HTML

NIH Publication No: 06-4514

BACKGROUND REVIEW DOCUMENT

Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants:

Isolated Rabbit Eye Test Method

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

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

THE INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS and THE NTP INTERAGENCY CENTER FOR THE EVALUATION OF ALTERNATIVE TOXICOLOGICAL METHODS

In 1997, the National Institute of Environmental Health Sciences (NIEHS), one of the National Institutes of Health (NIH), established the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to:

- Coordinate interagency technical reviews of new and revised toxicological test methods, including alternative test methods that reduce, refine, or replace the use of animals

- Coordinate cross-agency issues relating to validation, acceptance, and national and international harmonization of new, modified, and alternative toxicological test methods

On December 19, 2000, the ICCVAM Authorization Act (42 U.S.C. § 2851-2, 2851-5 [2000]) established ICCVAM as a permanent interagency committee of NIEHS under the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

ICCVAM is comprised of representatives from 15 U.S. Federal regulatory and research agencies that use, generate, or disseminate toxicological information. ICCVAM conducts technical evaluations of new, revised, and alternative methods with regulatory applicability. The Committee promotes the scientific validation and regulatory acceptance of toxicological test methods that more accurately assess the safety or hazards of chemicals and products and that refine (i.e., decrease or eliminate pain and distress), reduce, and/or replace animal use. NICEATM administers ICCVAM and provides scientific and operational support for ICCVAM-related activities. More information about ICCVAM and NICEATM can be found on the ICCVAM/NICEATM web site (http://iccvam.niehs.nih.gov) or obtained by contacting NICEATM (telephone: [919] 541-2384, e-mail: iccvam@niehs.nih.gov).

The following Federal regulatory and research agencies are ICCVAM members:

Consumer Product Safety Commission
Department of Agriculture
Department of Defense
Department of Energy
Department of Health and Human Services
- Agency for Toxic Substances and Disease Registry
- Food and Drug Administration
- National Cancer Institute
- National Institute of Environmental Health Sciences
- National Institutes of Health, Office of the Director
- National Institute of Occupational Safety and Health
- National Library of Medicine

Department of the Interior
Department of Labor
- Occupational Safety and Health Administration

Department of Transportation
Environmental Protection Agency

On the Cover
The ICCVAM/NICEATM graphic symbolizes the important role of new and alternative toxicological methods in protecting and advancing the health of people, animals, and our environment.

Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants:

Isolated Rabbit Eye Test Method

Background Review Document

Prepared by
The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences (NIEHS)
P.O. Box 12233
Mail Drop: EC-17
Research Triangle Park, NC 27709

March 2006
NIH Publication No. 06-4514

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

This document is available electronically at:
http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ire

Printed: 08/08/2006

TABLE OF CONTENTS

LIST OF TABLES ..... **viii**
LIST OF FIGURES ..... **xi**
LIST OF ACRONYMS AND ABBREVIATIONS ..... **xii**
ACKNOWLEDGEMENTS ..... **xix**
PREFACE ..... **xxi**
EXECUTIVE SUMMARY ..... **xxv**

1.0 Introduction and Rationale for the Proposed Use of *In Vitro* Test Methods to Identify Ocular Corrosives and Severe Irritants **1-1**
1.1 Introduction 1-1
1.1.1 Historical Background of *In Vitro* Ocular Irritation/Corrosion Test Methods and Rationale for Their Development 1-1
1.1.2 Peer Reviews of the IRE Test Method 1-6
1.2 Scientific Basis for the IRE Test Method 1-6
1.2.1 Purpose and Mechanistic Basis of the IRE Test Method 1-6
1.2.2 Similarities and Differences of Modes and Mechanisms of Action Between the IRE Test Method and Ocular Irritancy in Humans and/or Rabbits 1-7
1.2.2.1 The Mammalian Eye: Common Anatomy of the Human and Rabbit Eye 1-7
1.2.2.2 Differences Between Human and Rabbit Eyes 1-9
1.2.2.3 The *In Vivo* Rabbit Eye Test Method 1-10
1.2.2.4 Comparison of the IRE Test Method with the *In Vivo* Rabbit Eye Test Method 1-13
1.2.3 Intended Range of Substances Amenable to the IRE Test Method and/or Limits of the IRE Test Method 1-14
1.3 Regulatory Rationale and Applicability 1-14
1.3.1 Current Regulatory Testing Requirements and ICCVAM Prioritization Criteria 1-14
1.3.2 Intended Uses of the IRE Test Method 1-16
1.3.3 Similarities and Differences in the Endpoints Measured in the Proposed Test Method and *In Vivo* Reference Test Method 1-16
1.3.4 Use of Proposed Test Method in Overall Strategy of Hazard or Safety Assessment 1-16
1.4 Validation of the IRE Test Method 1-17
1.5 Search Strategies and Selection of Citations for the IRE BRD 1-17
1.6

2.0 IRE Test Method Protocol Components **2-1**
2.1 Overview of How the IRE Test Method is Conducted 2-1
2.2 Description and Rationale for the Test Method Components 2-1
2.2.1 Materials, Equipment, and Supplies Needed 2-2
2.2.1.1 Sources of Rabbit Eyes 2-2
2.2.1.2 Quality of Eyes 2-3
2.2.1.3 Preparation of the Eyes 2-3

2.2.1.4 IRE Experimental Setup 2-4
2.2.2 Dose-Selection Procedures, Including the Need for Any Dose Rangefinding Studies or Acute Toxicity Data Prior to Conducting a Study 2-4
2.2.3 Endpoints Measured 2-4
2.2.3.1 Corneal Opacity 2-5
2.2.3.2 Corneal Swelling 2-5
2.2.3.3 Fluorescein Penetration/Retention 2-5
2.2.3.4 Assessment of Epithelial Integrity 2-5
2.2.3.5 Additional Endpoints 2-6
2.2.4 Duration of Exposure 2-6
2.2.4.1 Pre-exposure Preparations 2-6
2.2.4.2 Test Substance Exposure Duration 2-6
2.2.4.3 Application and Amount of Test Substance 2-6
2.2.4.4 Number of Eyes Required per Test Substance and Controls 2-7
2.2.4.5 Concentration of Test Substance 2-7
2.2.5 Known Limits of Use 2-7
2.2.6 Nature of the Response Assessed 2-8
2.2.7 Appropriate Controls and the Basis for their Selection 2-8
2.2.7.1 Negative Controls 2-8
2.2.7.2 Solvent/Vehicle Controls 2-8
2.2.7.3 Positive Controls 2-8
2.2.7.4 Benchmark Controls 2-9
2.2.8 Acceptable Range of Control Responses and the Basis for the Acceptable Ranges 2-9
2.2.8.1 Negative/Solvent Controls 2-9
2.2.8.2 Positive Controls 2-9
2.2.8.3 Benchmark Controls 2-9
2.2.9 Nature of the Data to be Collected and the Methods Used for Data Collection 2-10
2.2.9.1 Corneal Opacity and Area of Involvement 2-10
2.2.9.2 Corneal Thickness and Calculation of Corneal Swelling 2-10
2.2.9.3 Fluorescein Penetration 2-11
2.2.9.4 Evaluation of Corneal Epithelial Integrity 2-12
2.2.9.5 Overall Scoring System for Identification of a Severe Irritant 2-12
2.2.10 Types of Media in which Data are Stored 2-13
2.2.11 Measures of Variability 2-13
2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data 2-14
2.2.13 Decision Criteria and the Basis for the Prediction Model Used to Classify a Test Chemical as a Severe Eye Irritant 2-14

2.2.14 Information and Data that Will Be Included in the Study Report and Availability of Standard Forms for Data Collection and Submission....2-14
2.3 Basis for Selection of the Test Method System....2-16
2.4 Proprietary Components....2-16
2.5 Basis for Number of Replicate and Repeat Experiments....2-16
2.6 Compliance with Good Laboratory Practice....2-18
2.7 Study Acceptance Criteria....2-18

3.0 Substances Used For Validation of the IRE Test Method....3-1
3.1 Rationale for the Substances or Products Selected for Use....3-1
3.1.1 CEC Collaborative Study (1991)....3-1
3.1.2 Balls et al. (1995)....3-2
3.1.3 Gettings et al. (1996)....3-2
3.1.4 Guerriero et al. (2004)....3-3
3.2 Rationale for the Number of Substances Tested....3-3
3.3 Chemicals or Products Evaluated....3-3
3.4 Coding Procedures Used in the Studies....3-5
3.4.1 CEC Collaborative Study (1991)....3-5
3.4.2 Balls et al. (1995)....3-5
3.4.3 Gettings et al. (1991, 1996)....3-5
3.4.4 Guerriero et al. (2004)....3-5

4.0 *In Vivo* Reference Data Used for an Assessment Of Test Method Accuracy....4-1
4.1 Description of Protocol Used to Generate *In Vivo* Data....4-1
4.1.1 Draize Rabbit Eye Test....4-1
4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols....4-2
4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems....4-3
4.2 Detailed Reference Data Used to Assess *In Vitro* Test Method Accuracy....4-9
4.3 *In Vivo* Classification Criteria Used for BRD Analysis....4-10
4.3.1 GHS Classification Rules Used for BRD Analysis....4-11
4.3.2 EPA Classification Rules Used for BRD Analysis....4-12
4.3.3 EU Classification Rules Used for BRD Analysis....4-13
4.4 Availability of Original Records for the *In Vivo* Reference Data....4-14
4.5 *In Vivo* Data Quality....4-14
4.6 Availability and Use of Toxicity Information from the Species of Interest....4-14
4.7 Information About Accuracy and Reliability of the *In Vivo* Test Method....4-16
4.7.1 Information About the Accuracy of the *In Vivo* Test Method....4-16
4.72 Information About the Reliability of the *In Vivo* Test Method....4-16

5.0 IRE Test Method Data And Results 5-1
5.1 Description of the IRE Test Method Protocols Used to Generate Data 5-1
5.1.1 CEC Collaborative Study (1991) 5-1
5.1.2 Balls et al. (1995)…. 5-1
5.1.3 Gettings et al. (1996)……..... 5-2
5.1.4 Guerriero et al. (2004)…… 5-2
5.2 Data Obtained to Evaluate the Accuracy and Reliability of the IRE Test Method..... 5-2
5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data 5-3
5.3.1 CEC Collaborative Study (1991) 5-3
5.3.2 Balls et al. (1995)…. 5-3
5.3.3 Gettings et al. (1996) 5-3
5.3.4 Guerriero et al. (2004)……..... 5-3
5.4 Summary of Results 5-4
5.5 Use of Coded Chemicals and Compliance with GLP Guidelines 5-4
5.6 Description of "Lot-to-lot" Consistency, Time Frame of Studies and Testing Laboratories 5-4
5.6.1 CEC Collaborative Study (1991) 5-5
5.6.2 Balls et al. (1995) 5-5
5.6.3 Gettings et al. (1996). 5-5
5.6.4 Guerriero et al. (2004)…. 5-5
5.7 Availability of Data for External Audit, if Requested 5-6

6.0 IRE Test Method Accuracy..... 6-1
6.1 Accuracy of the IRE Test Method 6-1
6.1.1 GHS Classification System: IRE Test Method Accuracy 6-3
6.1.1.1 Balls et al. (1995)..... 6-4
6.1.1.2 Gettings et al (1996)..... 6-4
6.1.1.3 Guerriero et al. (2004)..... 6-4
6.1.1.4 Expanded Data Set..... 6-6
6.1.1.5 Pooled Data Set..... 6-6
6.1.1.6 Discordant Results According to the GHS Classification System 6-6
6.1.1.7 Expanded Data Set..... 6-6
6.1.1.8 Pooled Data Set..... 6-6
6.1.2 EPA Classification System: IRE Test Method Accuracy..... 6-10
6.1.2.1 Balls et al. (1995)..... 6-10
6.1.2.2 Gettings et al. (1996)..... 6-10
6.1.2.3 Guerriero et al. (2004)..... 6-10
6.1.2.4 Expanded Data Set..... 6-12
6.1.2.5 Pooled Data Set..... 6-12
6.1.2.6 Discordant Results According to the EPA Classification System 6-12
6.1.3 EU Classification System: IRE Test Method Accuracy..... 6-15

6.1.3.1 CEC Collaborative Study (1991) 6-15
6.1.3.2 Balls et al. (1995) 6-17
6.1.3.3 Gettings et al (1996) 6-17
6.1.3.4 Guerriero et al. (2004) 6-17
6.1.3.5 Expanded Data Set 6-17
6.1.3.6 Pooled Data Set 6-17
6.1.3.7 Discordant Results According to the EU Classification System 6-17
6.2 Accuracy of the IRE Test Method for Identifying Ocular Corrosives and Severe Irritants - Summary of Results 6-21
6.2.1 Discordance Among Chemical Classes 6-21
6.2.2 Discordance Among Physical or Chemical Properties Of Interest 6-22

7.0 IRE Test Method Reliability 7-1
7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the IRE Test Method 7-1
7.2 Analyses of Repeatability and Reproducibility 7-2
7.2.1 Quantitative and Qualitative Assessments of Intralaboratory Repeatability 7-2
7.2.2 Quantitative and Qualitative Assessments of Intralaboratory Reproducibility 7-2
7.2.3 Assessment of Interlaboratory Reproducibility 7-3
7.2.3.1 Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System 7-3
7.2.3.2 Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System 7-5
7.2.3.3 Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System 7-6
7.2.3.4 Common Chemical or Product Classes Among Substances with Discordant Interlaboratory Results Using the GHS, EPA, and EU Classification Systems 7-9
7.2.4 Coefficient of Variation Analysis 7-10
7.2.4.1 CEC Collaborative Study (1991). 7-10
7.2.4.2 Balls et al. (1995). 7-12
7.2.5 Additional Analysis of Interlaboratory Reproducibility 7-17
7.3 Historical Positive and Negative Control Data 7-17
7.4 Conclusions 7-19

8.0 Test Method Data Quality 8-1
8.1 Adherence to National and International GLP Guidelines 8-1
8.2 Data Quality Audits 8-1

8.3 Impact of Deviations from GLP Guidelines 8-2
8.4 Availability of Laboratory Notebooks or Other Records..... 8-2
8.5 Need for Data Quality..... 8-2

9.0 Other Scientific Reports And Reviews 9-1
9.1 Summaries of IRE Data from Published and Unpublished Studies 9-1
9.1.1 Balls et al. (1995) 9-1
9.1.2 Chamberlain et al. (1997) 9-2
9.1.3 Cooper et al. (2001)..... 9-4
9.1.4 Gettings et al. (1996) 9-5
9.1.5 Guerriero et al. (2004) 9-6
9.1.6 Jacobs and Martens (1990) 9-6
9.1.7 Jacobs and Martens (1989) 9-7
9.1.8 Jacobs and Martens (1988) 9-7
9.1.9 Jones et al. (2001) 9-8
9.1.10 Koëter and Prinsen (1985)… 9-8
9.1.11 Lewis et al. (1994)..... 9-10
9.1.12 Price and Andrews. 9-10
9.1.13 Whittle et al. (1992)..... 9-11
9.1.14 York et al. (1982)… 9-12
9.2 Data Received in Response to the ICCVAM *Federal Register* Notice or from Study Authors 9-12
9.2.1 Jacobs and Martens (January 1987) 9-12
9.2.2 Jacobs and Martens (May 1987) 9-13
9.2.3 Proctor and Gamble (P&G) Submission from Drs. Daniel Marsman and Karen Acuff 9-13
9.2.3.1 Summary of P&G Confocal Ocular Test Method 9-13
9.2.3.2 P&G Data 9-14

10.0 Animal Welfare Considerations..... 10-1
10.1 How the IRE Test Method Will Refine, Reduce, or Replace Animal Use..... 10-1
10.2 Requirement for the Use of Animals 10-1

11.0 Practical Considerations 11-1
11.1 Transferability of the IRE Test Method..... 11-1
11.1.1 Facilities and Major Fixed Equipment 11-1
11.1.2 General Availability of Other Necessary Equipment and Supplies……..... 11-1
11.2 Training Considerations..... 11-2
11.2.1 Required Level of Training and Expertise Needed to Conduct the IRE Test Method..... 11-2
11.3 Cost Considerations 11-3
11.4 Time Considerations 11-3

12.0 References 12-1

13.0 Glossary 13-1

Appendix A Publicly Available Protocols for the IRE Test Method A-1
A1 INVITTOX Protocol 85. The Rabbit Enucleated Eye Test Method of Dr. Lesley Earl A-3
A2 Table of IRE Protocols from the Reviewed Literature A-13

Appendix B Chemical and Product Class Information for the Substances Tested in the IRE Test Method B-1
B1 Chemical and Product Class Information for the Substances Tested in the IRE Test Method. B-3
B2 Table of Formulation Components in Gettings et al. (1996)…..... B-13

Appendix C *In Vitro* Data for Substances Tested in the IRE Test Method C-1
C1 Balls et al. (1995) Data Sorted by Substance Name C-3
C2 Guerriero et al. (2004) Data Sorted by Substance Name C-13
C3 Gettings et al. (1996) - Data Sorted by Substance Name C-21
C4 CEC (1991) - Data Sorted by Reference C-27

Appendix D *In Vivo* and *In Vitro* Comparison of Ocular Irritancy Classification D-1
D1 Data Sorted by Reference D-3
D2 Data Sorted by Substance D-11

Appendix E Interlaboratory Correlation Coefficients from the EC/HO Validation Study (Balls et al. 1995) E-1

LIST OF TABLES

Table 1-1 Summary of Current U.S. Legislation Related to Ocular Health 1-2
Table 1-2 *In Vivo* Ocular Irritancy Classification Systems..... 1-5
Table 2-1 Evaluation of Corneal Irritation..... 2-10
Table 2-2 Fluorescein Penetration Scoring System..... 2-12
Table 2-3 Overall Scoring System for Corneal Damage and Irritation… 2-13
Table 3-1 Chemical Classes Tested in the IRE Test Method..... 3-4
Table 3-2 Product Classes Tested in the IRE Test Method 3-4
Table 4-1 Scale of Weighted Scores for Grading the Severity of Ocular Lesions 4-2
Table 4-2 Test Guidelines for *In Vivo* Ocular Irritation Test Methods 4-4
Table 4-3 Criteria for Classification of Rabbits According to the GHS Classification System..... 4-11
Table 4-4 Criteria for Classification of Substances According to the GHS Classification System (Modified from UN 2003). 4-12
Table 4-5 Criteria Required for Classification of Rabbits According to the EPA Classification System (EPA 1996). 4-13
Table 4-6 Criteria for Classification of Substances According to the EU Classification System (EU 2001) 4-13
Table 6-1 Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by Study 6-5
Table 6-2 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 Expanded Data Set) 6-7
Table 6-3 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) 6-9
Table 6-4 Evaluation of the Performance of the IRE Test Method in Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by Study 6-11
Table 6-5 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EPA Classification System (Analysis Based on the Expanded Data Set) 6-13
Table 6-6 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EPA Classification System (Analysis Based on the Pooled Data Set) 6-14

Table 6-7 Evaluation of the Performance of the IRE Test Method in Predicting Ocular Corrosives and Severe Irritants Compared to *In Vivo* Findings, as Defined by the EU Classification System, by Study 6-16
Table 6-8 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EU Classification System (Analysis Based on the Expanded Data Set) 6-19
Table 6-9 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EU Classification System (Analysis Based on the Pooled Data Set) 6-20
Table 7-1 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/ Nonirritants Using the GHS Classification System……7-4
Table 7-2 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EPA Classification System 7-5
Table 7-3 Interlaboratory Variability of CEC Collaborative Study (1991) for Substances Classified as Ocular Corrosives/Severe Irrittants or Nonsevere Irritants/Nonirritants Using the EU Classification System 7-7
Table 7-4 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EU Classification System .....……7-8
Table 7-5 Quantitative Evaluation of the Interlaboratory Variability of the IRE Test Method (CEC 1991). 7-11
Table 7-6 Quantitative Evaluation of the Interlaboratory Variability of the IRE Test Method (Balls et al. 1995) 7-13

Table 7-7 Interlaboratory Correlation Ranges Determined for Various Subsets of Tested Substances in Balls et al. (1995).....7-18
Table 9-1 *In Vitro*/*In Vivo* Correlation Coefficients from Balls et al. (1995).....9-3
Table 9-2 Relationship Between MAS *In Vitro* and the Ability to Cause More Than 15% Corneal Swelling *In vitro* (Chamberlain et al. 1997).....9-4
Table 9-3 Comparison of IRE Test Method Results With *In Vivo* Data (Cooper et al. (2001). 9-4
Table 9-4 The Results of the CTFA Evaluation of *In Vitro* Alternatives to the Draize Primary Eye Irritation Test (Phase III): Surfactant-Based Formulations (Gettings et al. 1996)..... 9-5
Table 9-5 Comparison of IRE *In Vitro* Irritancy Grades to *In Vivo* Rabbit Eye Test Irritancy Classifications (Koëter and Prinsen 1985)..... 9-9
Table 9-6 Irritancy of Surfactant-Based Products Using P&G *In Vivo* and ExRET *In Vitro* Confocal Microscopy Test Methods. 9-15
Table 9-7 Irritancy of General Chemicals Using P&G *In Vivo* and ExRET *In Vitro* Confocal Microscopy Test Methods.....9-16

Table 9-8 Irritancy of Surfactant-Based Liquid Dishwashing Formulations Using LVET and Histopathology *In Vivo* and P&G ExRET Confocal Microscopy *In Vitro* Test Method 9-17
Table 9-9 Irritancy of Bleach-Containing Laundry Additive Products Using LVET *In Vivo* and P&G ExRET *In Vitro* Confocal Microscopy Test Methods 9-17

LIST OF FIGURES

Figure 1-1 Anatomy of the Human Eye. 1-8
Figure 1-2 GHS Testing Strategy for Serious Eye Damage and Eye Irritation…… 1-18
Figure 2-1 Sample Scoring Form for the IRE Test Method.. 2-17

