	0% - 40%	20% - 61%	0% - 78%

Abbreviations: CO = corneal opacity; CS = corneal swelling; CV = coefficient of variation (standard deviation/mean), expressed as a percentage; FP = fluorescein penetration.
[1]Balls et al. (1995): %CV values based on four testing laboratories.
[2]CEC (1991): %CV values based on three testing laboratories.

Acknowledgements

NICEATM and ICCVAM would like to thank the generous contributions of the companies and individuals who provided data for this review of the IRE test method:

- Cosmetics, Toiletry, and Fragrance Association (Dr. Carol Eisenmann)
- ECVAM (Dr. Chantra Eskes)
- GlaxoSmithKline (Mr. Frederick Guerriero)
- National Institute of Health Sciences (Japan) (Dr. Yasuo Ohno)
- Proctor and Gamble (Dr. Daniel Marsman)
- SafePharm Laboratories (Mr. Robert Guest, Dr. Andrew Wittingham)
- Unilever (Ms. Penny Jones)
- U.S. Food and Drug Administration (Ms. Donnie Lowther)

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.

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