LIST OF ACRONYMS AND ABBREVIATIONS

3 R's	Replacement, Reduction, and Refinement
AAALAC	Association for Assessment and Accreditation of Laboratory Animal Care
Anal.	Analytical method used to transform the sample data into BCOP classification
ANOVA	Analysis of variance
ASTM	American Society for Testing Materials
BP	British Pharmacopeia
BCOP	Bovine Corneal Opacity and Permeability
BRD	Background Review Document
C	Corrosive
CAS	Chemical Abstract Service
CASRN	Chemical Abstracts service Registry Number
CEC	Commission of European Communities
CGRP	Calcitonin Gene Related Peptide
CO	Corneal opacity
CPSC	U.S. Consumer Protection Safety Commission
CS	Corneal swelling
CTFA	Cosmetic Toiletry and Fragrance Association
CV	Coefficient of variation
DOI	Depth of injury
EC	European Commission
EC/HO	European Commission/British Home Office
ECETOC	European Centre for Ecotoxicology and Toxicology
ECVAM	European Centre for the Validation of Alternative Methods
EDTA	Ethylenediaminetetraacetic acid
EEC	European Economic Community
EPA	U.S. Environmental Protection Agency
EPA OPPTS	U.S. Environmental Protection Agency Office of Prevention, Pesticides and Toxic Substances

ESL	Environmental Safety Laboratory
EU	European Union
ExRET	Ex Vivo Rabbit Eye Test
FDA	U.S. Food and Drug Administration
FHSA	Federal Hazardous Substance Act
FIFRA	Federal Insecticide, Fungicide and Rodencide Act
FRt	Fluorescein retention
FR	*Federal Register*
FRAME	Fund for Replacement of Animals in Medical Experiments
g	Gram
GHS	Globally Harmonized Classification System
GLP	Good Laboratory Practice
GSK	GlaxoSmithKline
HET-CAM	Hen's Egg Test Chorioallantoic membrane
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye
IL-1	Interleukin-1
IRE	Isolated Rabbit Eye
IACAC	Institutional Animal Care and Use Committee
I.H.E.	Institute for Hygeine and Epidemiology
Kg	Kilogram(s)
L	Liter(s)
LVET	Low Volume Eye Test
MAS	Maximum Average Score
mg	Milligram(s)
mg/kg	Milligram(s) per kilogram
mL	Milliliter(s)
mL/min	Milliliter(s) per minute
mm	Millimeter(s)
MMAS	Modified Maximum Average Score
MMPs	Matrix metalloproteinases

MT	Mottling
NA	Not applicable
NDI	Normalized Depth of Injury
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NS	Not specified
OECD	Organization for Economic Co-Operation and Development
TG 405	Test Guideline 405
OPPTS	Office of Prevention, Pesticides, and Toxic Substances/U.S. EPA
OSHA	Occupational Safety and Hazardous Administration
OT	Other
PBS	Phosphate buffered saline
P.L.	Public law
PLA_2	Phospholipase A_2
pos	Positive
PT	Pitting
QA	Quality assurance
r	Correlation coefficient
REET	Rabbit Enucleated Eye Test
RN	Registry number
Sample ID	Sample identification
SD	Standard deviation
SL	Sloughing
SOT	Society of Toxicology
ST	Stippling
TG	Test Guideline
TNF	Tumor necrosis factor
TNO-CIVO	Institute CIVO – Toxicology and Nutrition (TNO)
TSCA	Toxic Substance Control Act
U.S.	United States
UL	Ulceration

μL	Microliter(s)
UN	United Nations
WHO	World Health Organization
YSI	Yellow Springs Instruments
Zeneca CTL	Central Testing Laboratory

ACKNOWLEDGEMENTS

The following individuals are acknowledged for their contributions to the In Vitro Ocular Test Method review process:

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Ocular Toxicity Working Group (OTWG)

Consumer Product Safety Commission (CPSC)

Marilyn Wnd, Ph.D.
Kailash Gupta, Ph.D.

Department of Defense (DOD)

Harry Salem, Ph.D.

Department of Transportation (DOT)

Steve Hwang, Ph.D.

Environmental Protection Agency (EPA)

Karen Hamernik, Ph.D. (OTWG, Co-Chair)
Angela Auletta, Ph.D.
Meta Bonner, Ph.D.
Andrew Geller, Ph.D.
Karen Hicks
Marianne Lewis
Debbie McCall
Mark Perry, Ph.D.
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration (FDA)

Leonard Schechtman, Ph.D. (ICCVAM Chair)
Jill Merrill, Ph.D. (OTWG, Co-Chair)
Bob Bronaugh, Ph.D.
Paul Brown, Ph.D.
Wiley Chambers, M.D.
Zhou Chen
Abby Jacobs, Ph.D.
Donnie Lowther
Josie Yang, Ph.D.

National Institute for Occupational Safety & Health (NIOSH)

Surender Ahir, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)

Raymond (Buck) Grissom, Ph.D.
William Stokes, D.V.M., D.A.C.L.A.M. (Executive Director, ICCVAM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

David Allen, Ph.D.
ILS, Inc.

Bradley Blackard, M.S.P.H.
ILS, Inc.

Sue Brenzel
ILS, Inc.

Thomas A Burns, Jr. M.S.
ILS, Inc.

Patricia Ceger, M.S.
ILS, Inc.

Jeffrey Charles, Ph.D., M.B.A., D.A.B.T.
ILS, Inc.

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

Christina Inhof, M.S.P.H.
ILS, Inc. (until March 2005)

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

William Stokes, D.V.M., Diplomate, A.C.L.A.M. (Director)
NIEHS

Judy Strickland, Ph.D., D.A.B.T.
ILS, Inc.

Raymond Tice, Ph.D. (Deputy Director)
NIEHS

James Truax, M.A.
ILS, Inc.

Additional Reviewers for the *In Vitro* Ocular Corrosion and Irritation Test Methods Background Review Documents

Chantra Eskes, Eng., Ph.D.
ECVAM
Ispra, Italy

Robert L Guest Bsc, CBiol, MlBiol
SafePharm Laboratories, Ltd.
Derby, United Kingdom

John Harbell, Ph.D.
Institute for *In Vitro* Sciences
Gaithersburg, Maryland

Joe Haseman, Ph.D.
Consultant
Raleigh, North Carolina

Penny Jones
Unilever Research
Sharnbrook, United Kingdom

Menk Prinsen
TNO Nutrition & Food Research Institute
The Netherlands

Horst Spielmann, Dr. med.
ZEBET
Berlin, Germany

Companies and Individuals that Provided *In Vitro* and/or *In Vivo* Data for the IRE Test Method Background Review Document

NICEATM gratefully acknowledges the generous contributions of the companies and individuals who provided data for this review of the IRE test method. Their time and efforts to sort through study archives and compile data for this document are greatly appreciated.

ECVAM
Chantra Eskes, Ph.D.

GlaxoSmithKline
Frederick J. Guerriero

The Proctor and Gamble Company
Dan Marsman, Ph.D.

Wetenschappelijk Instituut Volksgezondheit
Guido Jacobs

[This Page Intentionally Left Blank]

PREFACE

During the past 60 years, government regulatory agencies have implemented safety-testing requirements to identify potential hazards of various chemicals and products to protect human health and the environment. Testing results are used for hazard classification and labeling and to identify appropriate risk management practices necessary to reduce or avoid human injury, disease, disability, and/or death. The first standardized toxicity test method developed for assessing the safety of a chemical ingredient or new product was for chemically induced eye injuries (Draize et al. 1944). The U.S. Food and Drug Administration (FDA) developed this test in response to new laws implemented as a result of permanent eye injuries from various cosmetic products in the 1930s (Calabrese 1983). Various national and international regulatory authorities now require updated versions of this test method to assess whether substances can potentially cause eye irritation or corrosion. The U.S. Consumer Product Safety Commission (CPSC), the U.S. Environmental Protection Agency (EPA), FDA, and the U.S. Occupational Health and Safety Administration (OSHA) have testing requirements and guidelines in place for assessing the ocular irritation of various substances such as pesticides, hazardous household products, pharmaceuticals, cosmetics, and other agricultural and industrial chemicals.

While ocular safety assessments have clearly supported appropriate protection of consumers and workers, there have been concerns raised about the humane aspects of this test method. Various modifications to the Draize rabbit eye test (Draize et al. 1944) have now been adopted by regulatory authorities that reduce the numbers of animals used and that reduce the potential pain and distress associated with the procedure. Significant progress has been made during the last decade, with only one to three rabbits now required per test compared to six rabbits in the original protocol, and addition of provisions that allow for humane euthanasia of animals with severe lesions or discomfort. In addition, a number of scientists and organizations began to develop nonanimal alternatives in the early 1980s that might be useful in further reducing or replacing the need for animals for the assessment of ocular irritancy and corrosion. Although a great deal of progress has been made, there is currently no accepted nonanimal alternative test method for ocular irritancy in the United States.

Cognizant of various *in vitro* methods that had been developed and have undergone some degree of validation, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) recommended in August 2003 that ICCVAM give high priority to reviewing the validation status of *in vitro* test methods proposed for identifying ocular irritants/corrosives. In October 2003, the EPA formally nominated several ocular irritation test methods and related activities for evaluation by ICCVAM. This included review of the validation status of four *in vitro* methods for identifying potential ocular corrosives and severe irritants in a tiered testing strategy. Validation[1] of a test method is a prerequisite for it to be considered for regulatory acceptance (ICCVAM 1997, 2003). The four test methods were the Bovine Corneal Opacity and Permeability (BCOP) assay, the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay, the Isolated Chicken Eye

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

(ICE) assay, and the Isolated Rabbit Eye (IRE) assay.

ICCVAM, which is charged with coordinating the technical evaluations of new, revised, and alternative test methods with regulatory applicability (ICCVAM Authorization Act of 2000, Public Law [P.L.] 106-545), unanimously agreed that the four nominated *in vitro* test methods should have a high priority for evaluation. An ICCVAM Ocular Toxicity Working Group (OTWG) was established to work with the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to carry out these evaluations. ICCVAM and NICEATM also collaborate closely with the European Centre for the Validation of Alternative Methods (ECVAM), a component of the European Commission's Joint Research Centre. Accordingly, an ECVAM liaison was designated for the ICCVAM OTWG to ensure input and contributions during the evaluation and review process.

NICEATM, which administers the ICCVAM and provides scientific support for ICCVAM activities, subsequently prepared four comprehensive background review documents (BRDs) that provided information and data about the current validation status of the four nominated *in vitro* test methods (i.e., BCOP, HET-CAM, ICE, and IRE) for detecting ocular corrosives and severe irritants. These draft BRDs were based on published studies using the identified test methods, and other data and information submitted in response to a 2004 *Federal Register* (*FR*) request (Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), and were made available to the public on November 1, 2004 (Available: http://iccvam.niehs.nih.giv/methods/ocudocs/ocu_brd.htm). Notification for data also was made through the ICCVAM electronic mailing list.

ICCVAM subsequently convened an Expert Panel meeting on January 11-12, 2005, to independently assess the validation status of these four *in vitro* test methods for identifying ocular corrosives or severe irritants. Prior to this meeting, public comments on the Addendum were received from three organizations and provided to the Expert Panel for their consideration. Public comments at the meeting revealed that additional relevant data was available that had not previously been provided in response to earlier requests for data. The Expert Panel recommended that the additional data be requested and that a reanalysis of the accuracy and reliability of each test method be conducted, where appropriate (the Expert Panel report from this meeting is available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm).

In response to this recommendation, an *FR* notice was published on February 28, 2005 (Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), which requested all available *in vitro* data on these four *in vitro* ocular irritancy test methods and corresponding *in vivo* rabbit eye test method data, as well as any human exposure data (either via ethical human studies or accidental exposure). A request for relevant data was re-sent directly to the primary developers or users of each test method. In response to these requests, additional *in vitro* test method data and corresponding *in vivo* rabbit eye test results were submitted for the BCOP, HET-CAM, and ICE test methods. These additional data were used to update the performance statistics of the test methods. Several U.S. Federal agencies (OSHA, CPSC, and the National Institute for Occupational Safety and Health [NIOSH]), along with the US Eye

Injury Registry (USEIR) were also contacted directly for data resulting from accidental human exposures. However, given the lack of details about the specific nature of the substances reported and their associated exposure conditions, these types of accidental human exposure injury data were not useful for evaluating the accuracy of the IRE test method for predicting human ocular hazard.

Further clarification of hazard classification rules for severe irritants also was obtained subsequent to the release of the four draft BRDs. This change resulted in a small number of substances previously classified as nonsevere irritants now being classified as severe irritants (from 10 to 15, depending on the test method and the classification system used). This change necessitated a reanalysis of the accuracy and reliability of all four of the test methods previously evaluated.

The original draft BRDs also provided an evaluation of the accuracy of each test method by chemical class. Subsequent to the release of the draft BRDs, the chemical classes assigned to each test substance were revised based on a chemical classification system consistent with the U.S. National Library of Medicine's Medical Subject Headings (MeSH; Available: http://www.nlm.nih.gov/mesh), an internationally recognized standardized classification scheme. This scheme was used to ensure consistency in classifying substances by chemical class among all the *in vitro* ocular test methods under consideration, and resulted in some chemicals being re-classified into different chemical classes. As a result, the accuracy of each test method by chemical class was reanalyzed.

To incorporate the additional data submitted, the changes in irritancy classification, and the revised chemical classes, a BRD Addendum was developed. The purpose of this document was to highlight changes in the performance statistics due to the above noted updates. The BRD Addendum was released on July 26, 2005, with notification of its release via an *FR* notice and notification through the ICCVAM electronic mailing list (and is available in electronic format on the ICCCVAM/NICEATM website, http://iccvam.niehs.nih.gov/methods/ocudocs/reanalysis.htm). The Expert Panel was subsequently reconvened via public teleconference on September 19, 2005 to discuss the BRD Addendum. Prior to this meeting, public comments on the Addendum were received from three organizations and provided to the Expert Panel for their consideration (no public comments were provided during the public teleconference). The Expert Panel then provided final endorsement regarding the effects, if any, of the information in the BRD Addendum on their original evaluation from the January 11-12, 2005 meeting (the Expert Panel report from this meeting is available at http://iccvam.niehs.nih.gov/methods/ocudocs/EPreport/EPrptAddend.htm).

NICEATM has subsequently prepared revised BRDs to reflect a compilation of the updated information for each test method. Each BRD provides a comprehensive summary of the current validation status of the *in vitro* test method, including what is known about its reliability and accuracy, and the scope of the substances tested. Raw data for these test methods will be maintained for future use. Therefore, the performance statistics of these test methods will be updated as additional information becomes available.

The ICCVAM and its OTWG will consider both Expert Panel reports, the updated performance statistics presented in the BRDs, and any public comments in preparing its final test method recommendations for these *in vitro* ocular test methods. These recommendations will be made available to the public and provided to the U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (Public Law 106-545) (Available: http://iccvam.niehs.nih.gov/about/PL106545.pdf).

We want to acknowledge the excellent cooperation and contributions from the many organizations and scientists who provided critical data and information necessary for this BRD. The efforts of the many individuals who contributed to the preparation of this BRD also are gratefully acknowledged. These include David Allen, Ph.D., Bradley Blackard, M.S.P.H., Thomas A. Burns, Jr. M.S., Jeffrey Charles, Ph.D., M.B.A., D.A.B.T., Neepa Choksi, Ph.D., and James Truax, M.A. of Integrated Laboratory Systems (ILS), Inc., the NICEATM Support Contractor, as well as the members of the ICCVAM OTWG and ICCVAM representatives who reviewed various drafts. We also want to thank Raymond Tice, Ph.D., Deputy Director of NICEATM, for his extensive efforts on this project. Finally, we want to recognize the excellent leadership of the OTWG Co-chairs, Dr. Karen Hamernik (U.S. Environmental Protection Agency) and Dr. Jill Merrill (U.S. Food and Drug Administration).

William S. Stokes, D.V.M. Diplomate A.C.L.A.M.
Director, NICEATM
Executive Director, ICCVAM

Leonard Schechtman, Ph.D.
U.S. Food and Drug Administration
National Center for Toxicological Research
Chairman, ICCVAM

March, 2006

EXECUTIVE SUMMARY

This Background Review Document (BRD) reviews available data and information regarding the validation status of the Isolated Rabbit Eye[1] test method for identifying ocular corrosives and severe irritants. The test method was reviewed for its ability to predict ocular corrosives and severe/irreversible effects as defined by the U.S. Environmental Protection Agency (EPA) (EPA 1996), the European Union (EU) (EU 2001), and the United Nations (UN) Globally Harmonized System (GHS) of Classification and Labelling of Chemicals (UN 2003). The objective of this BRD is to describe the current validation status of the IRE test method, including what is known about its accuracy and reliability, the scope of the substances tested, and the availability of a standardized test method protocol.

The information summarized in this BRD is based on publications obtained from the peer-reviewed literature, as well as unpublished information submitted to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in response to two *Federal Register* (*FR*) Notices requesting high quality *in vivo* rabbit eye test data and *in vitro* ocular irritation data for the IRE, Isolated Chicken Eye (ICE), Bovine Corneal Opacity and Permeability (BCOP), and the Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) test methods. An online literature search identified 14 publications that contained IRE test method results and protocol information; of these publications, detailed *in vivo* data were obtained for four studies. Submitted IRE and detailed *in vivo* data for these four studies allowed for an evaluation of test method accuracy[2] and reliability[3].

Other published and unpublished IRE test method studies are reviewed in **Section 9.0** (Other Scientific Reports and Reviews). This section discusses IRE studies that could not be included in the performance analyses because of the lack of appropriate study details or test method results and/or the lack of appropriate *in vivo* rabbit eye reference data.

The IRE test method was developed by Burton et al. (1981) and proposed as a preliminary *in vitro* screen for the assessment of severe eye irritants. This organotypic test method is also referred to as the Rabbit Enucleated Eye Test (REET) (e.g., Guerriero et al. 2004). The principal advantage of the IRE test is that it eliminates the use of live animals for ocular irritancy testing and thus the pain and suffering potentially associated with the *in vivo* Draize rabbit eye test. Another advantage of the IRE test method is that it typically uses eyes isolated from euthanized rabbits used for other research purposes or from animals sacrificed commercially as a food source. In the IRE, liquid or solid substances are placed directly on the corneal surface of isolated rabbit eyes, which are held and maintained in a temperature-

---

[1] Exposure of the isolated rabbit eye to irritants can produce corneal opacity, corneal swelling, an increase in permeability to sodium fluorescein dye, and observable effects on the corneal epithelium. These endpoints can be quantified and used to evaluate the potential eye irritation of substances.

[2] (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance". The term is often used interchangeably with "concordance."

[3] A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

controlled chamber. After a 10-second exposure, followed by rinsing, the treated eye may be evaluated for corneal opacity, corneal swelling, fluorescein penetration, and effects on the corneal epithelium at various times over a four-hour observation period. Various decision criteria based on use of one to four ocular endpoints have been employed to identify ocular irritants of varying levels of severity depending on the nature of the test substances used (e.g., surfactant-based or chemically diverse) (Burton et al. 1981; Chamberlain et al. 1997; Cooper et al. 2001; Jones et al. 2001; Gettings et al. 1996). However, Guerriero et al. (2004) provided decision criteria (prediction model) using all four of these endpoints to specifically identify chemically diverse test substances as ocular corrosives or severe irritants.

The IRE test method has not yet been considered by U.S. Federal agencies for regulatory use where submission of testing data is required. However, some companies have found the IRE test method useful for the identification of ocular corrosives and severe irritants in a tiered testing strategy on a case-by-case basis. Negative results and suspected false positive *in vitro* results proceed to standard *in vivo* testing or to validated *in vitro* test methods that are capable of detecting false negative corrosives and severe irritants.

The IRE test method protocols used in the various studies considered in this BRD are similar, but not identical. The essential principles of the test method protocol include procurement and enucleation of the eyes, a preliminary assessment of the integrity of the isolated eye (especially the corneal surface), equilibration of the eyes in a physiological environment, application of the test substance, incubation, temporal quantization of corneal damage using various endpoints (i.e., corneal opacity score, corneal swelling calculation, fluorescein penetration score, and evaluation of epithelial integrity), evaluation of data in relation to a prediction model, and assignment of an irritancy level based on graded responses (e.g., nonirritant, mild, moderate, or severe irritant) or as all or none responses (e.g., either a nonsevere irritant or a corrosive/severe irritant). However, given the various uses and applications of the IRE test method by different investigators and laboratories, and the evolution of the test method over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method. Variations in the publicly available IRE protocols include evaluation of one to four endpoints, different prediction models or *in vitro* classification systems, and differences in the number of controls, among other methodological variations.

Some of the published *in vivo* rabbit eye test data on the substances used to evaluate the accuracy of IRE for detecting ocular corrosives and severe irritants was limited to average score data or a reported irritancy classification based on a laboratory specific classification scheme. 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 were necessary to calculate the appropriate EPA (1996), EU (2001), and GHS (UN 2003) ocular irritancy hazard classifications. Thus, a portion of the test substances for which there was only limited *in vivo* data could not be used for evaluating test method accuracy as described in this BRD.

Only a few of the reports provided original *in vitro* test result data. However, summary *in vitro* data were available for all of the test substances evaluated, such that they could be

assigned *in vitro* irritancy classifications for comparison to the available *in vivo* reference data.

A total of 149 substances and formulations obtained from four studies that included a variety of chemical and product classes have been tested in the IRE test method. The chemical classes with the greatest amount of *in vitro* IRE data are alcohols, amines, carboxylic acids, esters, formulations, heterocyclic compounds, ketones, onium compounds (e.g., ammonium nitrate), and sulfur compounds. The formulations tested include hair shampoos, soaps, personal care cleansers, and detergents. Most common product classes tested in the IRE assay are active pharmaceutical ingredients, chemical/synthetic intermediates, cleaners, raw materials, soaps and detergents, solvents, and surfactants.

The existing database of substances tested using the four ocular endpoints needed to identify a severe irritant (corneal opacity, corneal swelling, fluorescein penetration, and epithelial integrity) was limited to the Guerriero et al. (2004) study. Because this was a small dataset (n=38), substances in the CEC (1991), Balls et al. (1995), and Gettings et al. (1996) studies that had been identified as ocular corrosives/severe irritants using appropriate decision criteria for identification of a severe irritant (i.e., a corneal opacity score greater than or equal to 3, or a corneal swelling equal to or greater than a 25%) were considered together with the test results obtained by Guerriero et al. (2004). This database is referred to as the "Expanded Data Set." This database has limitations, however, since negative (i.e., true or false negative) outcomes are not considered in those studies using fewer than four ocular endpoints.

Substances that were identified as ocular corrosives/severe irritants based on *in vitro* results by any single endpoint were, therefore, included as part of the expanded data set. Substances in the CEC (1991), Balls et al. (1995), and Gettings et al. (1996) studies that were identified as nonsevere irritants or nonirritants, based on *in vitro* results, were not included in the expanded data set. These substances were not included because an evaluation that included any of the omitted endpoints might have resulted in a severe irritant classification. For example, a substance that did not produce ≥ 25% corneal swelling might have produced a corneal opacity score, fluorescein penetration score, or damage to the epithelium that would have classified it as a severe irritant had these endpoints been evaluated.

A pooled data set consisting of substances from all available studies within a regulatory classification system was also analyzed. For example, using the GHS classification system, data from the Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004) studies were pooled for this analysis. While this pooled data set included all available data within a classification system, it was also limited by variability in the number of ocular endpoints. With the exception of the Guerriero et al. (2004) data in which four endpoints were used, the number of endpoints ranged from one (i.e., corneal swelling) to three (i.e., corneal opacity, corneal swelling, and fluorescein retention) in the other studies. Having less than four ocular endpoints could potentially reduce the likelihood of a positive response using the BRD all-or-none decision criteria.

The accuracy evaluation of the IRE test method was limited to the substances evaluated in four *in vitro-in vivo* comparative studies. The ability of the IRE test method to correctly

identify ocular corrosives and severe irritants, as defined by the EPA (1996), the EU (2001), and the GHS (UN 2003) was evaluated using two approaches. In the first approach, the accuracy of IRE was assessed separately for each *in vitro-in vivo* comparative study using the decision criteria (prediction model) of Guerriero et al. (2004), where possible, to identify corrosives/severe irritants. In the second approach, the accuracy of IRE was assessed after pooling data across *in vitro-in vivo* comparative studies that used similar protocols, same method of data collection, and the decision criteria of Guerriero et al. (2004). While there were some differences in results among the three hazard classification systems evaluated (i.e., EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that IRE test method performance was comparable among the three hazard classification systems. The overall accuracy of the IRE test method obtained by pooling all studies ranged from 64% to 69%, depending on the classification system used. Sensitivity and specificity ranged from 69% to 76% and 60% to 65%, respectively. The false positive rate ranged from 35% to 40%, while the false negative rate ranged from 24% to 30%. When the analysis is restricted to Guerriero et al. (2004) in which the four ocular endpoints were used in the decision process, an accuracy of 79%, a sensitivity of 100%, and a specificity of 70% were obtained across all classification systems. In this analysis, the false positive rate was 30% and the false negative rate was 0% across all classification systems.

For the expanded data set and using the GHS ocular hazard classification system, the accuracy was 68% (52/76), the false positive rate was 56% (24/43), and the false negative rate was 0% (0/33). The expanded data set used for this evaluation include the 38 substances evaluated by Guerriero et al. (2004) and an additional 38 substances tested by Balls et al. (1995) and Gettings et al. (1996) and classified by IRE as severe irritants, 22 of which were also severe irritants *in vivo* and 16 of which were nonsevere irritants or nonirritants *in vivo*. The performance of the expanded data set is potentially confounded by the exclusion of substances with true negative outcomes (matching *in vivo* and *in vitro* nonsevere or nonirritant classifications), which would affect both specificity and the false negative rate.

Using the expanded data set, the chemical classes that were overpredicted (i.e., were false positives) in the IRE test method according the GHS classification system were ketones (67%, [4/6]), esters (67%, [4/6]), and alcohols (60%, [6/10]). Among the 10 surfactants tested, the false positive rate was 67% (2/3) and the false negative rate was 0% (0/7). The seven cationic surfactants included in this group had a false positive rate of 100% (1/1) and a false negative rate of 0% (0/6). Twelve surfactant-based formulations had a false positive rate of 100% (2/2) with no false negative outcomes (0/10).

For the pooled data set (when results were compared to the GHS ocular hazard classification system) the accuracy was 65% (70/107), the false negative rate was 30% (14/47), and the false positive rate was 38% (23/60). The pooled data set used for this evaluation includes 38 substances evaluated by Guerriero et al. (2004), 54 substances tested by Balls et al. (1995), and 24 tested by Gettings et al. (1996). For nine substances tested in common, consensus regulatory calls were used for comparison of *in vitro* and *in vivo* data.

In order to further evaluate discordant responses of the IRE test method relative to the *in vivo* hazard classification, several accuracy subanalyses were performed using both the expanded

data set and the pooled data set. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., pesticides, surfactants, pH, physical form). Because the international community will soon adopt the GHS classification system for hazard labeling (UN 2003), and considering that there were only modest differences in overall IRE test method accuracy among the three regulatory classification systems (i.e., EPA, EU, GHS), these sub-analyses are focused only on the GHS classification system, using the expanded data set.

Using the expanded data set, with regard to physical form of the substances overpredicted by the IRE test method, liquids had a higher overprediction rate (83%, [19/23]) than solids (25%, [5/20]). The highest false positive rate, based on pH was 33% (2/6) for substances with pH > 7.

No substances in the expanded data set were underpredicted (i.e., were false negatives) by the IRE test method. Thus, an analysis of underprediction based on chemical class, physical form, pH, or NICEATM GHS Category I subclassification was not possible.

Using the pooled data set, the chemical classes that were overpredicted (i.e., were false positives) in the IRE test method according to the GHS classification system were ketones (67%, [4/6]), alcohols (55%, [6/11]), and amines (50%, [3/6]). Among the 13 surfactants tested, 40% (2/5) were overpredicted and 12% (1/8) were underpredicted. Of 25 surfactant-based formulations, 25% (2/8) were overpredicted and 38% (6/16) were underpredicted.

Using the pooled data set, with regard to physical form of the substances overpredicted by the IRE test method, liquids had a higher overprediction rate (49%, (18/37) than solids (22%, [5/23]). The highest false positive rate, based on pH was 33% (2/6) for substances with pH > 7.

In the pooled data set, the highest underprediction rate (i.e., were false negatives) was for carboxylic acids (67%, [4/6]) and organic compounds (50%, [3/6]). The underprediction rate for liquids and solids were similar at 29% (8/28) and 32% (6/19), respectively. The underprediction rate for surfactants was 12% (1/8) and for surfactant-based formulations was 38% (6/16). Underprediction rates of 25-37% (1/4 to 7/19) were obtained for Category 1 subgroups 1 to 4.

In the original draft IRE BRD (NICEATM 2004), no data was provided for the assessment of intralaboratory repeatability and reproducibility. Therefore, an analysis of intralaboratory reliability still could not be conducted.

The original IRE test method reliability analysis included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. While the quantitative analysis was unaffected by the reclassification of the ocular irritancy of some test substances, the qualitative analysis (correct classification as an ocular corrosive/severe irritant or as a non-corrosive/non-severe irritant) of the individual laboratory test results obtained for the

EC/HO validation study (Balls et al., 1995) and for the CEC (1991) collaborative study was affected.

Overall, in the Balls et al. (1995) study, the number of substances with 100% agreement among the four participating laboratories was 59 to 63% (35 to 37/59). The number of substances with 75% agreement among laboratories was 22 to 25% (13 to 15/59). The number of substances with 50% agreement among laboratories was 15% (9/59).

Overall, in the CEC (1991) study, the number of substances with 100% agreement among the three participating laboratories was 81% (17/21). The number of substances with 67% agreement among laboratories was 14% (3/21), while the number of substances with 33% agreement was 5% (1/21).

As stated above, this BRD provides a comprehensive summary of the current validation status of the IRE test method, including what is known about its reliability and accuracy, and the scope of the substances tested. Raw data for the IRE test method will be maintained for future use, so that these performance statistics may be updated as additional information becomes available.

1.0 INTRODUCTION AND RATIONALE FOR THE PROPOSED USE OF *IN VITRO* TEST METHODS TO IDENTIFY OCULAR CORROSIVES AND SEVERE IRRITANTS

1.1 Introduction

1.1.1 Historical Background of *In Vitro* Ocular Irritation/Corrosion Test Methods and Rationale for Their Development

The location of the eye and its anatomy predisposes it to exposure to a variety of environmental conditions (e.g., ozone, pollen) and substances on a daily basis. Injury from ocular exposure to a variety of chemical agents can lead to a range of adverse effects with the most extreme being blindness. Societal concern for evaluating consumer products for ocular irritation and/or corrosion was heightened in 1933 when a 38 year old woman went blind after her eyelashes and eyebrows were tinted with a product containing paraphenylenediamine, a chemical with the potential to cause allergic blepharitis, toxic keratoconjunctivitis, and secondary bacterial keratitis[1] (Wilhelmus 2001).

In 1938, the U.S. Congress responded to these concerns by enacting the Federal Food, Drug, and Cosmetic Act of 1938, which included extending the regulatory control of the U.S. Food and Drug Administration (FDA) to cosmetics (FDA 1938). This legislation required manufacturers to evaluate product safety before marketing their products (Wilhelmus 2001). Several additional legislative statutes were later enacted to enable government agencies to regulate a variety of substances that could pose a risk to ocular health. **Table 1-1** provides a synopsis of current U.S. regulatory laws that pertain to eye irritation and corrosion.

Table 1-1 Summary of Current U.S. Legislation Related to Ocular Health[1]

Legislation (Year of Initial Enactment)	Agency	Substance
Food, Drug and Cosmetic Act (1938)	FDA	Pharmaceuticals and cosmetics
FIFRA (1947) and Federal Environmental Pesticide Control Act (1972)	EPA	Pesticides
FHSA (1964)	CPSC	Household products
FHSA (1964) and TSCA (1976)	Department of Agriculture and EPA	Agricultural and industrial chemicals
Occupational Safety and Health Act (1970)	OSHA	Occupational materials
Clean Air Act Amendments (1990)	Chemical Safety and Hazard Investigation Board and EPA	Accidentally released chemicals and air pollutants

[1]Adapted from Wilhelmus (2001).
Abbreviations: CPSC = U.S. Consumer Product Safety Commission; EPA = U.S. Environmental Protection Agency; FDA = U.S. Food and Drug Administration, FHSA = Federal Hazardous Substances Act; FIFRA = Federal Insecticide, Fungicide, and Rodenticide Act; TSCA = Toxic Substances Control Act.

---

[1] Allergic blepharitis (also referred to as blepharitis): inflammation of the eyelids; Toxic keratocojunctivitis (also referred to as contact, irritative, or chemical keratoconjuctivitis): inflammation of the cornea and conjunctiva due to contact with an exogenous agent; Secondary bacterial keratitis: inflammation of the cornea that occurs secondary to another insult that compromised the integrity of the eye (Vaughn et al. 1999; Chambers W, personal communications).

Exposure of the eye of a rabbit to a test substance is the primary method for assessing the hazard potential of substances that may come in contact with or be placed near the eye of a human. The rabbit eye test method currently accepted by U.S. Federal and international regulatory agencies (FHSA 1964; EPA 1998; OECD 2002) is based on a method developed by Draize and colleagues in 1944 (Draize et al. 1944). This technique involves placing a test substance into the lower conjunctival sac of one eye of a rabbit. The contralateral eye serves as a negative control. The rabbit is then observed at selected intervals for up to 21 days after exposure for adverse effects to the conjunctiva, cornea, and iris.

The current rabbit eye test method identifies both irreversible (e.g., corrosion) and reversible ocular effects. It also provides scoring that allows for relative categorization of severity for reversible effects such as mild, moderate, or severe irritants (e.g., see U.S. Environmental Protection Agency [EPA] Ocular Classification System discussed below). Current EPA ocular testing guidelines and the United Nations (UN) Globally Harmonized System (GHS) of Classification and Labeling of Chemicals (UN 2003) indicate that if serious ocular damage is anticipated (e.g., irreversible adverse effects on day 21), then a test on a single animal may be considered. If serious damage is observed, then no further animal testing is necessary (EPA 1998; UN 2003). If serious damage is not observed, additional test animals (1 or 2 rabbits) may be evaluated sequentially until concordant irritant or nonirritant responses are observed (UN 2003).

Depending on the legislative mandate of various regulatory agencies and their goals for protecting human health, the classification of irritant responses evaluated by each agency varies (**Table 1-2).** The EPA ocular irritation classification regulation and testing guidelines (EPA 1996, 1998) are based on the most severe response in one animal in a group of three or more animals. This classification system takes into consideration the kinds of ocular effects produced, as well as the reversibility and the severity of the effects. The EPA classifies substances into four ocular irritant categories, ranging from I to IV (**Table 1-2**) (EPA 1996). Category I substances are defined as corrosive or severe irritants, while classification from II to IV is based on decreasing irritation severity, as well as the time required for irritation to clear. Irritation that clears in 8 to 21 days is classified as Category II, while irritation that clears within seven days is classified as Category III. For Category IV substances, irritation clears within 24 hours. The U.S. Federal Hazardous Substances Act (FHSA) guideline for ocular irritation classification (CPSC 1995) categorizes a test substance as corrosive, irritant, or nonirritant. The definition of a corrosive, according to the FHSA, is a substance that causes visible destruction or irreversible alterations in the tissue at the site of contact (CPSC 2004). FHSA classification depends on the incidence of test animals exhibiting a positive ocular response within 72 hours after application of the test substance in the conjunctival sac. Hazard classification of ocular irritants in the European Union (EU) corresponds to two risk phrases: 1) R36 denotes “Irritating to eyes”; 2) R41 denotes “Risk of serious damage to the eyes” (EU 2001). These risk phrases are based on whether the levels of damage, averaged across the 24-, 48- and 72-hours observation times for each ocular lesion, fall within or above certain ranges of scores. For the purpose of harmonizing the classification of ocular irritants internationally, the GHS (UN 2003) includes two harmonized categories, one for irreversible effects on the eye/serious damage to the eye (Category 1), and one for reversible effects on the eye (Category 2). Reversible effects are further subclassified, based on the duration of

Table 1-2 *In Vivo* Ocular Irritancy Classification Systems

Regulatory Agency (Authorizing Act)	Number of Animals	Minimum Observation Times (after treatment)	Mean Score Taken?	Positive Response	Irritant/Nonirritant Classification
EPA (FIFRA; TSCA; and The Federal Environmental Pesticide Control Act)	At least 3	1 hour, 1, 2, 3, 7, 14, and 21 days	No	- Maximum score in an animal used for classification - Opacity or Iritis ≥ 1 or Redness or Chemosis ≥ 2	One or more positive animals needed for classification in categories below. <u>Category</u>: I = Corrosive, corneal involvement, or irritation persisting more than 21 days II= Corneal involvement or irritation clearing in 8 to 21 days III = Corneal involvement or irritation clearing in 7 days or less IV = Minimal effects clearing in less than 24 hours
European Union	**Current Directive:** 1 if severe effects are suspected or 3 if no severe effects are suspected **Prior Directive:** 3 or 6 animals used to assign risk phrases	1, 2, 3 days (observation until Day 21)	Yes	(1) <u>6 animals</u> Mean study values (scores averaged over all animals in study over Days 1, 2, and 3) of: Opacity or Chemosis ≥ 2, Redness ≥ 2.5, or Iritis ≥ 1 **OR** (2) <u>3 animals</u> Individual animal mean values (scores for each endpoint are averaged for each animal over Days 1, 2, and 3) of: Opacity or Chemosis ≥ 2, Redness ≥ 2.5, or Iritis ≥ 1	**R36 Classification** (1) Mean study value (when more than 3 animals are tested) where: 2 ≤ Opacity < 3 or 1 ≤ Iritis < 1.5 or Redness ≥ 2.5 or Chemosis ≥ 2 (2) If 2 of 3 tested animals have individual animal mean values that falls into one of the following categories: 2 ≤ Opacity < 3 1 ≤ Iritis < 2 Redness ≥ 2.5 Chemosis ≥ 2 **R41 Classification** (1) Mean study value (when more than three animals are tested) where: Opacity ≥ 3 or Iritis > 1.5 (2) If 2 of 3 tested animals have individual animal mean values that fall into one of the following categories: Opacity ≥ 3 or Iritis = 2 (3) At least one animal where ocular lesions are still present at the end of the observation period, typically Day 21
GHS-Irreversible Eye Effects	3	1, 2, 3 days (observation	Yes	Mean animal values (over Days 1, 2, and 3) of:	- At least 2 positive response animals = Eye Irritant Category 1 - At least 1 animal where Opacity, Chemosis, Redness, or Iritis

Regulatory Agency (Authorizing Act)	Number of Animals	Minimum Observation Times (after treatment)	Mean Score Taken?	Positive Response	Irritant/Nonirritant Classification
		until Day 21)		Opacity ≥ 3 and/or Iritis ≥ 1.5	> 0 on Day 21 = Eye Irritant Category 1
GHS-Reversible Eye Effects	3	1, 2, 3 days (observation until Day 21)	Yes	Mean animal values (over Days 1, 2, and 3) of: Opacity or Iritis ≥ 1 or Redness or Chemosis ≥ 2 and the effect fully reverses in 7 or 21 days	- At least 2 positive response animals and the effect fully reverses in 21 days = Eye Irritant Category 2A - At least 2 positive response animals and effect fully reverses in 7 days = Eye Irritant Category 2B
CPSC (FHSA [provided under the authority of the Consumer Products Safety Act]), FDA (Food, Drug, and Cosmetics Act), and OSHA (Occupational Safety and Health Act)	6 (12, 18 possible)	1, 2, 3 days	No	Opacity or Iritis ≥ 1 or Redness or Chemosis ≥ 2 for any animal on any day	1 or more animals with destruction or irreversible alterations in the tissue at the site of contact = Corrosive 1st Tier: 4 or more positive animals = Irritant 2-3 positive animals = Go to 2nd Tier 1 positive animal = Negative 2nd Tier 3 or more positive animals = Irritant 1-2 positive animals = Go to 3rd Tier 3rd Tier 1 positive animal = Irritant

Abbreviations: CPSC = U.S. Consumer Products Safety Commission; EPA = U.S. Environmental Protection Agency; FDA = U.S. Food and Drug Administration; FIFRA = Federal Insecticide, Fungicide, and Rodenticide Act; GHS = United Nations Globally Harmonized System; OSHA = Occupational Safety and Health Administration; TSCA = Toxic Substances Control Act

persistence as Category 2A ("irritating to eyes") (reverses within 21 days) and Category 2B ("mildly irritating to eyes") (reverses within seven days). The GHS (UN 2003) categories are based on severity of the lesions and/or the duration of persistence. The GHS, the US, and the EU *in vivo* ocular irritancy classification systems are described in greater detail in **Section 4.1.3**.

Concerns about animal welfare, the cost and time to conduct ocular irritation assessments, the reproducibility of the currently used *in vivo* rabbit eye test, as well as scientific interest in understanding eye injury at the tissue and cellular level have led researchers to develop and evaluate alternative *in vitro* test methods. Recently, the EPA requested the evaluation of four *in vitro* test methods -- Isolated Chicken Eye (ICE), Isolated Rabbit Eye (IRE), Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) and Bovine Corneal Opacity and Permeability (BCOP) -- for their ability to identify ocular corrosives and severe irritants. As part of this evaluation process, a Background Review Document (BRD) has been prepared for each test method that describes the current validation status of the *in vitro* test method, including what is known about its reliability and accuracy, its applicability domain, the numbers and types of substances tested, and the availability of a standardized protocol.

This BRD evaluates the ability of the IRE test method to identify ocular corrosives and severe irritants. The IRE test method was developed by Burton et al. (1981) and proposed as a preliminary *in vitro* screen for the assessment of severe eye irritants. The principal advantage of the IRE test is to eliminate the use of live animals for ocular irritancy testing and eliminate the pain and suffering potentially associated with the *in vivo* Draize rabbit eye test. Another advantage of the IRE test method is that it typically uses eyes isolated from rabbits previously euthanized in other laboratories for research purposes in which the eyes were not specifically used or are culled from animals sacrificed commercially as a food source. In the IRE, liquid or solid substances are placed directly on the corneal surface of isolated rabbit eyes, which are held and maintained in a temperature-controlled chamber. After a 10-second exposure, followed by rinsing, the corneal thickness score, corneal swelling measurement, fluorescein penetration score, and effects on the corneal epithelium are taken over a four-hour observation period. Substances that exceed a defined cut-off score for any single parameter are considered severe irritants.

For current regulatory applications, the IRE test method could potentially be used to identify the irreversible, corrosive, and severe irritation potential of products, product components, individual chemicals, or substances in a tiered testing strategy (e.g., GHS [UN, 2003]). In the GHS stepwise approach, substances that are predicted by IRE as ocular corrosives/severe irritants could be classified as Category 1 eye irritants without the need for animal testing. Substances that are negative in the IRE test method for severe/irreversible effects would then undergo additional testing to confirm that they are not false negatives, and to determine the type, if any, of reversible effects that may occur. The test method also may be useful in a battery of *in vitro* eye irritation methods that collectively predicts the eye irritation potential of a substance *in vivo*. However, the predictivity of a battery approach will first require the assessment of the performance of each individual component.

The IRE test method is currently being used or contracted by some U.S. and European companies (e.g., pharmaceutical, cosmetic, and personal care product companies) as an in-house assay to assess the ocular irritation and corrosion potential of a wide range of substances for which there could be accidental exposures in the workplace or home. Substances are tested either individually, as mixtures, or in product formulations. The IRE test method is currently being used as an assay to evaluate industrial chemicals or products and formulations such as those used in cosmetics or in shampoos and conditioners (Balls et al. 1995; Chamberlain et al. 1997; Jones et al. 2001; Guerriero et al. 2004).

Although the IRE test method is not yet validated, the EU national regulatory authorities accept positive outcomes from this test method for eye irritation for classifying and labeling severe eye irritants (R41). Where a negative result is obtained, an *in vivo* test is subsequently required, as IRE has not been shown to adequately discriminate between eye irritants and nonirritants (European Communities 2004).

1.1.2 Peer Reviews of the IRE Test Method

Studies have been conducted in recent years to assess the validity of the IRE test method as a complete replacement for the *in vivo* ocular irritation and corrosion test method (e.g., Balls et al. 1995; Gettings et al. 1996). Previous validation efforts may have failed because: 1) they attempted to support the utility of an *in vitro* alternative as a full replacement for the *in vivo* rabbit test, rather than as a component in a tiered testing strategy; and/or 2) data generated with the *in vitro* test method(s) have typically been compared to *in vivo* maximum average scores (MAS).

However, there have been no formal evaluations of the ability of the IRE test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003) and the EPA (EPA 1996). This BRD was prepared for use by an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) expert panel review of IRE as a method to identify ocular corrosives and severe irritants. Parallel reviews of the ICE, HET-CAM, and BCOP test methods were conducted. Results of the Expert Panel Report, combined with the analyses presented in the BRDs, were used to support ICCVAM recommendations on the proposed standardized test method protocols, proposed list of recommended reference substances, and additional optimization and/or validation studies that may be necessary to further develop and characterize the usefulness and limitations of these methods.

1.2 Scientific Basis for the IRE Test Method

1.2.1 Purpose and Mechanistic Basis of the IRE Test Method

The IRE test method is an *in vitro* organotypic assay used to assess the ocular irritancy potential of substances that may be exposed topically to the eye (Burton et al. 1981; Köeter and Prinsen 1995; Chamberlain et al. 1997; Cooper et al. 2001; Jones et al. 2001). This test method was originally developed as a refinement of the *in vivo* rabbit eye test to avoid application of severe ocular irritants or corrosives to rabbits' eyes and to obviate associated pain and suffering (Burton et al. 1981). The IRE test method was later modified to identify less severe ocular irritants (Cooper et al. 2001; Jones et al. 2001).

In the IRE test method, the isolated eye is physiologically maintained and the corneal effects of an applied test substance are assessed over time (usually up to 4 hours) using multiple endpoints. These endpoints are corneal opacity (area and intensity), corneal swelling, fluorescein penetration (area and intensity), and an assessment of epithelial integrity (e.g., ulceration, mottling, stippling, etc.). If such data exist, corneal opacity in the IRE test method can be compared to that produced in the *in vivo* rabbit eye test. However, not all of the endpoints evaluated in the IRE test method are conducted in the *in vivo* rabbit eye test. For example, corneal thickness is measured quantitatively in the IRE using pachymetry, which may then be used for calculating corneal swelling (i.e., percent increase in corneal thickness over time). Because it is not required in regulatory testing (e.g., according to OECD TG 405 [OECD 2002]), corneal thickness is not typically measured in the *in vivo* rabbit eye test. Although qualitative, fluorescein staining is also measured in the IRE to estimate the degree of corneal penetration. Again, this endpoint is not typically measured in the rabbit eye test.

Histopathology or confocal microscopy would allow for a more accurate assessment of the extent of corneal injury. Maurer et al. (2002) proposed that the extent of ocular injury, as measured by confocal microscopy, has the greatest impact on the outcome of such an injury. Live/dead cell staining methods evaluated with confocal microscopy have also been used to determine the extent or depth of corneal injury *in vivo* (Maurer et al. 1997) and in an *ex vivo* corneal button assay (Jester et al. 2001). These studies prompted the authors to suggest that the extent of corneal injury could be used as the basis for developing alternative methods to predict the level of damage produced by ocular irritants.

1.2.2. Similarities and Differences of Modes and Mechanisms of Action Between the IRE Test Method and Ocular Irritancy in Humans and/or Rabbits

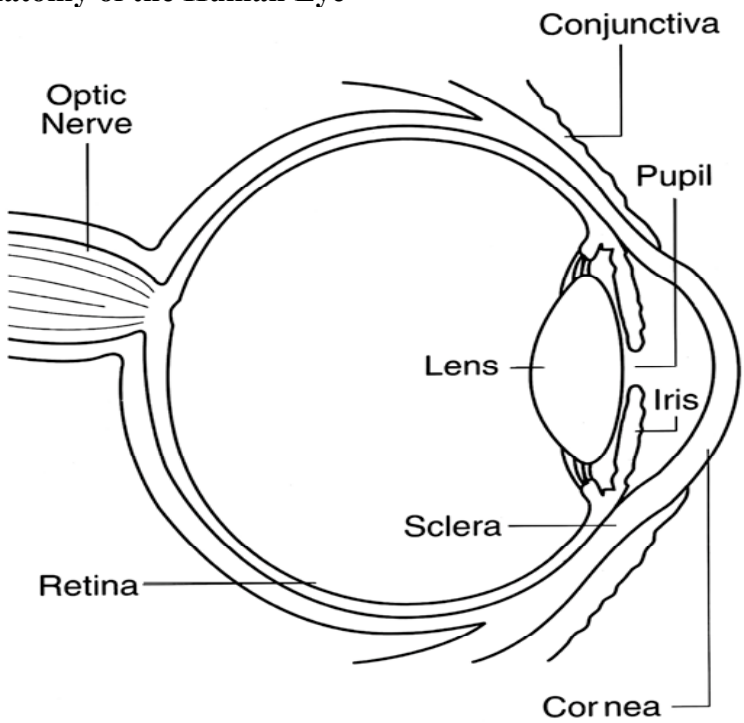
1.2.2.1 *The Mammalian Eye: Common Anatomy of the Human and Rabbit Eye*

The eyeball is a fibrovascular globe, which is surrounded by a bony orbit that is impenetrable to light (Bruner 1992). The anterior portion of the eyeball is the only portion that is exposed to the environment, while the remainder of the eye is protected by the eyelids and the bony orbit. The eyeball is composed of three concentric tunics (the fibrous tunic, the vascular tunic, and the neuroectodermal tunic) that can be further subdivided. The fibrous tunic is the outermost layer of the eye comprised of the transparent cornea and the opaque sclera. The middle vascular tunic is comprised of the choroids, the ciliary body, and the iris (which can be referred to as the uvea). The neuroectodermal tunic is the innermost layer and is comprised of the retina, which contains photoreceptors and is connected to the central nervous system (Wilkie and Wyman 1991; Bruner 1992).

The fibrous tunic provides the primary framework for the eye. The cornea is the transparent surface of the eye, and is comprised of three major layers: the epithelium, the stroma, and the endothelium (**Figure 1-1**). The human cornea is a hydrated, nonvascularized structure. Corneal stroma contains 78% water and hydration is a requisite for the capacity of the stroma to swell in response to an irritant (Duane 1949). The cornea is nutritionally maintained in a homeostatic state by the aqueous humor, tear film, and the surrounding vascularized tissues.

Proper function of squamous or cuboidal cells in the endothelial layer is required to remove water from the cornea.

Figure 1-1 Anatomy of the Human Eye



Figured obtained at http://www.nei.nih.gov/photo/eyean/index.asp

The cornea is the major refracting element in the optical path, which flows from the light source through the cornea (70% of refractive power) to the lens (30% of refractive power) and into the retina (Duane 1949; Mishima and Hedbys 1968a). Therefore, corneal transparency is an important factor in optimal eye functioning. For maximum refractive power, the anterior surface of the cornea, composed of layers of translucent epithelial cells, is maintained in a smooth configuration by the tear film. The corneal stroma, composed of translucent keratocytes interspersed with collagen fibrils, requires uniformity and proper spacing of the collagen fibrils to maintain an appropriate corneal refractive index with minimal light scattering (Maurice 1957). This combination of structure and cellular morphology serves to maintain corneal transparency.

The eye is critically dependent on the highly vascularized middle coat (uvea) for regulation of blood and ocular permeability barriers, maintenance of intraocular pressure in the aqueous humor, and drainage of ocular fluid (Unger 1992). The uveal tract is richly innervated by somatic sensory neurons, derived from the ophthalmic division of the trigeminal nerve. Importantly, alterations to any of these features (e.g., edema, cell destruction, vascularization, cell proliferation) can cause corneal opacity and concomitant loss of function (Parish 1985; Wilkie and Wyman 1991; Bruner 1992).

The sclera is comprised primarily of three layers of irregularly arranged collagen fibrils of varying diameter. The irregular arrangement of the fibrils produces the white color that is seen on eyeballs. The conjunctiva is a mucous membrane that covers the exposed scleral surface (bulbar conjunctiva) and the inner surface of the eyelids (palpebral conjunctiva). The conjunctiva contains blood vessels, nerves, conjunctival glands, and inflammatory cells. As part of the inflammatory response in the conjunctiva, dilation of the blood vessels, fluid leakage and cellular leakage occurs (Bruner 1992).

The major component of the vascular tunic is the iris. The iris sits in front of the lens and the cilliary body, which also are considered part of the vascular tunic. Contraction of the iridal muscles alters the diameter of the pupil and thus regulates the amount of light entering the eye (Bruner 1992).

1.2.2.2 *Differences Between Human and Rabbit Eyes*

There are several anatomical and physiological differences between the rabbit eye and the human eye. One difference is the presence of a nictitating membrane, or third eyelid, in the rabbit. As this membrane slides horizontally across the eye, it is proposed that it aids removing and/or excluding irritating substances from the corneal surface (Calabrese 1983). It also is proposed that the kinetic removal of a substance from a rabbit eye may occur at a rate different than in humans, due to the presence of the nictitating membrane, although this has not been documented in comparative studies (Curren and Harbell 1998). Another difference is the larger conjunctival sac in the rabbit, which allows for larger test volumes to be instilled, perhaps more than could be accounted for on accidental exposure (Curren and Harbell 1998).

The rabbit cornea is thinner than that found in humans, and rabbits tend to have less tear production (Curren et al. 1998; Cooper et al. 2001). The thicknesses of structural components of the cornea also are different between the two species. For example, Descemet's membrane is proposed to be about 5 to 10 μm in humans and 7 to 8 μm in rabbits (Calabrese 1983). Furthermore, the area of the cornea in relation to the total surface of the globe varies significantly between species; in humans the relationship is 7%, while in rabbits the relationship is 25% (Swanston 1985). Finally, young rabbits have the ability to regenerate damaged corneal endothelium, while humans do not (Chambers W, personal communication).

The relationship between species differences in eye anatomy and physiology and the sensitivity to ocular irritants has not been clearly established. It has been proposed that the larger conjunctival sac, thinner cornea, larger proportion of the cornea to the eyeball as well as other differences in the rabbit eye lead to an increased sensitivity to irritants (Calabrese 1983; Swanston 1985). However, other differences (e.g., the presence of the nictitating membrane, low blink frequency rate) indicate that the rabbit is as sensitive as humans to irritants. Comparisons of human exposure experiences to results in the *in vivo* test method indicate that in some cases the rabbit eye is more sensitive to some irritants, while in other cases the human eye is more sensitive (McDonald et al. 1987).

1.2.2.3 *The In Vivo Rabbit Eye Test Method*

The current *in vivo* rabbit eye irritation test method evaluates the cornea, the iris, and the conjunctiva for adverse effects after exposure to a potential irritant (See **Section 4.0** for a discussion of the *in vivo* scoring system for lesions at these sites). The cornea is visually observed both for the degree of corneal opacity and the area of the cornea in which opacity is involved. The iris is assessed for inflammation, iridal folds, congestion, swelling, circumcorneal injection, reaction to light, hemorrhage, and gross destruction. The conjunctiva is evaluated for the degree of redness, chemosis (swelling), and discharge (Draize et al. 1944). Draize and colleagues (1944) developed an analysis method where the severities of the effects are weighted differently, with corneal effect being weighted the most. The effects of a test substance on the cornea, conjunctiva, and iris play a role in severe ocular irritant and corrosive labeling and classification in classification systems used by some regulatory agencies (CPSC 1995; EPA 1998; EU 2001; UN 2003).

Irritation responses and the degree of the response in the cornea, iris, and conjunctiva differ due to the specific functions and anatomy of each structure. Development of slight corneal opacity can be due to loss of superficial epithelial cells and epithelial edema. Comparatively, more severe corneal opacity may be observed if an ocular irritant produces its effects deeper into the cornea. The ensuing repair process can lead to scar development on the cornea and vision impairment. Irritation responses in the iris are typically due to direct exposure to a substance, which has passed through the cornea and sclera, or due to extension of significant surface inflammation. Acute inflammation of the uvea tract is characterized by edema, vessel dilation, and the presence of exudates, while severe inflammation of the uvea tract is characterized by accumulation of blood or leukocytes in the anterior chamber. Conjunctival inflammatory responses can produce vasodilation, edema, subconjunctival hemorrhage, and lacrimal secretions (Bruner 1992).

The extent of corneal injury resulting from an ocular irritant also is dependent on the physicochemical characteristics (e.g., acids and bases with pH extremes, solvent-induced protein or DNA precipitation, surfactant-induced saponification of membranes), and chemical reactivity of the substances when in contact with individual ocular cells or structures (e.g., alkylation, hydrolysis, oxidation, reduction, hydroxylation) (Grant 1974; McCulley 1987; Berta 1992; Nourse et al. 1995; Fox and Boyes 2001). Direct or indirect ocular injury may result from the impact of these physicochemical effects on normal homeostatic cellular mechanisms and from consequent edema, inflammation, apoptosis, necrosis, and reparative processes (e.g., collagen deposition and scarring) (Unger 1992; Pfister 2005). In the normal eye, test substances may disrupt the tear film, reach the epithelium, and penetrate through Bowman's layer into the stroma, through Descemet's membrane, and into the endothelium (Pasquale and Hayes 2001). Damage to the endothelium may be irreparable.

The tear film consists of an inner layer of mucous, a middle layer of water, and an outer film of oil. The tear film contains lactoferrin, peroxidase, lysozyme, immunoglobulins and complement factors to eliminate potentially offensive material (Unger 1992). In conjunction with the neurogenically controlled blink reflex and tear producing cells, the tear film serves as a protective barrier against an ocular irritant for the corneal epithelium. The

physicochemical properties (e.g., hydrophilicity, hydrophobicity, hypertonicity, hypotonicity, oxididation, reduction) in addition to the chemical and biochemical properties of an applied test substance impact its ability to breach the tear film, or interact with its components and impact the corneal epithelium. The tear film and the aqueous humor also provide nourishment (e.g., glucose and oxygen) to the nonvascularized cornea. The extent of damage to the tear film by an applied substance therefore impacts the ability of the tear film to nourish dependent corneal tissue. Changes in the distribution, physical structure, or secretion rate of the tear film by an applied test substance might have significant nutritional, refractory, chemical and physical impacts on corneal tissue (Mishima and Hedbys 1968a; Mishima and Hedbys 1968b).

Either direct (e.g., caustic or corrosive) or indirect (e.g., inflammatory mediator release) effects of chemicals in contact with the anterior corneal surface may result in perturbation of the optical elements needed to maintain the appropriate index of refraction in the cornea (e.g., uniformity and proper spacing of collagen fibrils), resulting in significant light scattering and impairment of vision (McCulley 1987; Berta 1992; Nourse et al. 1995; Wilson et al. 2001). Corneal injury may result in opacification, swelling, damage extending from the epithelium into the stroma or possibly through the endothelium, and changes in corneal morphology (e.g., ulceration, scarring, pitting, mottling).

Opacification of the cornea may result from: 1) direct or indirect damage to the epithelial cells with or without penetration into the stroma; 2) protein denaturation of the epithelial cells such as that produced by alcohols, alkalis, or organic solvents; 3) alkylation of protein or DNA; 4) membrane saponification by surfactants; 5) inflammatory cell infiltration; 6) collagen deposition; 7) swelling of corneal epithelial cells or corneal stroma; 8) displacement or rearrangement of collagen fibrils; or 9) degradation of the extracellular matrix (Grant 1974; McCulley 1987; Thoft 1979; York et al. 1982; Fox and Boyes 2001; Kuckelkorn et al. 2002; Eskes et al. 2005; Pfister 2005).

Corneal swelling results from disruption of the anterior barrier membrane formed by the epithelial cell layer and Bowman's layer. This results in disruption of stromal collagen fibril uniformity, loss of proteoglycans, and cell death, which leads to bullae formation, stromal cloudiness, and increased hydrostatic pressure (which may extend posteriorly throughout the corneal stroma, penetrating into Descemet's layer and into the endothelium) (Mishima and Hedbys 1968a; Mishima and Hedbys 1968b). Osmotic changes induced by these effects may further damage keratocytes and the collagen matrix.

Corneal damage also may be characterized by morphological changes (e.g., described as stippling, ulceration, mottling, pannus, neovascularization).

Corneal injury also is dependent on the type and concentration of applied chemical. Alkalis penetrate more readily than acids do, and the depth of penetration is dependent on alkali concentration (McCulley 1987). With alkali injury, the hydroxyl ion saponifies the fatty acid components of the cell membrane, disrupting cellular contents and resulting in cell death. The cation is responsible for the penetration process (Grant 1974). Acids tend to penetrate less deeply than alkalis, with the exception of hydrofluoric and sulfuric acids. The hydrogen

ion causes damage due to pH alteration, while the anion precipitates and denatures protein in the corneal epithelium and superficial stroma (Freidenwald et al. 1946). Limbal ischemia is a significant consequence of even mild alkali or acid burns (Kuckelkorn et al. 2002).

While not in the direct optical path, the Palisades of Vogt, located in the sclero-corneal limbus, are thought to house corneal stem cells and serve as a generative organ for normal replacement of dead corneal epithelial cells for re-epithelialization during repair of corneal injury. Depletion or partial loss of the limbal stem cell population may result in corneal vascularization due to loss of the barrier function of the limbus, which serves to prevent conjunctival epithelial cells from migrating to the corneal surface (Dua and Azuara-Blanco 2000).

Neutrophils are recruited in response to acid and alkali injury as well as in response to other ocular toxicants (Pfister 2005). Neutrophil migration is stimulated by the release of chemotatic factors (e.g., interleukins, growth factors, etc.) from damaged or chemically activated local resident epithelial cells or stromal keratocytes (Wilson et al. 2001). Loss of keratocytes following either chemical or mechanical epithelial injury may be mediated by apoptosis, perhaps by release of IL-1 and TNFα (Wilson et al. 2001). Resident mast cells may release biogenic amines that perturb the hydrostatic balance and permit inflammatory or edemagenic mediators into the locally inflamed area. Migrated neutrophils release additional cytokines (e.g., IL-1 and TNF-α) and enzymes such as proteases, collagenases, kinases, and phospholipaseA_2 (PLA_2). PLA_2 produces edemagenic and vasoactive mediators such as prostaglandins and leukotrienes from arachidonic acid in cellular membranes.

This cascade of events ultimately facilitates repair by stimulating fibrin deposition and granuloma formation. However, migrating inflammatory cells such as neutrophils also may be involved in the release of collagenases (e.g., matrix metalloproteinases [MMPs]), which have been implicated in corneal ulcer formation. Acetylcysteine, L-cysteine, and EDTA have been shown to reduce corneal ulceration in response to alkali injury while inhibiting MMPs (Pfister 2005). Other inflammatory cells such as macrophages and T-lymphocytes may be found up to 24 hours after injury. Once an area is damaged and devoid of keratocytes, proliferation and migration occurs as part of the wound healing process. This process may be mediated in part by numerous growth factors (Wilson et al. 2001).

Although variable responses occur among species, neuropeptides (e.g., Calcitonin Gene Related Peptide [CGRP] and substance P) have profound effects on the anterior portion of the highly innervated eye, particularly in lower mammals such as the rabbit (Unger 1992). CGRP appears to affect vascular smooth muscle (Oksala and Stjernschantz 1988), whereas substance P may be involved in meiosis (Unger 1990). Loss of functional sympathetic innervation reduces or eliminates presynaptic catecholamine reuptake sites resulting in denervation supersensitivity. This also may result in enhanced sensitivity to noxious stimuli.

Applied test substances also can adversely affect homeostasis within the cornea. As oxygen is absorbed into the cornea from the atmosphere, interference with oxygen uptake may lead to corneal swelling (Mishima and Hedbys 1968a). The cellular respiratory needs of the endothelium and epithelium are similar, both requiring carbohydrate metabolism. Glucose

metabolism in the cornea occurs by glycolysis and oxidation through the tricarboxylic acid cycle as well as through the hexose-monophosphate shunt (Kinoshita 1962). Glucose within the cornea is used to supply glycogen, which is stored in the epithelium. Applied substances that modulate any of these processes may be associated with ocular toxicity.

1.2.2.4 *Comparison of the IRE Test Method with the In Vivo Rabbit Eye Test Method*

Both the IRE test method and the Draize rabbit eye irritation test focus primarily on irritation or corrosion to the corneal surface. The degree of penetration of damage to the corneal surface by an irritant or corrosive into the epithelium, stroma or endothelium can be assessed in the IRE, but is difficult to assess in the *in vivo* test. In the isolated eye, fluorescein retention and histology can also be assessed to confirm the degree of penetration within the corneal tissue. In addition, the morphological abnormalities of the corneal surface (i.e., stippling, mottling, pitting, etc.) and the area of corneal involvement can be evaluated in the IRE as it can in the *in vivo* test method. The primary disadvantage of the IRE test method in comparison to the *in vivo* test method is that the iridal tissue *in vitro* is devoid of blood flow and nerves controlling the musculature and blood vessels may be compromised. Thus, iridal damage due to inflammatory or neuromuscular components cannot be evaluated. Similarly, the conjunctival tissue is not present in the isolated eye. However, in the Draize test, the corneal score is weighted to represent 73% of the total score. The iris is weighted at 9% of the total ocular score and the conjunctiva is weighted at 18%. The combined contribution of the iridal and conjunctival tissue to the *in vivo* score is 27%. Thus, for the identification of a severe irritant or corrosive, damage to the cornea would be expected to have the greatest impact on the Draize score. Furthermore, other *in vitro* test methods can be used in a tiered-testing approach to assess vascular changes produced by an irritant or corrosive (e.g., HET-CAM) which would be more likely related to conjunctival damage. The isolated rabbit eye cannot be used to assess inflammatory changes induced by an irritant or corrosive, since the vascular components are not functional. Substances that might produce irritancy through a delayed inflammatory response may not be picked up in the IRE test method, since the test is only run for four hours.

In the isolated rabbit eye, neurogenic components that drive tear film production are not functional. Although the cornea is constantly hydrated with a saline drip in the IRE test method, the lack of a tear film is considered a limitation. In fact, the saline drip eventually removes the residual tear film, which cannot subsequently be regenerated. When compared with an *in vivo* rabbit eye study, application of a test substance in the absence of this protective barrier might be expected to cause an increase in false positive outcomes. One of the conclusions from a workshop on mechanisms of eye irritation highlighted the need for additional research on the impact of chemicals on tear film and the consequences of tear film disruption (Bruner et al. 1998).

Corneal opacification in both the *in vivo* rabbit eye test and the IRE test method is visually observed or may be assessed using a slit-lamp.

In the IRE test, corneal swelling is assessed quantitatively, using an ultrasonic or optical pachymeter to measure the increase in corneal thickness during an experiment. It is

expressed as a percent increase in corneal thickness over time relative to the pre-treatment measurement.

1.2.3 Intended Range of Substances Amenable to the IRE Test Method and/or Limits of the IRE Test Method

Studies indicate that the IRE test method is amenable to use with a broad range of solid and liquid substances. However, the IRE appears to be less predictive with solids than it is for liquid substances (Whittle et al. 1992). Substances amenable to testing include, but are not limited to, inorganic, aliphatic, aromatic, and heterocyclic chemicals, surfactants, polymers, and mixtures/formulations.

The major limitation of the IRE test method is the inability to evaluate the effect of test substances on the iris and conjunctival tissue. For the evaluation of severe irritants, this is not a major limitation of the test method, since most severe irritants damage the corneal tissue.

Another potential limitation of the test method is that it can be used only for short-term assessments of the irritancy of a test substance. The currently accepted *in vivo* test method usually observes the rabbits for up to 21 days after treatment to assess reversibility of any of the observed endpoints and to evaluate test substances that produce eye effects over an extended time period. Comparatively, the observation period for evaluating effects in the IRE test method post-treatment is up to four hours. Therefore, potential reversibility of the affected endpoint beyond four hours, or an effect with a delayed onset (e.g., slow-acting irritants) due to inflammation or other factors cannot be adequately evaluated with this test method.

1.3 Regulatory Rationale and Applicability

1.3.1 Current Regulatory Testing Requirements and ICCVAM Prioritization Criteria

The following section reviews and summarizes the extent to which the five ICCVAM prioritization criteria apply to the IRE test method (ICCVAM 2003).

Criteria 1. The extent to which the proposed test method is (a) applicable to regulatory testing needs and (b) applicable to multiple agencies/programs.

The IRE assay has been proposed as a method to identify ocular corrosives or severe irritants, as is required by several U.S. laws. **Table 1-1** identifies the U.S. agencies and programs, which classify and label substances for eye irritation and corrosion. These agencies are the FDA, the EPA, Department of Agriculture, Department of Labor, the Consumer Product Safety Commission (CPSC), and the Chemical Safety and Hazard Investigation Board. Therefore, the IRE test method is applicable to the regulatory testing needs of multiple U.S. Federal agencies and programs.

Criteria 2. Warranted, based on the extent of expected use or application and impact on human, animal, or ecological health.

Current regulatory testing needs require the *in vivo* assessment of the eye irritancy or corrosivity hazard associated with the use of chemicals/products for labeling purposes.

These testing needs require the use of laboratory rabbits. Alternative *in vitro* eye irritation and corrosion test methods could be applied to these testing needs.

Criteria 3. The potential for the proposed test method, compared to current test methods accepted by regulatory agencies, to (a) refine animal use (decrease or eliminate pain and distress), (b) reduce animal use, or (c) replace animal use.[2]

The IRE test method has the potential to refine or reduce animal use in eye irritation testing. Substances that are identified as ocular corrosives or severe irritants would be excluded from testing *in vivo*, which would reduce the number of rabbits used for ocular testing and also spare animals the pain and distress of exposure to severe eye irritants.

Criteria 4. The potential for the proposed test method to provide improved prediction of adverse health or environmental effects, compared to current test methods accepted by regulatory agencies.

Based on its long history of use and acceptance by U.S. Federal and international regulatory agencies, the current system of ocular hazard assessment, which is based on the rabbit eye test (i.e., CPSC 1995; EPA 1998; OECD 2002), appears to have adequately protected public health. However, use of the rabbit eye test to predict the ocular irritation potential of substances for humans is not without controversy (e.g., intra- and inter-laboratory variability, qualitative evaluation of ocular lesions). The accuracy of the currently used *in vivo* rabbit eye test for predicting severe eye irritants in humans and the limitations of the method for predicting the irritancy of specific chemical and/or product classes are not known due to the lack of comparative data. Therefore, the potential of the proposed test method to provide improved prediction of adverse human health effects is unknown.

Criteria 5. The extent to which the test method provides other advantages (e.g., reduced cost and time to perform) compared to current methods.

The IRE test method could reduce the time needed to assess a substance, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days and can be extended up to 21 days. The irritancy of a test substance in the IRE assay is measured over a four-hour time period. The complete assay with dissection and equilibration can be completed (after eye procurement) in eight hours. Regarding comparative costs (based on conducting Good Laboratory Practice [GLP] compliant studies), the standard IRE assay conducted with concurrent positive and negative controls costs $1074 per test substance at SafePharm Laboratories (2004). The current cost of a GLP compliant EPA OPPTS Series 870 Acute Eye Irritation (EPA 1998) or OECD Test Guideline 405 (OECD 2002) test at MB Research Laboratories (Spinnerstown, PA) ranges from $765 for a 3 day/3 animal study up to $1665 for a 21 day/3 animal study (MB Research Laboratories, personal communication). While the cost of an IRE assay includes concurrent positive controls, the *in vivo* rabbit test method

---

[2] Refinement alternative is defined as a new or revised test method that refines procedures to lessen or eliminate pain or distress to animals, or enhances animal well-being; Reduction alternative is defined as a new or revised test method that reduces the number of animals required; Replacement alternative is defined as a new or revised test method that replaces animals with non-animal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate) (ICCVAM 1997).

does not include equivalent controls. The cost of the IRE assay is reduced significantly by the reduction in time involved in running the assay relative to the *in vivo* test and by the use of donated rather than purchased animals.

1.3.2 Intended Uses of the Proposed IRE Test Method

In vitro ocular irritation testing methods (e.g., ICE, IRE, BCOP, and HET-CAM) have been proposed for identification of ocular corrosives and severe irritants (e.g., Ocular Irritant Class I per the EPA classification system [EPA 1996], Ocular Irritant Class R41 per the EU classification system [EU 2001], or Ocular Irritant Class 1 per the GHS classification system [UN 2003]).

1.3.3 Similarities and Differences in the Endpoints Measured in the Proposed Test Method and *In Vivo* References Test Method

As mentioned in **Section 1.1.1**, the *in vivo* rabbit eye test method in current use by U.S. Federal and international agencies is based on a method developed by Draize and colleagues in 1944. This test method involves instillation of the test substance into the lower conjunctival sac of the rabbit eye, and evaluates the cornea, the iris, and the conjunctiva for adverse effects after exposure to the potential irritant. The cornea is evaluated both for the degree of corneal opacity and the area of the cornea in which opacity is involved. The iris is assessed for inflammation, iridal folds, congestion, swelling, circumcorneal injection, reaction to light, hemorrhage, and gross destruction. The conjunctiva is evaluated for the degree of redness, chemosis (swelling), and discharge (Draize et al. 1944).

As detailed in **Section 1.2.1**, the IRE is an organotypic model that uses an isolated rabbit eye for ocular irritation testing. This *in vitro* assay is designed to identify the effect of severe irritants/corrosives on the corneal tissue. Corneal opacity, corneal swelling, penetration of fluorescein into the corneal tissue and assessment of the morphology of the corneal epithelium are measured or scored. In the *in vivo* Draize rabbit eye test, scores are obtained for corneal opacification and the area affected in response to an irritant, irritation of the iris and redness, swelling and discharge in the conjunctival tissue. The scores are weighted with the effects on the cornea receiving the highest weight (73% of total). The IRE test method is designed to measure corneal effects only, since the iris and conjunctival tissue are dissected out in the preparation of the eye. However, given the increased weight of the *in vivo* Draize score for corneal effects over iridal or conjunctival irritation and the general observation that most severe irritants have an effect on corneal opacity (Parish 1985; Sina and Gautheron 1994), the endpoints measured in the IRE test method are comparable to those used for evaluation of corneal damage in the *in vivo* method.

1.3.4 Use of Proposed Test Method in Overall Strategy of Hazard or Safety Assessment

The IRE test method is being considered for use in the identification of ocular corrosives and severe irritants in a tiered testing strategy (GHS [UN 2003]). The GHS proposes a tiered testing and evaluation strategy for serious eye damage and eye irritation using available data from dermal irritation studies, knowledge of structure activity relationships, and pH screening. As shown in **Figure 1-2**, the GHS also allows for use of validated and accepted *in vitro* methods to identify severe ocular irritants/corrosives without further testing. If a test substance is classified in a validated *in vitro* method as an ocular corrosive or severe irritant,

then no further testing would be required and the test substance would be appropriately labeled. If a test substance is not classified as an ocular corrosive or severe irritant using a validated *in vitro* method (i.e., the test substance remains unclassified), then current regulatory agency regulations for ocular testing would be followed. It is noted that the current testing strategy is proposed for use for regulatory classification and labeling purposes.

1.4 Validation of *In Vitro* IRE Test Methods

The ICCVAM Authorization Act (Sec. 4(c)) mandates that "[e]ach Federal Agency … shall ensure that any new or revised … test method … is determined to be valid for its proposed use prior to requiring, recommending, or encouraging [its use]." (Public Law 106-545).

Validation is the process by which the reliability and relevance of an assay for a specific purpose are established (ICCVAM 1997). Relevance is defined as the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 1997). For the IRE test method described in this BRD, relevance is restricted to how well the assay identifies substances that are capable of producing corrosive or severe irritant effects to the eye. Reliability is defined as the reproducibility of a test method within and among laboratories and should be based on performance with a diverse set of substances that are representative of the types of chemical and product classes that are expected to be tested and the ranges of response that need to be identified. The validation process will provide data and information that will allow U.S. Federal agencies to develop guidance on the development and use of the IRE test method as part of a tiered-testing approach to evaluating the eye irritation potential of substances.

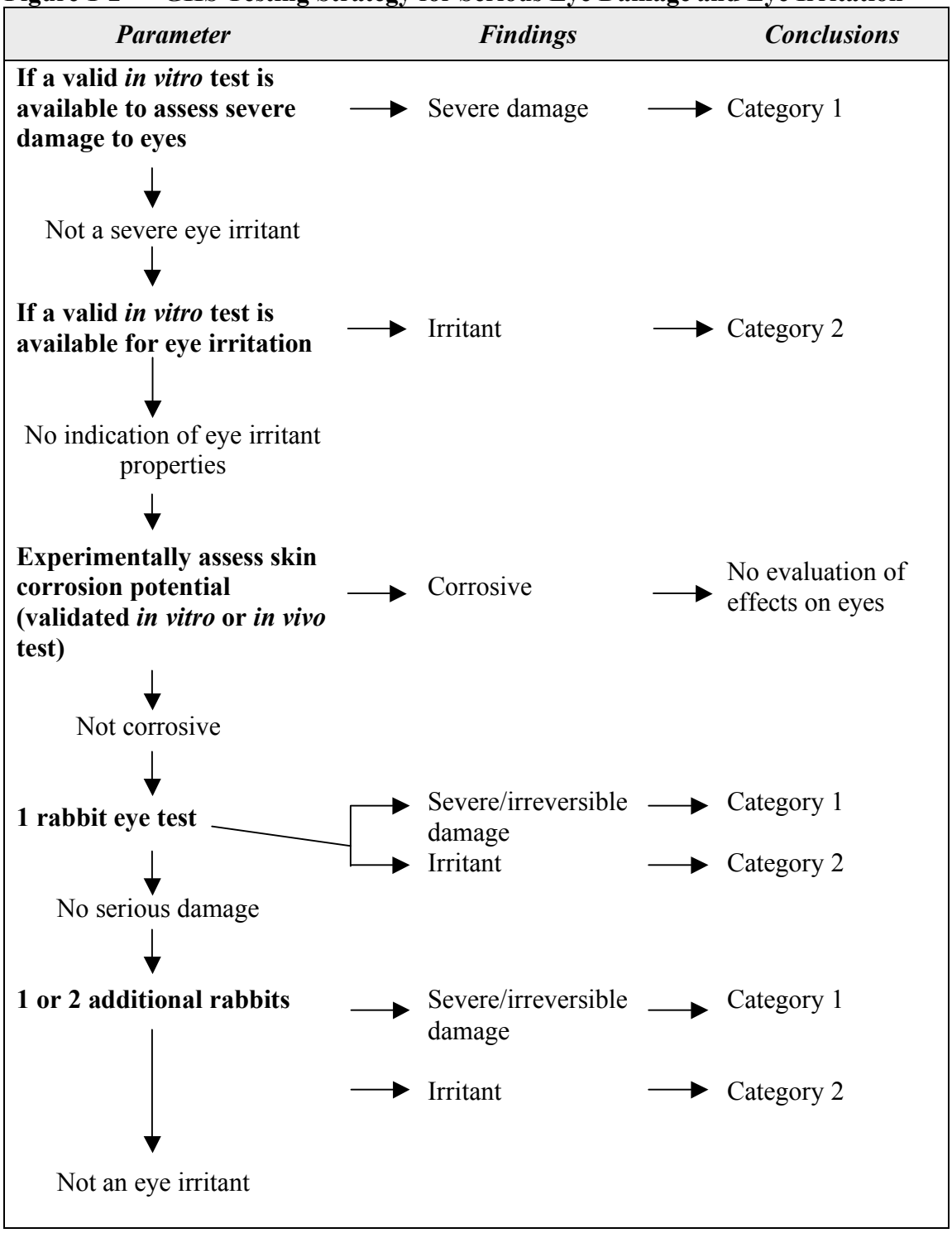
The first stage in this evaluation is the preparation of a BRD that presents and evaluates the relevant data and information about the assay, including its mechanistic basis, proposed uses, reliability, and performance characteristics (ICCVAM 1997). This BRD summarizes the available information on the various versions of the IRE test method that have been published. Where adequate data are available, the qualitative and quantitative performances of the assays are evaluated and the reliability of each version of the test method is compared with the reliability of the other versions. If there are insufficient data to support the recommendation of a standardized protocol for IRE, this BRD will aid in identifying essential test method components that should be considered during its development and validation.

1.5 Search Strategies and Selection of Citations for the IRE BRD

The IRE test method data summarized in this BRD are based on information found in the peer-reviewed scientific literature. An online literature search of entries in MEDLINE, TOXLINE, and Web of Science was conducted to retrieve database records on publications reporting on *in vitro* testing of substances for their ocular irritancy potential using the IRE test method. Specifically, records were sought containing the keywords "isolated" or "enucleated" and "rabbit" and "eye." Each database record included authors, bibliographic

citation, and indexing terms. Most records also included abstracts. Of the 28 records obtained from the search last updated in February 2004, seven contained results from an IRE

Figure 1-2 GHS Testing Strategy for Serious Eye Damage and Eye Irritation



Adapted from UN (2003).

test method, four were review articles and two were background articles related to the IRE test method. Additional information including an abstract on the IRE test method was obtained at the Society of Toxicology Meeting in March, 2004. Abstracts of selected titles were reviewed, and the relevant articles were selected and retrieved from the literature for analysis. A database of the literature citations was established using bibliographic database software.

[This Page Intentionally Left Blank]

2.0 IRE TEST METHOD PROTOCOL COMPONENTS

2.1 Overview of How the IRE Test Method is Conducted

The IRE test was developed by Burton and his colleagues at Unilever Research Laboratory, Colworth, United Kingdom as an *in vitro* alternative to the *in vivo* Draize rabbit eye test method for the assessment of eye irritation (Burton et al. 1981). In the IRE test method, liquid test substances are spread using a syringe and solids are pulverized and applied as a powder over the corneas of enucleated rabbit eyes. The principal advantages of this test method are that the animals are euthanized prior to ocular irritancy testing, eyes from animals used for other toxicological purposes or from the food chain can be used, and testing is performed on the cornea, the part of the eye that is generally given the highest weight for scoring ocular irritancy in the Draize test. The effects of the test substance on the cornea of the isolated eye are measured quantitatively as an increase in thickness (swelling), subjectively as scores for corneal opacity, the area of corneal involvement, and fluorescein penetration, and descriptively as morphological changes to the corneal epithelium. However, the number of ocular parameters and the number of time points measured varies from study to study. Two additional refinements of the IRE test method may be incorporated into the protocol or used *ad hoc* to supplement existing data. One is the use of histopathology to confirm or identify the extent of irritancy at the cellular level, especially when the degree of irritancy falls between moderate and severe. Another is the use of confocal microscopy to determine the extent and depth of ocular injury (Maurer et al. 2002). Many studies using the IRE test method evaluate single or multiple ocular endpoints at various times and then assign irritancy classifications to the substances tested (CEC 1991; Köeter and Prinsen 1995; Cooper et al. 2001; Jones et al. 2001), while others use mean data from one or more ocular endpoints assessed at various times after application of the test substance, typically 0.5 to 4 hours (Balls et al. 1995; Gettings et al. 1996). One protocol for the IRE test method was designed to specifically identify severe eye irritants (Guerriero et al. 2004). In this study, cut-off values for each ocular parameter tested were predetermined. If these cut-off values were achieved or exceeded in any single parameter over a period of 0.5 to 4 hours, including a significant change in the corneal epithelium, the test substance was classified as a severe eye irritant with potential to cause serious or irreversible damage to the human eye. Protocols developed and used at SafePharm and Unilever in the United Kingdom were provided (Jones P and Whittingham A, personal communications) and information on additional IRE protocols was obtained from reports in the literature (Gettings et al. 1966; Burton et al. 1981; Price and Andrews 1985; INVITTOX 1994; Balls et al. 1995; Chamberlain et al. 1997; Cooper et al. 2001; Jones et al. 2001; Guerriero et al. 2004). These protocols are compared in **Appendix A.**

2.2 Description and Rationale for the Test Method Components

Currently, there is no widely accepted, standardized IRE test method for detecting ocular corrosives and severe irritants. Evaluation of the IRE test method for its usefulness as a partial or full replacement for the Draize rabbit eye test has been confounded by the lack of a standardized protocol. Although initially developed by Burton et al. (1981) for the

assessment of severe eye irritants using a relatively small set of eleven test substances, the IRE test method has been modified for use in the assessment of either selective types of irritants (e.g., severe irritants) or for specific classes of chemical substances or products (e.g., surfactant-based chemicals, cosmetic and hair care products) (Gettings et al. 1966; Chamberlain et al. 1997; Cooper et al. 2001; Jones et al. 2001). In other studies, protocols were geared to evaluate a wider range of chemical classes over the entire range of irritancy for test method assessment or validation purposes (Price and Andrews 1985; Köeter and Prinsen 1985; CEC 1991; Balls et al. 1995; Gettings et al. 1996) or for interlaboratory trials (Whittle et al. 1992). Guerriero et al. (2004) modified the original IRE test method protocol to refine assessment of pharmaceutical worker safety by using decision criteria (Prediction Model) designed to identify severe eye irritants using a chemical database of 30 pharmaceutical ingredients, chemical intermediates, and raw materials and an additional 14 reference chemicals from ECETOC (1998).

The following sections describe in detail the essential components of the IRE test method for the identification of ocular corrosives or severe eye irritants. For each section, a summary is provided of the information obtained from reviewed reports and personal communications with expert scientists knowledgeable about the assay. Many of the components of these protocols have been included based on historical use, and the rationale for their selection is not known. For each test method component, a summary is presented of information obtained from:

- SafePharm Laboratories, a toxicology laboratory that has performed the enucleated rabbit eye assay in a GLP-compliant testing facility since 1999.
- INVITTOX Protocol No. IP-85 (1994). This protocol was used by the lead laboratory for the Balls et al. (1995) IRE validation study.
- A literature search and review of publicly available IRE protocols; which are based on the methodology first reported by Burton et al. (1981). These protocols are summarized in **Appendix A**.
- Discussion and personal communication with Ms. Penny Jones (Unilever) and Mr. Robert Guest (SafePharm Laboratories), scientific experts currently using the isolated rabbit eye test method, with additional information provided by Dr. Andrew Whittingham (SafePharm Laboratories) and Mr. Frederick Guerriero (GlaxoSmithKline).

2.2.1 Materials, Equipment and Supplies Needed

2.2.1.1 *Source of Rabbit Eyes*

Typically, healthy New Zealand white rabbits weighing 2.5 to 4.0 Kg are used. Rabbits are usually purchased from laboratory animal suppliers. However, since a principal purpose of the IRE test method is to reduce animal use, eyes have been obtained from laboratory rabbits used for other purposes, such as skin testing, in which the eyes are not affected (e.g., mild or nonirritant substances or control eyes). However, where regulatory agencies do not permit animal reuse, it is possible to obtain eyes from an abattoir (e.g., PelFreeze, Rogers, AZ) where rabbits are routinely killed for food. Local abattoirs are available throughout the U.S. and Europe. There are 200,000 rabbit producers throughout the U.S. with turnover of 6 to 8 million rabbits per year [http://agalternatives.aers.psu.edu/other/rabbit/rabbit.pdf]. Eyes are typically shipped from a local laboratory or abattoir in a humidified container wetted with

saline or an appropriate buffer solution at room temperature for use within an hour or on ice for longer periods up to 24 hours. Eyes have been shipped overnight (i.e., PelFreeze, Rogers AR) on ice under conditions that do not have adverse effects on corneal transparency or physiological function when the abattoir is instructed on how to remove and package the eyes properly (Edelhauser H, personal communication). No ages have been reported for rabbits used in IRE test method in the literature. In terms of the weight range, there have been reports of differences in corneal thickness between rabbits that weigh less than 2000 grams and those that weigh more than 2000 grams (Burton et al., 1972). When two ranges of rabbit weights were compared with respect to corneal thickness, animals in the 1300 to 2000 gram range had corneal thickness measurements of 0.346 ± 0.02 mm (mean and standard deviation [SD], 156 eyes) versus 0.382 ± 0.017 mm (mean and SD, 18 eyes) for rabbits in the 2000-2700 gram range. However, there have been no reports regarding differences in the ability of the IRE test method to detect ocular corrosives and severe irritants depending on rabbit weight or corneal thickness. Published IRE studies typically report on the use of rabbits in the 2500 to 4000 gram weight range. Although corneal thickness in rabbits depends on animal weight, no studies have been conducted to evaluate whether differences in corneal thickness would alter the performance characteristics of the IRE test method. Furthermore, there are no reported studies comparing use of rabbit strains other than New Zealand White in the IRE test method and the consequence of the use of other strains (e.g., California, New Zealand Red) are unknown.

2.2.1.2 *Quality of Eyes*

Currently, there are no standardized criteria for the selection of rabbit eyes for the IRE assay. Most IRE studies reported that eyes were carefully examined visually for defects, including opacity, scratches or pitting, pannus or neovascularization, once they had arrived at the laboratory. A few studies also noted use of stereomicroscopes or loupes to assist in identifying damaged corneas. Some laboratories reported use of fluorescein to assist in the identification of corneal epithelial barrier defects.

The quality of the eyes is typically evaluated at later steps in the assay, as well. For example, an increase in corneal thickness measured just before and/or after equilibration of greater than 7 to 10% relative to that of the corneal thickness measurement taken during the initial eye examination following enucleation would result in rejection of the use of that eye in an assay.

2.2.1.3 *Preparation of the Eyes*

In general, rabbits are euthanized by an intravenous injection of a lethal dose of sodium pentobarbitone (approximately 200 mg/kg) into the marginal ear vein. The corneas are kept moist after sacrifice with drops of physiological saline (prewarmed from 31 to 32°C) applied throughout the dissection process. Although the dissection process is not typically described in the literature reports, scientists with expertise in performing the dissection have provided details of the procedure (Jones P, Guest R, personal communication). The nictitating membrane is deflected away using forceps and the conjunctivae are cut using angled forceps and curved scissors. The eyeball is proptosed by applying gentle pressure with fingers above and below the orbit. The remaining conjunctival tissue, the orbital muscles and the optic nerve (leaving approximately a 5-10 mm section to prevent loss of intraocular pressure) are

removed and the eyeball is lifted from the orbit. Any tissue adhering to the globe is then removed by careful dissection, and the eyeball is gently rinsed with a stream of physiological saline to remove any adherent debris. The eyes are prepared for immediate use or for shipment as described in **Section 2.2.1.1**.

2.2.1.4 *IRE Experimental Setup*

Burton et al. (1981) provided a description and drawing of the original IRE experimental eye incubation apparatus, which was termed a superfusion chamber, and most studies to date have used slight variations of this original instrumental setup, usually expanding the number of eyes that can be accommodated by the apparatus for a single experiment. The superfusion apparatus is a large Perspex chamber that has a water-jacketed surface that maintains the temperature of multiple (usually six to eleven) individual cells that house the isolated eyes during the experiment. The dissected eyes are mounted in specially designed Perspex, plastic or metal holders with rings or studs on both the upper and lower jaws of the holding clamp, which provide just enough pressure to keep the eyes from slipping while maintaining intraocular pressure. The holders are designed to fit into the individual cells of the superfusion apparatus with the eyes maintained in a vertical position. The holders can be readily removed and placed in a horizontal position for test substance application. The individual chambers are typically blackened to permit slit-lamp observations. A saline drip tube is mounted over the eye within the chamber and a steady drip of warm saline is used to maintain the eye in a hydrated condition. The temperature maintained in the cells of the superfusion apparatus is typically 31 to 32°C with a range of approximately 1.5°C. A water bath and two peristaltic pumps are used to heat and circulate the water and saline used for temperature control. The flow rates vary due to changes in ambient conditions in the laboratory, but are typically around 4 liters/minute for the water and range from 0.1-0.4 mL/min for the saline drip.

2.2.2 Dose-Selection Procedures, Including the Need for Any Dose Rangefinding Studies or Acute Toxicity Data Prior to Conducting a Study

As described below in **Section 2.2.4**, test substances are typically evaluated undiluted at a fixed volume of 0.1 mL liquid or a fixed weight of 100 mg powdered solid (prewetted if necessary). Because a fixed quantity of a substance is tested, dose selection is not a relevant issue.

2.2.3 Endpoints Measured

2.2.3.1 *Corneal Opacity*

The original developer of the IRE test method (Burton et al. 1981) noted changes in corneal opacity visually, by slit-lamp, and by fluorescein staining to assess the extent of corneal injury (i.e., effects on corneal stroma and/or epithelium). Corneal opacity was not formally scored in the original report, but was used in conjunction with corneal swelling measurements (see **Section 2.2.3.2)** to produce an arbitrary irritancy rating (e.g., negligible, slight, moderate, severe) for comparison of data on 10 test substances with *in vivo* rabbit eye data. The *in vivo* Draize et al. (1944) scoring system or a slightly modified version of it for assessment, although subjective, is now routinely used to score corneal opacity and area (Prinsen and Köeter 1985; Whittle et al. 1992; Balls et al. 1995; Jones et al. 2001). In addition, the McDonald-Shadduck ocular scoring system (Hacket and McDonald 1991) was

based on slit-lamp observations and proposed as an *in vivo* alternative to the Draize method to reduce subjectivity. Guerriero et al. (2004) used the McDonald-Shadduck scoring system for the identification of severe ocular irritants and it is the *in vitro* scoring method used routinely for isolated rabbit eye irritation testing at SafePharm Laboratories (Guest R, Whittingham A, personal communication).

2.2.3.2 *Corneal Swelling*

Burton et al. (1981) used a depth-measuring attachment on a slit-lamp microscope to measure corneal thickness in the isolated rabbit eye and then used relative changes in corneal thickness 4 hours after application of the test substance to provide a quantitative measurement of corneal swelling. Since then, corneal swelling is routinely used as an ocular endpoint at various times after application of a test substance (York et al. 1982; Price and Andrews 1985; Prinsen and Köeter 1985; CEC 1991; Balls et al. 1995; Gettings et al. 1996; Jacobs and Martens 1988; Cooper et al. 2001; Guerriero et al. 2004).

Corneal swelling may be calculated from corneal thickness measurements using the following equation:

$$\left(\frac{\textit{corneal thickness at time } t \quad - \quad \textit{corneal thickness at time} = 0}{\textit{corneal thickness at time} = 0}\right) \times 100$$

In general, corneal thickness is measured as a quantitative endpoint and corneal swelling is typically calculated at 0.5, 1, 2, 3 and 4 hours.

2.2.3.3 *Fluorescein Penetration/Retention*

Although fluorescein staining is used routinely to assess the integrity of eyes used in the IRE test method, it is not used routinely as a scored endpoint. However, several investigators report the use of fluorescein retention or penetration as a scored ocular endpoint to supplement general observations regarding corneal opacity and to provide an indication of the area of the cornea affected and the type of lesion produced (e.g., diffuse, stippled, focal) or as an indicator of the depth of penetration of the injury to the cornea (i.e., intensity of fluorescein stain). Fluorescein staining has been applied by some investigators at 0.5 hours (CEC 1991), but is more often applied at 4 hours after the other endpoints are evaluated for any effects produced by the test substance (CEC 1991; Prinsen and Köeter 1985; Guerriero et al. 2004). Scoring systems for fluorescein area and intensity staining of isolated rabbit eyes are generally arbitrary, but well defined (e.g., scales of 0 to 3 or 0 to 4) in a manner similar to ocular opacity and area scores (Prinsen and Köeter 1985; Guerriero et al. 2004).

2.2.3.4 *Assessment of Epithelial Integrity*

Guerriero et al. (2004) reported the evaluation of epithelial integrity visually and by slit-lamp as an indicator of severe ocular corrosion or irritation. Any significant indication of corneal epithelial stippling, mottling, pitting, ulceration, pannus, or other significant or irreversible

corneal epithelial defects are considered sufficient to indicate that a substance is an ocular corrosive or severe irritant.

2.2.3.5 *Additional Endpoints*

Routine or selective use of histopathology may provide useful information regarding: 1) the depth of ocular injury; 2) characteristics of the injury at the cellular level; and 3) integrity of critical and irreplaceable cellular components such as endothelial or stem cells. Furthermore, histopathology can be combined with modern staining techniques (e.g., vital dyes, immunohistopathology, biochemical markers) to provide information on possible mechanisms of ocular toxicity. Unfortunately, not all laboratories are equipped to perform histopathology, or do not have access to or a collaboration with other laboratories that have this capability. Furthermore, the added cost of routine histopathology might impact consideration of the use of histopathology by an ocular toxicity-testing laboratory, especially if the burden of the increased cost falls on the client.

Confocal microscopy has been used with vital dyes to measure the depth of corneal injury as the level of penetration of the live/dead cell layer (Jester et al. 1996; Jester et al. 2001). The extent of this corneal injury has been suggested as the mechanistic basis for ocular irritation (Maurer et al. 2002). Again, the increased cost of a confocal microscope may be a significant burden to many laboratories.

2.2.4 Duration of Exposure

2.2.4.1 *Pre-exposure Preparations*

In most published studies, once the isolated rabbit eyes are assessed for their utility in the assay and deemed free of ocular defects, they are equilibrated in a superfusion apparatus based on the one originally described by Burton et al. (1981) for a period of 30 to 60 minutes at temperatures ranging from 31 to 32 °C. The eyes are mounted in special plastic or metal holders fitted with rings or pins used to gently clamp the eye in place and to prevent them from slipping and to minimize changes in intraocular pressure. Using these holders, the eyes are removed from the superfusion apparatus for application of the test substance, then returned to the original vertical position under the saline drip tube for incubation.

2.2.4.2 *Test Substance Exposure Duration*

Following equilibration, the isolated rabbit eye is typically exposed to the test substance for a total of 10 (± 2) seconds at which time it is gently rinsed off with a volume of 20 mL of physiological saline (prewarmed to 31 or 32°) using a syringe or other means of delivery. The 10-second exposure period is the standard time used by most investigators to identify and assess the ocular effects of severe eye irritants. A note is recorded if any particles of solid or precipitated material remain on the surface of the cornea after rinsing. Some authors have increased the time of exposure to 1 minute (Cooper et al. 2001; Jones et al. 2001) when evaluating select products such as hair shampoos that are generally in the mild to moderate range of eye irritation.

2.2.4.3 *Application and Amount of the Test Substance*

A volume of 0.1 mL of a neat liquid test substance or 0.1 gram of a solid ground to a fine powder is typically applied to the isolated rabbit eye using a syringe. The isolated eye is

removed from the equilibration chamber and placed in a horizontal position with the cornea facing upward for application of the test substance. This is the standard scientific practice for volume or weight used by most investigators in the *in vivo* rabbit eye test in the relevant literature and it is the application volume and weight currently accepted by the U.S. (EPA 1998) and EU (EU 2001, UN 2003 [GHS]) regulatory agencies for *in vivo* studies.

2.2.4.4 *Number of Eyes Required per Test Substance and Controls*

Historically, in IRE studies, one to three isolated rabbit eyes have been used to assess the ability of a test substance to induce corrosion or irritation; one isolated rabbit eye has been used as the negative control and, if included, one isolated rabbit eye has been used to assess the ability of a positive control substance to induce an appropriate response (Whittle et al. 1992; Balls et al. 1995; Gettings et al 1996; Jones et al. 2001). Some authors include benchmark controls (Jones et al. 2001).

2.2.4.5 *Concentration of Test Substance*

For regulatory purposes, substances are generally tested neat in the IRE. However, there have been reports comparing neat liquid test substances with 10% dilutions (Cooper et al. (2001; Jones et al. 2001) at 10 and 60 second exposure times to differentiate mild and moderate eye irritants for select product types such as hair shampoos.

2.2.5 Known Limits of Use

The IRE was designed as an *in vitro* test method to measure effects on the isolated eye; namely effects on the cornea. However, scoring of irritancy of the iris similar to that reported in the Draize assay such as swelling, injection, reaction to light, and hemorrhage is not possible, because the physiological mechanisms for their production (principally blood flow and muscular activity) are not present in the isolated eye. In addition, the dissection and removal of the conjunctiva during removal of the eye precludes testing irritant effects on this tissue. Although a severe ocular irritant may produce significant corneal damage, the EPA has documented cases in which severe conjunctival irritancy persisted for 21 days in the absence of significant corneal opacity, and the test substance was therefore labeled a severe ocular irritant according to the EPA (EPA 1996) classification system (Lewis M, personal communication). Thus, severe irritancy of a test substance resulting from severe effects on the conjunctiva cannot be identified in the IRE assay. Finally, reversible ocular effects cannot be evaluated in the IRE test method.

2.2.6 Nature of the Response Assessed

As noted in **Section 2.2.3**, the corneal endpoints observed in the IRE are opacity, swelling, fluorescein retention, and morphological effects on the epithelium. The severity of each response is graded at each time point (with the exception of fluorescein retention which is generally assessed only at four hours so as not to interfere with the other endpoint evaluations). The data to be collected includes both numerical and descriptive data. The numerical data includes scores for corneal opacity, corneal thickness, and fluorescein retention, while the descriptive data represents morphological and/or histopathological findings. Alternative endpoints such as histopathology and confocal microscopy for evaluation of depth of corneal injury are available if it becomes necessary to differentiate a moderate response from a severe response, when the existing endpoints do not permit this

level of differentiation. Use of vital dyes alone or with confocal microscopy, immunohistopathology, or biochemical markers may be useful to assess mechanistic aspects of a severe irritant.

2.2.7 Appropriate Controls and the Basis for their Selection

2.2.7.1 Negative Controls

The negative control provides a baseline for the assay endpoints, ensures that the experimental conditions do not inappropriately result in an irritant response, and permits detection of nonspecific changes in the test system. The most frequently reported negative control in published IRE studies is isotonic saline. This would appear to be the most suitable control since the test method is conducted using isotonic saline to bathe the rabbit eyes as well as for the requisite rinsing steps. Treating the negative control eyes with isotonic saline ensures that any mechanical alterations (i.e., those not related to the test substance) are properly controlled. There have been no formal studies to assess whether results obtained using buffered salt solutions (e.g., Ringer's) would be similar to those using isotonic saline. For example, it is not known whether buffered salt solutions would be an impediment to proper evaluation of an acidic or basic test substance due to pH control, or if they would increase or reduce the false positive or false negative rates and impact accuracy.

2.2.7.2 Solvent/Vehicle Controls

Based on a review of published IRE studies, it appears that concurrent solvent controls have not been used. However, it is scientifically critical to know that the vehicle for the test material, if different from isotonic saline, has an impact on the outcome of the study by producing irritancy on its own and possibly impeding the detection of irritancy of the test substance.

2.2.7.3 Positive Controls

As discussed by Harbell and Curren (2002), the function of the positive control is to ensure that the test method is operating within normal limits and that each experiment is properly executed, such that the toxic effects of interest can be properly detected. A concurrent positive control substance is included in each experiment to develop a historical database. Results from the concurrent positive control are compared to the historical control range, which is used to determine whether a particular experiment is acceptable. Because the positive control should allow for detection of an over- or under-response in the assay, the selected positive control should not produce responses at either the extreme low or the extreme high end of assay response. In the literature, positive controls have not historically been used in the IRE test method, because positive controls would typically be severe irritants such as 10% sodium hydroxide, by nature of the design of the original IRE test method (Burton et al. 1981). As discussed by Harbell and Curren (2002), the importance of a positive control cannot be overemphasized, and perhaps severe irritants producing less than a maximal level of ocular damage could be considered as positive controls to permit assessment of variability over time and to insure the integrity of the test system and its proper execution.

2.2.7.4 Benchmark Controls

Benchmark substances are often used during the testing of substances of unknown toxicity potential. The toxicity of the benchmark substance is generally well characterized (i.e., adequate human or animal toxicity data are available). A benchmark is selected to match the chemical or product type of the unknown substance, and is used to set an upper or a lower limit of response against which the unknown is compared (Harbell and Curren 2002). Benchmark substances are often selected from a list of reference chemicals for the assay and have the following properties:

- consistent and reliable source(s)
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in the *in vivo* rabbit eye test
- known potency in the range of the desired response

They are useful for evaluating the ocular irritancy potential of unknown chemicals of a specific chemical or product class, or for evaluating the relative irritancy potential of an ocular irritant within a specific range of irritant responses.

2.2.8 Acceptable Range of Control Responses and the Basis for the Acceptable Ranges

2.2.8.1 *Negative/Solvent Controls*

Negative controls are generally considered acceptable if they produce a nonirritant response. Solvent controls are not typically used in the IRE test method, but would also be expected to produce a nonirritant response. A positive solvent response (mild or moderate irritation) would generally require replacement of the irritating solvent with a nonirritating solvent, unless it was part of a formulation that could not be changed. The basis for the acceptable range of negative controls were derived from observations made by laboratories experienced in the performance of the IRE assays (Jones P, Prinsen M, Harbell J, personal communications) and from information in articles that describe the IRE test method in the literature.

2.2.8.2 *Positive Controls*

Because positive controls have not been traditionally employed in this test method, a defined range of response has not been described previously. However, the positive control substance should produce a response that is appropriate based on its historical classification as a severe irritant in the *in vivo* rabbit eye test. If adequate historical IRE test method data are not available for a particular positive control, pilot studies may have to be conducted to provide this information. Positive controls should produce the anticipated response in order to ensure that the test method is performing correctly.

2.2.8.3 *Benchmark Controls*

Benchmark substances may be useful in demonstrating that the test method is functioning properly for detecting the ocular irritancy potential of chemicals of a specific chemical class or a specific range of response, or for evaluating the relative irritancy potential of an ocular irritant. Therefore, benchmark substances should produce an irritation response that is within acceptable limits of historical data.

2.2.9 Nature of the Data to be Collected and the Methods Used for Data Collection

2.2.9.1 *Corneal Opacity and Area of Involvement*

Corneal opacity and area are typically evaluated visually and by use of a slit-lamp microscope. The most common scheme used to quantify corneal opacity and area of involvement uses the Draize scoring system (Draize et al. 1944). In this method, the severity of corneal cloudiness and the area of the cornea involved are graded and a score is assigned for each parameter using various scoring schemes. However, not all authors include an assessment of the area of corneal involvement in the IRE test method. The reason for this is unclear. The method described by Hackett and McDonald (1991) is an updated version of the original McDonald-Shadduck scoring system (McDonald and Shadduck (1977). The updated version of this scoring system is presented in **Table 2-1**. This method is similar in scoring to the Draize method, but is more specifically targeted to slit-lamp observations and describes corneal effects in terms of what is actually observed with the slit-lamp at each increasing level of corneal damage and score. Like the Draize assay, this method also allows for separate examination and scoring of the area of corneal involvement. Raw data are typically recorded in notebooks and electronically.

Table 2-1 Evaluation of Corneal Irritation[1]

Appearance	Score
Normal cornea. Appears with the slit-lamp as having a bright grey line on the epithelial surface and a bright grey appearance on the stroma.	0
Some loss of transparency. Only the anterior half of the stroma is involved as observed with an optical section of the slit-lamp. The underlying structures are clearly visible with diffuse illumination, although some cloudiness can be readily apparent with diffuse illumination.	1
Moderate loss of transparency. In addition to involving the anterior stroma, the cloudiness extends all the way to the endothelium. The stroma has lost its marble-like appearance and is homogenously white. With diffuse illumination, underlying structures are clearly visible.	2
Involvement of the entire thickness of the stroma with endothelium intact. With the optical section, the endothelial surface is still visible. However, with diffuse illumination the underlying structures are just visible.	3
Involvement of the entire thickness of the stroma with endothelium damaged. With the optical section cannot clearly visualize the endothelium. With diffuse illumination, the underlying structures cannot be seen.	4
Area	**Score**
Normal cornea with no area of cloudiness	0
1 to 25% area of stromal cloudiness	1
26 to 50% area of stromal cloudiness	2
51 to 75% area of stromal cloudiness	3
76 to 100% area of stromal cloudiness	4
Overall Corneal Opacity/Area	**Product Score**
Corneal Opacity x Area[2]	Maximum of 16

[1] From: Hackett and McDonald (1991)

[2]The overall corneal opacity score is the product of the corneal opacity score and the corneal area score. The product of individual scores of 1 and 4 (Product Score of 4) or 2 and 2 (Product Score of 4), for example, would each qualify for a severe irritant rating based on the overall corneal opacity/area score.

2.2.9.2 *Corneal Thickness and Calculation of Corneal Swelling*

Corneal thickness is measured quantitatively using an optical pachymeter (Attachment No. 1) for the Haag-Streit slit-lamp biomicroscope (e.g., Haag-Streit AG or equivalent, Liebefeld-

Bern, Switzerland), or by an ultrasonic pachymeter (e.g., DGH Technology Inc., Solana Beach, California, USA) (Jones P, Guest R, personal communication). The optical pachymeter measures degree of reflection or refraction from a normal light slit imposed into the corneal surface from the slit-lamp. The ultrasonic pachymeter measures the transit time of high -frequency sound pulses beamed into the eye and reflected off tissue interfaces where high impedance gradients are encountered (Jacobs and Martens 1988). From the known acoustic velocity of tissue, transit times may be converted to distance. Corneal swelling results induced by a variety of test substances using both measuring systems were found to be comparable (Jacobs and Martens 1988). One advantage of the ultrasonic pachymeter is that measurement of corneal thickness is usually possible even when corneal opacity has been induced, while this may not always be possible when using the optical pachymeter. Another advantage is that measurement of corneal thickness can be conducted at any position on the cornea, which is not possible with the optical pachymeter. Corneal thickness can therefore be measured at various positions on the cornea and a mean value obtained for each eye. However, if injury to the cornea is restricted to a small area, it may be more appropriate to measure corneal thickness at this position.

Corneal swelling is measured as the percent increase in thickness at each time point relative to the measurement at T0 (after equilibration, before treatment) as follows:

$$\left(\frac{\textit{corneal thickness at time } t \quad - \quad \textit{corneal thickness at time} = 0}{\textit{corneal thickness at time} = 0}\right) \times \quad 100$$

The level of corneal swelling needed for a test substance to be considered an irritant varies in the literature depending on what type of instrument is used and the experience of the laboratory performing the study. Levels of irritancy may be assigned to a test substance based on 20 or 25% swelling over an entire range of time (e.g., 0 to 4 hours) (Balls et al. (1995; Gettings et al. 1996; Guerriero et al. 2004) or based on differences in swelling over time (CEC 1991; Prinsen and Köeter 1995). For example, 20-25% swelling in one hour may produce a higher irritancy rating than 20-25% in 5 hours. For the purposes of the analyses used in this BRD, a decision criterion for identification of a severe irritant is a corneal swelling value equal to or exceeding 25%.

2.2.9.3 Fluorescein Penetration

Fluorescein is used as an aid to further define epithelial damage (Norn 1971). In the IRE literature, a drop or two of a 1 to 2% solution of fluorescein sodium (sufficient to cover the cornea) is generally applied and left for several seconds followed by rinsing with isotonic saline. Fluorescein penetration is typically measured at 4 hours after application of the test substance. Most authors use a scoring system similar to that shown in **Table 2-2**. However, some protocols use a broader range of fluorescent terms (e.g., moderately bright, extremely bright) and scores from 0 to 5. Although the use of 1 to 2% fluorescein is recommended, variations in concentration of fluorescein from batch to batch may require some adjustment to achieve the desired corneal effect (Chambers W, personal communication).

Table 2-2 Fluorescein Penetration Scoring System[1]

Description	Scores
	Area/Intensity
Negligible – No staining.	0
Slight staining confined to small focal area. Some loss of detail in underlying structures with diffuse illumination.	1
Moderate staining confined to a small focal area. Some loss of detail in underlying structures on diffuse illumination.	2
Marked staining involving a larger portion of the cornea. Underlying structures are barely visible, but not completely obliterated with diffuse illumination	3
Extreme staining with no visibility of underlying structures.	4
Overall Fluorescent Area/Intensity	**Product Score**
Fluorescent Area x Intensity[2]	Maximum score of 16

[1]From: Hackett and McDonald (1991)

[2] Fluorescent area and intensity scores are determined individually and the overall Fluorescent Penetration Score is the product of both measurements. For example, an area score of 1 and intensity score of 4 would produce an overall score of 4, which meets the criteria for a severe irritant. An area score of 2 and intensity score of 2 would produce a product score of 4, which also exceeds the cut-off for a severe irritant, although each individual score did not.

2.2.9.4 Evaluation of Corneal Epithelial Integrity

As described by Guerriero et al. (2004), the cornea may also be observed macroscopically or microscopically using a slit-lamp to evaluate any injury to the epithelium. Stippling, pitting, mottling, sloughing, ulceration, or any other unusual effects on the epithelium are noted and reported.

Observations of the integrity of the corneal epithelium may be done visually or with a slit-lamp when laboratories are not equipped to perform routine histology. However, histology may also be used as an additional method for more precise evaluation of the integrity of the corneal epithelium.

2.2.9.5 *Overall Scoring System for Identification of a Severe Irritant*

Assignment of irritant classification or categories to test substances evaluated in the IRE test method varies from study to study. For example, in the CEC (1991) study, irritancy ratings of A, B, C, or D were assigned ranging from least to most severe that were arbitrarily based on the results from a combination of endpoints (corneal opacity, corneal swelling, and fluorescein retention). In the Balls et al. (1995) study, irritancy of a test substance was based on the mean corneal swelling measurement or corneal opacity score and then ranked accordingly. In the Gettings et al. (1996) study, an irritancy rating was assigned based on a predetermined cutoff using the percentages of corneal swelling. In the Guerriero et al. (2004) study, an overall scoring system for the identification of severe irritants was based on a test substance meeting or exceeding predetermined cutoff values in any of four ocular endpoints evaluated (corneal opacity and area, corneal swelling, fluorescein area and intensity, and integrity of the epithelium) (**Table 2-3**).

Table 2-3 Overall Scoring System for Corneal Damage and Irritation[1]

Ocular Parameter	Cutoff Value to Detect Severe Eye Irritants
Maximum Corneal Opacity[2] (Cloudiness x Area)	Greater than or equal to a score of 3
Maximum Fluorescein Uptake[3] (Intensity x Area)	Greater than or equal to a score of 4
Mean Corneal Swelling[4] 0.5 hours 1 hour 2 hours 3 hours 4 hours	Greater than or equal to 25%
Corneal Epithelial Observations[5]	Any pitting, mottling, stippling, sloughing, or ulceration of epithelium

[1] From: Guerriero et al. (2002)
[2] Represents maximum score obtained in three eyes
[3] Represents maximum score obtained in three eyes
[4] Represents mean swelling calculated for three eyes
[5] Represents information obtained for any single animal

2.2.10 Types of Media in Which Data are Stored

Although not specifically mentioned in published IRE protocols, it is reasonable to assume that data from studies performed in compliance with GLP guidelines (Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004) were stored in a manner suitable for GLP compliant studies. It would seem appropriate that data from the IRE be stored and archived in a manner consistent with international GLP guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines are nationally and internationally recognized rules designed to produce high-quality laboratory records. These guidelines provide a standardized approach to report and archive laboratory data and records, and information about the test protocol, to ensure the integrity, reliability, and accountability of a study (EPA 2003a,b; FDA 2003). Materials that should be retained include, but are not limited to, raw data, documentation, protocols, final reports, records and reports of the maintenance and calibration of apparatus, validation documentation for computerized systems, the historical file of all Standard Operating Procedures, and environmental monitoring records. The archives should be organized and indexed so that retrieval of all information can be done expediently and conditions of storage should minimize deterioration of the documents. An individual should be identified as responsible for these data archives. All raw data from the experiment should be recorded using a system that meets institutional and GLP requirements.

2.2.11 Measures of Variability

Both numerical and descriptive data are generated using IRE. Variability of numerical data is typically assessed through calculation of the mean along with the standard deviation for each numerical endpoint. Other descriptive statistics (e.g., coefficient of variation or CV) may be used in the analysis of variability. These values allow for an assessment of the performance of the test conducted and whether the observed variability between replicates or groups of replicates is greater than would be considered acceptable. Descriptive data may also provide an additional subjective measurement of variability.

2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data

For statistical analysis, most studies rely on calculation of the mean and standard deviation of individual endpoint data produced by a test substance (CEC 1991; Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). Other descriptive statistical methods may also be used in analysis of *in vitro* data, or for comparison of the *in vitro* data with *in vivo* rabbit eye test data (e.g., coefficient of variation or CV, ANOVA, regression, rank correlation). As an example, in the EC/HO validation study, Balls et al. (1995) used regression analysis of MMAS scores with mean results from the IRE test data (i.e., corneal opacity or corneal swelling at two time points) and obtained Spearman's rank correlation coefficients to compare *in vivo* and *in vitro* test results for a set of 59 test substances.

2.2.13 Decision Criteria and the Basis for the Prediction Model Used to Classify a Test Chemical as a Severe Eye Irritant

Once the individual mean endpoint data are obtained, studies vary in the methods used to assign an irritation classification based on the degree of severity of the ocular response using composite endpoint data. The irritant classifications assigned may be either descriptive (e.g., nonirritant, mild, moderate or severe) (Cooper et al. 2001; Jones et al. 2001) or scaled rankings of increasing or decreasing irritancy (e.g., numerical [0 to 4] or alphabetical [A to D]), based on predetermined, arbitrary endpoint values and, are occasionally, time dependent (i.e., corneal swelling) (CEC 1991; Köeter and Prinsen 1995). In addition, identification of severe irritants may be based on meeting or exceeding predetermined cutoff values (Guerriero et al. 2004). These predetermined endpoint values may be selected on the basis of statistically derived decision criteria (Prediction Model) using biostatistical approaches such as discriminant analysis. These decision criteria (Prediction Model) may be targeted to either a general population or to a select population of test substances (e.g., surfactant-based products) that may vary with respect to prevalence of a particular level or range of severity of ocular irritation. The statistical methods used for the determination of these decision criteria are not usually provided in publications. An example of the decision criteria used by Guerriero et al. (2004) for the identification of severe ocular irritants is shown in **Table 2-3**.

2.2.14 Information and Data That Will be Included in the Study Report and Availability of Standard Forms for Data Collection and Submission

It would seem appropriate that the test report include the following information, if relevant to the conduct of the study:

Test and Control Substances

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known
- The CAS Registry Number (RN), if known
- Purity and composition of the substance or preparation (in percentage[s] by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- Stability, if known

Information Concerning the Sponsor and the Test Facility

- Name and address of the Sponsor
- Name and address of the test facility
- Name and address of the Study Director

Justification of the Test Method and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time (e.g., periodic testing of proficiency substances, use of historical negative and positive control data)

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data
- If applicable, acceptable concurrent benchmark control ranges based on historical data

Test Conditions

- Experimental starting and completion dates
- Details of test procedure used
- Test concentration(s) used
- Description of any modifications of the test procedure
- Reference to historical data of the model (e.g., negative and positive controls, proficiency substances, benchmark substances)
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., irritancy scores for the test substance and the positive, negative, and benchmark controls, including data from replicate repeat experiments as appropriate, and means ± the SDs for each experiment)

Description of Other Effects Observed

Discussion of the Results

Conclusion

A Quality Assurance Statement for Good Laboratory Practice (GLP)-Compliant Studies

- This statement indicates all inspections made during the study, and the dates any results were reported to the Study Director. This statement also serves to confirm that the final report reflects the raw data.

Additional reporting requirements for GLP-compliant studies are provided in the relevant guidelines (e.g., OECD 1998; EPA 2003a, 2003b; FDA 2003).

The use of a standardized scoring form may be appropriate (**Figure 2-1**).

2.3 Basis for Selection of the Test Method System

The IRE test method was designed as an *in vitro* assay to measure the direct effects of severe irritant test substances on the corneal tissue of a rabbit eye (Burton et al. 1972, 1981). The purpose of the IRE was to use rabbits that had been previously euthanized, thus preventing the pain and suffering associated with application of severe eye irritants to live animals. Furthermore, rabbits used for this assay could be obtained from other laboratories (e.g., rabbits used for skin irritancy testing or physiology studies in which the eyes were unaffected) minimizing the need for additional animals. For larger numbers of animals, the rabbits could be obtained from a local abattoir where the animals are bred and used as a food source. Therefore, the use of the IRE as a prescreen or as a replacement assay could reduce the number of animals used in the *in vivo* Draize rabbit eye test. Furthermore, the assay is refined by elimination of pain and suffering. As discussed in **Section 1.1.1**, the cornea is one of the main tissues targeted during accidental eye exposures. In addition, corneal effects are weighed heavily in the original *in vivo* ocular irritancy scoring systems (e.g., 80 out of a possible 110 points in the Draize eye test scoring system). Therefore, although conjunctival and iridal tissue are not available for use as endpoints in the IRE, most of the scoring capacity of the *in vivo* rabbit test method is maintained and other endpoints such as corneal swelling and fluorescein penetration may be incorporated. Furthermore, use of histopathology and/or confocal microscopy can be used to qualitatively or quantitatively assess the depth of penetration of ocular injury in the IRE.

2.4 Proprietary Components

The IRE assay does not employ any proprietary components.

2.5 Basis for the Number of Replicate and Repeat Experiments

The irritancy of a test substance is normally determined using three rabbit eyes if the distribution of the data is within acceptable limits (no single value exceeds a statistically acceptable deviation from the mean group response). The experiment is typically repeated when an individual data point is outside the range of random distribution as determined by appropriate statistical methods or failure to meet predetermined acceptance criteria, such as the various controls are outside the range of historically acceptable data, corneal swelling in a negative control eye exceeds 7 to 10%, and/or when equivocal results are obtained for the test substance (Jones P, Guest R, personal communication).

Figure 2-1 Sample Scoring Form for the IRE Test Method

Substance Name:	**Test Facility:**	**Study No.:**
CASRN:	**Chem Class:**	**Prod Class:**
Date:	**pH:**	
Color of Material:	**Liquid or Solid:**	**Viscosity of Material**

	Corneal Opacity/Area Score							**Fluorescein Penetration Score**	
	Hours after Application of Test Substance								
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**	**0.5**	**2**
1									
2									
3									
4									
5									
6									
	Corneal Thickness (Instrument Units)							**Additional Comments**	
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**		
1									
2									
3									
4									
5									
6									
	Corneal Swelling (Percent of Time 0)								
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**		
1									
2									
3									
4									
5									
6									
	Corneal Epithelium Damage Assessment (See Legend Below)							**Epithelium Notations**	
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**		
1									
2									
3									
4									
5									
6									

Corneal Epithelium Damage Assessment: Pitting = PT; Stippling = ST; Mottling = MT; Sloughing; SL; Ulceration = UL; Other = OT (Describe Other Effects in Epithelium Notations)

2.6 Compliance with Good Laboratory Practice

GLP compliant studies are performed in compliance with regulatory GLP Guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003) to increase confidence in the quality and reliability of the test data. For potential submission of data using these test methods, compliance with appropriate GLP guidelines would be required.

2.7 Study Acceptance Criteria

A test is acceptable if the positive control gives a score for each ocular test parameter that falls within two SDs of the current historical mean, which should be updated on a regular basis. The negative/solvent control responses should be nonirritating and corneal swelling in each of negative control eyes should not exceed 7 to 10% (Jones P, Guest R, personal communication). As described in previous sections in detail, the McDonald-Shadduck scoring methodology from Hackett and McDonald (1991) is used to assess corneal opacity and is based on a description of slit-lamp observations of corneal damage from the epithelium to the endothelium. The decision criteria were designed to identify severe versus nonsevere irritants and are based on exceeding maximal cut-off values in any of four ocular test parameters. The cut-off values are based on a maximum corneal opacity score (opacity x area), maximum fluorescein penetration score (area x intensity), maximal corneal swelling, and observation of the corneal epithelium in which any indication of epithelial damage (e.g., pitting, mottling, stippling, sloughing, or ulceration) constitutes a severe irritant classification. A positive control, in addition to the negative control, is needed to ensure that operation of the test system is within normal limits. Benchmark controls should also be used to demonstrate test method function within an applicability domain (e.g., surfactant formulations). Ideally, a set of quality reference substances should be used for validation efforts such as that used by Balls et al. (1995).

3.0 SUBSTANCES USED FOR VALIDATION OF THE IRE TEST METHOD

3.1 Rationale for the Substances or Products Selected for Use

In vitro ocular test method validation studies should, ideally, evaluate an adequate sample of test substances and products from chemical and product classes that would be evaluated using the *in vivo* rabbit eye test method. Test substances with a wide range of *in vivo* ocular responses (e.g., corrosive/severe irritant to nonirritant) also should be assessed to determine any limit to the range of responses that can be evaluated by the *in vitro* test method.

Of the seventeen IRE reports considered in developing this BRD, four contained or the authors provided sufficient *in vitro* and *in vivo* data for an accuracy analysis[1]. These four reports are the CEC Collaborative Study (1991), Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004).

A total of 149 substances and formulations were evaluated in the four studies, of which 25 were commercial products or formulations. In the Guerriero et al. (2004) SOT study, substances that were unspecified substituted chemicals, such as pyridines, were initially placed under the single test substance name "substituted pyridines," but were assigned code numbers for differentiation with respect to data analysis. However, following the individual animal and *in vitro* IRE data submission from that study, GlaxoSmithKline granted permission to use the actual chemical names of the tested substances and provided that information along with MSDS sheets for the tested substances. **Sections 3.1.1** through **3.1.4** address the rationale for the chemicals or products tested in each of these studies.

3.1.1 CEC Collaborative Study (1991)

The Commission of the European Communities sponsored a collaborative study on possible alternative methods to the *in vivo* eye irritation test. This study was commissioned by the Division Control of Chemicals, Industrial Risks and Biotechnologies of Directorate General Environment, Nuclear Safety, Civil Protection and the Health and Safety Directorate of Directorate General Employment Industrial Relations and Social Affairs. The aim of this pilot study was to obtain reliable information about the relationship between the *in vivo* eye irritation test of Annex V and several alternative test methods (five *in vitro* assays, including an *ex vivo* IRE test method and a HET-CAM model). Twenty-one test substances were chosen to cover a full range of irritation potential. These test substances were supplied to each of three participating laboratories by the Fund for Replacement of Animals in Medical Experiments (FRAME) via a single supplier (Aldrich Chemical Company Limited, UK). Each test substance was derived from a single chemical batch. Ten of the 21 chemicals were selected from a list of 30 supplied by FRAME and the other 11 chemicals were selected because they were tested in a previous skin irritation study by the EC.

[1] The ability of the IRE test method to accurately identify test substances classified as corrosive or severe irritants is provided in **Section 6.0**. A description of the criteria and guidelines used by regulatory agencies to classify a substance as a corrosive or severe irritant is provided in **Section 4.0**.

3.1.2 Balls et al. (1995)

In the European Commission (EC)/British Home Office (HO) validation study, the test substances were initially selected from the 1992 European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Reference Data Bank for ocular irritation (ECETOC 1992) based on the following criteria:

- Substances should be single chemicals (no mixtures).
- Substances should be available at high purity and stable when stored.
- The *in vivo* rabbit eye test data should have been generated since 1981 according to the OECD Test Guideline (TG) 405 and in compliance with GLP guidelines.

Other criteria specific to the conduct of the studies are noted in the study report (Balls et al. 1995).

Originally, 60 substances were found in the ECETOC data bank that met the established criteria. However, this selection was determined to be inadequate due to the relatively low number of solid substances, the insufficient number of moderate to severe irritants, and the lack of pesticides. To avoid additional animal testing, the validation study management team attempted to locate high quality rabbit eye study data within the commercial sector. Subsequently, based on the availability of additional data (primarily from unpublished studies) that met the established criteria, the original list was modified to include more solids, some pesticides, and substances representing moderate to severe degrees of irritation. During the validation study, it was discovered that 14 of the reference substances had been tested by a protocol that involved rinsing or removal of the solid material from the eye one hour after application (rather than being allowed to remain continuously). Thus, the study protocols for these substances had not adhered to OECD TG 405. These 14 substances were retested *in vivo* and it was found that one, thiourea, was extremely toxic, killing the three rabbits on which it was tested. Based on this response, thiourea was excluded from the list of reference substances.

The final list of test substances included a total of 51 substances, four of which were tested at two different concentrations and two of which were tested at three concentrations, for a total of 59 different tests.

3.1.3 Gettings et al. (1996)

This report described results from Phase III of the CTFA Evaluation of Alternatives Program, a three-phase program that evaluated promising *in vitro* alternative test methods in relation to the *in vivo* rabbit eye test. Each phase of the program evaluated a specific product type; Phases I and II evaluated hydro-alcoholic and oil/water formulations, respectively, while Phase III evaluated surfactant-based personal care cleansing formulations. The rationale for using these surfactant-based formulations was that most commercial personal care products consist of several ingredients, and that there was a need in industry to predict correctly the irritation potential of complex mixtures. The 25 products tested in Phase III were representative surfactant-containing cleansing formulations, such as hair shampoos, liquid soap, eye make-up remover, and bubble bath. The selected formulations were chosen to provide a range of ocular irritancy responses in the *in vivo* rabbit eye test (from non-irritating

to moderately irritating, which is the highest level of irritancy generally achieved by this class of products. However, ten of the formulations with irritancy in the upper end of the desired MAS range (0 to 45) in a single animal test were diluted to 25% (v/v in distilled water) of 10 of the products to provide a wider distribution of irritant responses.

3.1.4 Guerriero et al. (2004)

Guerriero and his colleagues at GlaxoSmithKline (GSK) and SafePharm Laboratories (Derbyshire, United Kingdom) presented a study at the 2002 Society of Toxicology (SOT) Meeting that evaluated 30 pharmaceutical process materials (e.g., lactam, quinidine, acetophenone, sulfonamide, benzylamine, guanidine, piperazine) using the IRE test method as described in this BRD and compared the results to data obtained *in vivo* using the Draize rabbit eye test conducted concurrently. At the 2004 SOT meeting, Guerriero et al. (2004) presented a study using the IRE test method in which they tested 14 additional substances from the ECETOC database (ECETOC 1998). The rationale for the use of these pharmaceutical process chemicals was based on the potential exposure of pharmaceutical process workers to these substances and concern for worker safety in this environment. The ECETOC substances were used to expand the database. Although the test substances reported in the Guerriero et al. (2002, 2004) studies were originally coded and generic chemical names were used (e.g., substituted pyridine) as described in **Section 3.1,** the data obtained in that study and the actual names of the chemicals were eventually provided to NICEATM with permission from GlaxoSmithKline. Chemical and product classes were assigned to the test substances, and this information was used in the performance analyses.

3.2 Rationale for the Number of Substances Tested

No rationale was provided for the number of substances tested in any of the studies.

3.3 Chemicals or Products Evaluated

Physicochemical properties for each of the substances tested was obtained from information provided in the published reports and submitted data. No attempt was made to review original records to determine additional information about the test substances. Information, including substance name, Chemical Abstracts Service Registry Number (CASRN), chemical and/or product class, physicochemical properties, and literature reference for the substances tested in the IRE test method are shown in **Appendix B**. A chemical class was assigned for each test substance based on information found in the literature reference. If a chemical class was not assigned in the literature reference, the information was retrieved from the National Library of Medicine's ChemID Plus database.

As shown in **Table 3-1**, the chemical classes with the greatest amount of *in vitro* IRE data are surfactant-based formulations (25), alcohols (21), heterocyclic/aromatic compounds (18), acids (17), organic compounds (16), and carboxylic acids (16). Other chemical classes tested include inorganic chemicals, alkalis, ketones, esters, ethers, amines, amides, aldehydes, carboxylic acids, hydrocarbons, organometallics, and an organophosphate. The formulations tested include hair shampoos, personal care cleansers, detergents, bleaches, and a fabric

softener. Of the 149 substances included in **Appendix B**, all were categorized within one or more chemical class.

As shown in **Table 3-2**, the most common product classes tested in the IRE assay are chemical intermediates/raw materials (38), solvents (38), soap and surfactant-based products (28), solubilizers/emulsifiers/lubricants (14), shampoo and hair care products (12), herbicides/pesticides (12). fungicide/germicide (11), flavor additives/food ingredients (8) and detergents (8). Of the 149 substances included in **Appendix B**, all were categorized within one or more product class.

Table 3-1 Chemical Classes Tested in the IRE Test Method

Chemical Class	# of Substances	Chemical Class	# of Substances
Acetate/Ester	13	Halogenated compound	1
Acid	17	Hydrocarbon	2
Alcohol	21	Imide	2
Acyl halide	1	Inorganic	9
Aldehyde	2	Ketone	8
Alkali	4	Lactone	1
Amide	5	Onium	12
Amine	14	Organic	16
Amino acid	1	Organometallic	2
Amidine	2	Organophosphate	1
Boron compound	1	Nitrile	2
Carboxylic acid	16	Nitro compound	4
Cyclic hydrocarbon	1	Sulfur containing	9
Glycol	0	Polycyclic compound	1
Ester	13	Surfactant, anionic	2
Ether	10	Surfactant, cationic	8
Heterocyclic/Aromatic	18	Surfactant, nonionic	5
Formulation	25	Surfactant-based formulations	25

Table 3-2 Product Classes Tested in the IRE Test Method

Product Class	# of Substances	Product Class	# of Substances
Active pharmaceutical ingredient	6	Fungicide/Germicide	11
Antiseptic/Disinfectant	4	Household cleaner	1
Caustic agent	7	Plasticizer	7
Chemical intermediate/ Raw material	38	Shampoo/Haircare	12
Detergent	8	Soap/Surfactant	28
Herbicide/Pesticide	12	Solubilizer/Emulsifier/ Lubricant	14
Flavor additive/Food ingredient	8	Solvent	38
Fragrance/Perfume	5	Tanning agent	1

3.4 Coding Procedures Used in the Studies

The coding procedures used in the IRE validation studies were reported in the literature. No attempt was made to review original records to assess these procedures. Based on the available information, the only reports that identified using coded chemicals were Balls et al. (1995) and Gettings et al. (1996).

3.4.1 CEC Collaborative Study (1991)

Substances evaluated in the CEC collaborative pilot study were coded, but the identity of those substances was provided to each laboratory before the study.

3.4.2 Balls et al. (1995)

In the EC/HO study, Balls et al. allocated a numeric code for each test substance. In addition, each participating laboratory in the study was allocated a unique code number to permit analysis of the data without knowledge of which laboratory actually performed the test. The number of aliquots required for each substance by all of the participating laboratories was carefully determined and random codes were generated by computer software for the total number of substances and each sample was then assigned a unique number

3.4.3 Gettings et al. (1991; 1996)

A two-part system was developed to ensure that the identity of the test substances remained unknown during testing. The first part of the identification consisted of a sample ID that was unique for each test material to be distributed. The sample ID consisted of a two letter and one number combination. If additional samples were needed, the number was increased in sequence. The two letter code was chosen at random, but was unique to each sample and laboratory. The second part of the identification consisted of a sample number (which ranged from 1 to 12). The sample numbers corresponded to the 12 test substances provided in each shipment.

3.4.4 Guerriero et al. (2004)

Substances evaluated in the GSK studies by Guerriero and his colleagues were assigned generic nomenclature for proprietary reasons and were coded numerically for the purpose of differentiating similar chemicals with various chemical substitutions (e.g., substituted pyridine). The generic nomenclature, however, was provided to the testing laboratory before the study.

[This Page Intentionally Left Blank]

4.0 *IN VIVO* REFERENCE DATA USED FOR AN ASSESSMENT OF TEST METHOD ACCURACY

4.1 Description of Protocol Used to Generate *In Vivo* Data

4.1.1 Draize Rabbit Eye Test

The test method protocol most widely accepted by regulatory agencies for the evaluation of ocular eye irritants is based on the Draize rabbit eye test method. The methodology, originally described by Draize et al. (1944), involves instillation of 0.1 mL of the test substance (e.g., liquids, solutions, and ointments) into the conjunctival sac of an albino rabbit eye. In this test method, one eye is treated while the other eye serves as the untreated control. The eye is examined at selected time intervals after exposure and any injuries to the cornea, conjunctiva, and the iris are scored. Scoring is subjective and based on a discrete, arbitrary scale (**Table 4-1**) for grading the severity of ocular lesions. The scores for the observed ocular injuries range from 1 to 2 for iris effects, from 1 to 3 for conjunctival redness and discharge, and from 1 to 4 for corneal effects and conjunctival chemosis. A score of zero is assigned when the eye is normal and no adverse effects are observed. In the original protocol, the eyes were observed up to four days after application of the test substance. However in current practice, these time points vary according to the degree of irritation, the clearing time, and testing requirements imposed by the various regulatory agencies.

The original Draize protocol describes a scoring system in which each ocular parameter is graded on a continuous numerical scale. The scores may be weighted (as shown in **Table 4-1**); however, most classification systems today do not use a weighting factor. The weighting of the score by Draize et al. (1944) is biased more heavily for corneal injury, since injury to the cornea has the greatest probability of producing irreparable eye damage. To illustrate, each ocular parameter shown in **Table 4-1** is evaluated for each rabbit. The product of the opacity and area scores is obtained, then multiplied by a weighting factor of 5; the maximum corneal score is 80. The iris score is multiplied by a weighting factor of 5; the maximum score is 10. The scores for the three conjunctival parameters are added together and then the total is multiplied by a weighting factor of 2; the maximum score is 20. The overall score for each rabbit is calculated by adding the values for each parameter; the maximum total score is 110.

While the current test method is widely used, it has limitations. For example, because of reflexive pawing at the eye or tearing after instillation of a test substance, the exact dose and/or concentration of the test substance is unknown. Additionally, if observations are made at 24-hour intervals, it may not always be clear whether observed effects are associated with the test substance or an unobserved reflexive behavior.

Table 4-1 Scale of Weighted Scores for Grading the Severity of Ocular Lesions[1]

Lesion	Score[2]
Cornea	
A. Opacity – Degree of density (area which is most dense is taken for reading	
Scattered or diffuse area – details of iris clearly visible	1
Easily discernible translucent areas, details of iris slightly obscured	2
Opalescent areas, no details of iris visible, size of pupil barely discernible	3
Opaque, iris invisible	4
B. Area of cornea involved	
One quarter (or less), but not zero	1
Greater than one quarter, but less than one-half	2
Greater than one-half, but less than three quarters	3
Greater than three quarters up to whole area	4
Score equals A x B x 5 Total maximum = 80	
Iris	
A. Values	
Folds above normal, congestion, swelling, circumcorneal injection (any one or all of these or combination of any thereof), iris still reacting to light (sluggish reaction is positive)	1
No reaction to light, hemorrhage; gross destruction (any one or all of these)	2
Score equals A x 5 Total possible maximum = 10	
Conjunctiva	
A. Redness (refers to palpebral conjunctiva only)	
Vessels definitely injected above normal	1
More diffuse, deeper crimson red, individual vessels not easily discernible	2
Diffuse beefy red	3
B. Chemosis	
Any swelling above normal (includes nictitating membrane)	1
Obvious swelling with partial eversion of the lids	2
Swelling with lids about half closed	3
Swelling with lids about half closed to completely closed	4
C. Discharge	
Any amount different from normal (does not include small amount observed in inner canthus of normal rabbits	1
Discharge with moistening of the lids and hairs just adjacent to the lids	2
Discharge with moistening of the lids and considerable area around the eye	3
Score equals (A + B + C) x 2 Total maximum = 20	

[1]From Draize et al. (1944)
[2]Scores of 0 are assigned for each parameter if the cornea, iris, or conjunctiva are normal.

4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols

Since the original description of the *in vivo* rabbit eye test method, regulatory agencies in the U.S., as well as in other countries, have modified the test method protocol to suit their specific needs and goals in protecting human health (**Table 4-2**). Regulatory agencies generally recommend using healthy adult albino rabbits (e.g., White New Zealand). The eyes of each test rabbit are examined within 24 hours prior to test initiation. A quantity of 0.1 mL (for liquids) or 0.1 g (for pulverized solid, granular, or particulate test substances) is placed into the conjunctival sac of one eye of each rabbit, after pulling the lower lid away from the eyeball. The other eye remains untreated. The lids are held together for about one second to decrease loss of test substance from the eye. Although the observation period

varies, the eyes are typically examined at 24-hour intervals for at least 72 hours after application of the test substance for adverse effects to the cornea, conjunctiva, and iris. The length of the observation period should be sufficient to evaluate reversibility of any of the observed effects, but generally does not exceed 21 days. The ocular effects observed were usually those described by Draize et al. (1944) in **Table 4-1**. For current uses, other lesions, such as pannus[1] and herniation of the cornea, also are noted. Corneal, iris, and conjunctival lesions are scored using the individual numerical grades described in **Table 4-1**, but weighted scores and an overall score for irritation are not typically calculated or used for U.S. or European regulatory purposes.

Depending on the regulatory agency, the number of rabbits required for a study of ocular irritation can vary. To minimize pain and suffering of rabbits exposed to potentially corrosive agents, the EPA and European regulatory agencies suggest that, if a test substance is anticipated to produce a severe effect (e.g., corrosive effect), a test in a single rabbit may be conducted. If a severe effect is observed in this rabbit, further testing does not need to be conducted and classification and labeling of a test substance can proceed on the effects observed in a single rabbit. In cases where more than one rabbit is tested, at least three should be examined to classify the ocular effects produced by the test substance (EU 2004; EPA 1998). In contrast, regulations for other U.S. agencies (e.g., CPSC, FDA) require at least six rabbits be examined to classify the effects produced by a test substance (CPSC 2003). The differences in current *in vivo* test protocols in the U.S. appear to reflect each agency's objectives for eye irritation testing; EPA regulates industrial chemicals, while the CPSC and FDA regulate household consumer products, pharmaceuticals, cosmetics, and toiletries.

Various data transformations have been developed to compare and rate irritants of varying severity. One is the MAS, in which the Draize scores obtained at each time point are averaged and the highest score obtained is the MAS. The MAS value was later modified to the MMAS (Modified Maximum Average Score), which is the highest average MAS value beginning with the 24-hour time point (ECETOC 1998).

4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems

Although *in vivo* eye irritation test method protocols are similar across U.S. and international regulatory agencies, interpretation of the results from the *in vivo* test method varies considerably. Several classification systems are in use for regulatory ocular irritancy testing purposes (**Table 1-2**). In the United States, two major classification systems are currently used, the FHSA guideline (CPSC 1995), which is used by the FDA, OSHA, and CPSC, and the EPA guideline (EPA 1996).

[1] Pannus, also known as "chronic superficial keratitis", describes a specific type of corneal inflammation. Pannus is caused by a local inflammatory response that begins within the conjunctiva, and with time spreads to the cornea. On a cellular level, the inflammation is composed of brown melanin pigment, red blood vessels, and pink scar tissue.

Table 4-2 Test Guidelines for *In Vivo* Ocular Irritation Test Methods

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Evaluate existing animal and human eye data	NA	Yes	Yes[1]	NS	Yes
Results from dermal irritation study	NA	Yes	Yes[1]	Yes	Yes
Perform SAR for eye irritation	NA	Yes	Yes[1]	NS	Yes
Screen for pH	NA	Yes	Yes[1]	Yes	Yes
Results from validated alternative ocular methods	NA	Yes	Yes[1]	Yes	Yes
Rabbit model/Number of rabbits					
Rabbit species and strain	Albino rabbit	Healthy young adult albino rabbits.	New Zealand White rabbit	Healthy adult albino rabbits recommended. Other mammalian species may be substituted with justification.	Healthy young adult albino rabbits.
Sex and weight	NS	NS	Sex NS; 2.0-3.0 kg	NS	NS
Screen for severe effects	NS	1 rabbit – further testing not required if substance produces corrosive or severe effects.	NS	1 rabbit – further testing not required if substance produces corrosive or severe effects.	1 rabbit – further testing not required if substance produces corrosive or severe effects.
Main test/confirmatory test	NS	Up to 2 additional rabbits, tested sequentially. if irreversible effects are suspected. Test discontinued, if severe effects occur in 2nd rabbit. Additional rabbits may be needed to confirm weak or moderate responses.	A minimum of 6 rabbits, and up to 18 rabbits for confirmatory tests.	≥ 3 rabbits	Up to 2 additional rabbits, tested sequentially, if irreversible effects are suspected. Test discontinued if severe effects occur in 2nd rabbit.

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Test substance (amount and method of application)					
Liquids	0.1 mL	0.1 mL	0.1 mL	0.1 mL	0.1 mL
Solids, pastes, particulates	NS	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL or 100 mg
Aerosols	NS	Single burst of about 1 second sprayed at 10 cm.	NS	Single burst of about 1 second sprayed at 10 cm.	Single burst of about 1 second sprayed at 10 cm.
Pump sprays	NS		NS	0.1 mL	Should not be used for instilling substances directly into eye.
Application of test substance	Test substance is placed in the conjunctival sac.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.
Use of anesthetics prior to instillation of test substance	NS	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Local anesthetic may be used prior to instillation of test substance.	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Anesthetic may be used after 24 hours if it does not influence response of the eye to irritants.
Observation					
Observation Period	At least 48 hours. Extended if irritation persists.	At least 72 hours, except when rabbit shows severe pain or distress, or early severe/corrosive effects, upon which the rabbit is humanely killed. Otherwise, sufficient to evaluate reversibility or irreversibility within 21 days.	At least 72 hours. Extended if necessary.	At least 72 hours, but not more than 21 days. Should be sufficient enough to evaluate the reversibility or irreversibility of effects within a 21-day period.	At least 72 hours, except when rabbit shows severe pain or distress, or early severe/corrosive effects, upon which the rabbit is humanely killed. Can be extended up to 21 days if effects persist.

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Examination times after treatment	1, 24, 48 hours, and 4, 7 days.	1, 24, 48, 72 hours, 7, 14, 21 days.	24, 48, 72 hours, and 7 days.	1, 24, 48, and 72 hours. Extended up to 21 days to assess reversibility.	1, 24, 48, and 72 hours. Can be extended up to 21 days. Observations of mild to moderate lesions until they clear or for 21 days. Observations at 7,14 and 21 days to determine reversibility.
Observation aids	NS	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.
Irrigation					
Washout	NS	Generally, eyes may not be washed until after 24 hours post-treatment, except for solids, which may be removed with saline or water after 1 hour.	After 24 hours post-treatment, eyes may be washed with a sodium chloride solution.	After 24 hours post-treatment, eyes may be washed with water to show whether washing palliates or exacerbates irritation.	Generally, eyes may not be washed until after 24 hours pos-treatment, except for solids, which may be removed with saline or water after 1 hour.

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Additional testing to determine effects of timely irrigation	NS	Not recommended unless scientifically justified.	NS	Indicated when substances are shown to be irritating. At 30 seconds after exposure, the eyes are washed with water for 30 seconds	Possibility of washing out in case of immediate corrosive or irritating effects. Use of satellite group to investigate influence of washing is not recommended unless scientifically justified.

Abbreviations: CPSC = U.S. Consumer Product Safety Commission; EEC = European Economic Commission; EPA = U.S. Environmental Protection Agency; FDA = U.S. Food and Drug Administration; FIFRA = Federal Insecticide, Fungicide, and Rodenticide Act; NA = Not applicable; NS = Not specified; OECD = Organization for Economic Cooperation and Development; OPPTS = Office of Prevention, Pesticide, and Toxic Substances; OSHA = U.S. Occupational Safety and Health Administration; SAR = Structure activity relationships; TG = Test guideline; TSCA = Toxic Substances Control Act.

[1] Use of this information is not provided in the regulations cited, but in the CPSC Animal Testing Policy guideline (CPSC 1984) states that prior human experience, literature sources which record prior animal testing or limited human tests, and expert opinion may be used in making appropriate hazard determinations.

The FHSA guideline states that a test substance is considered an eye irritant if four or more of six rabbits have positive ocular scores in nonirrigated eyes within 72 hours after instillation of the test substance (CPSC 2003). A positive score is defined by corneal opacity or iritis scores of ≥ 1, or conjunctival redness or chemosis scores of ≥ 2. In addition, if only one of the six rabbits shows ocular effects within 72 hours, the test substance in considered nonirritating to the eye. If two or three rabbits have positive ocular scores, the test is repeated in a second group of six rabbits. Then, if the criteria for an ocular irritant for the second test (three or more positive rabbits) or a nonirritant (0 positive rabbits) are met, a classification is made. However, if only one or two rabbits have positive scores in the second test, the test is repeated a third and final time. If one or more rabbits have positive ocular scores in the third test, the test substance is classified as an ocular irritant. If none of the rabbits have positive ocular scores in the third test, the test substance is classified as a nonirritant (CPSC 2003).

The EPA classification guideline considers the kinds of ocular effects produced in the *in vivo* rabbit eye test, as well as the reversibility and the severity of the effects (EPA 1996). However, unlike the FSHA system, incidence is not considered, as classification is based on the rabbit that exhibits the most severe response in a group of three or more rabbits. Data from all observation times are used for EPA classification. Corneal opacity or iritis scores of

≥ 1, or conjunctival redness or chemosis scores of ≥ 2 define a positive score. EPA labeling regulations also require an assessment of the reversibility of positive scores. If a positive score persists for > 21 days, the substance is classified as a Category I eye irritant, which is defined as "corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for > 21 days." Substances that cause positive corneal opacity, iritis, or conjunctival scores that clear in 8 to 21 days are designated as Category II eye irritants. If positive scores induced by a substance clear within 7 days, the substance is labeled Category III. A minimal effect (i.e., inconsequential or complete lack of irritation) or an effect that clears within 24 hours of application is designated as Category IV.

In the current EU classification system for eye irritation, risk phrases are assigned based on whether (a) two or more of three rabbits exhibit a positive score, averaged across the 24-, 48- and 72-hour observation times, or (b) the score of four or more rabbits, averaged across the 24-, 48-, and 72-hour observation times, for each ocular lesion that falls within or above certain ranges of scores (**Table 1-2**) (EU 2001). Hazard classification in the EU system corresponds to the following risk phrases: (1) R36 denotes "Irritating to eyes"; (2) R41 denotes "Risk of serious damage to the eyes." An *in vivo* rabbit eye study that results in (1) a mean corneal opacity score ≥ 3; (2) a mean iris score of 2 in two or more of three rabbits; (3) an overall mean corneal opacity ≥ 3; or (4) a mean iris score ≥ 1.5 in four or more rabbits, would be assigned the R41 risk phrase. Additionally, if a positive score persists to ≥ 21 days, the substance is assigned the R41 risk phrase. Criteria for assigning the risk phrase R36 are provided in detail in **Table 1-2**.

The GHS for the classification and labeling of hazardous chemicals (UN 2003) is an initiative developed through the cooperative efforts of the International Labour Office, the OECD, and the UN to promote an internationally-harmonized approach for classifying chemicals according to their health hazards. For the purpose of harmonizing classification of ocular irritants, the UN adopted an approach put forth by the OECD in its *Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods* (OECD 1996). A tiered testing and evaluation strategy using available data from dermal irritation studies, data from validated alternative toxicological methods, knowledge of structure activity relationships, and screening for pH extremes (≤ 2 or ≥ 11.5; considering acid or alkaline reserve) has been proposed (UN 2003). In addition, a single harmonized hazard category is proposed for irreversible effects on the eye/serious damage to eye (Category 1). Irreversible effects according to the GHS system include grade 4 corneal lesions at any time during the *in vivo* test, positive responses on day 21 (e.g., score > 0 for any endpoint evaluated), and cases where two or more of three rabbits exhibit a mean score (24, 48, 72 hours) for corneal opacity ≥ 3 and/or iritis > 1.5. A single harmonized hazard category, Category 2, is proposed for reversible effects on the eye; however, for regulatory authorities that prefer to distinguish irritants in this group, subcategories have been developed based on whether effects reverse within 7 or 21 days. Category 2A is defined as an eye irritant with effects that fully reverse within 21 days. Category 2B is considered mildly irritating to the eyes, and is designated for substances whose effects reverse fully within 7 days. Reversible effects include positive responses in two or more of three rabbits, where the mean score (24, 48, 72 hours) for corneal opacity or iritis ≥ 1 (but < 3

or < 1.5, respectively), or conjunctival redness or chemosis ≥ 2. Additional details on the GHS classification system are provided in **Section 4.3**.

4.2 Detailed Reference Data Used to Assess *In Vitro* Test Method Accuracy

The IRE studies evaluated in this document include *in vivo* reference data generated using the basic procedures described above for the *in vivo* rabbit eye test method.

For the EC/HO validation study (Balls et al. (1995), MMAS were calculated for the 59 studies from existing and concurrently run *in vivo* studies, all of which were performed according to OECD TG 405 (OECD 2002) and following GLP guidelines. The data were generated since 1981 and met the following criteria.

- Normally used at least three New Zealand White rabbits tested at the same time.
- 0.1 mL or the equivalent weight of substance was instilled into the conjunctival sac.
- Anesthesia was not used.
- Observations were made at least at 1, 2, and 3 days after instillation.

The MMAS were calculated for each test substance. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances are available in the ECETOC Reference Chemicals data bank (ECETOC 1998). These substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For the CEC (1991) study, *in vivo* irritancy data was obtained from historical data on 21 chemically-diverse test substances and an irritancy classification assigned, to the extent possible, according to Directive 83/467/EEC, Part II (B) of Appendix 6, Dangerous Substances Directive, 5th Adaptation, using NI (nonirritating), R36 (irritating), or R41 (severely irritating, serious risk to eyes). A total of three substances were classified as either R36 and R41 (1), or NI and R36 (2), based on different results in different laboratories. However, the greater level of irritant classification was assigned for the accuracy analysis. The rationale for use of the 21 test substances was based on inclusion of as much chemical (e.g., acid, base, inorganic and organic salt, substituted benzene, heterocycle, surfactant) and product (e.g., pesticide, detergent, antimicrobial, solvent) diversity. It was also important to have test substances covering the complete range of irritancy from nonirritant (10), to irritant (4), and severe irritant (7).

For the CTFA study, data were obtained from a modified Draize eye test. Details of the protocol are provided in Gettings et al. (1991, 1994, 1996). Six rabbits (three male, three female) were used for each test substance. The right eye of each rabbit was anesthetized prior to instillation of 0.1 mL of test substance into the conjunctival sac. Ocular irritation was evaluated at 1 hour, and at 1, 2, 3, 4 and 7 days. If irritation persisted, ocular responses were observed at 7-day intervals up to a maximum of 21 days. MAS were determined according to Williams et al. (1982). Data were classified according to the scheme proposed by the FHSA (1988). MAS, maximum average total scores for each endpoint (i.e., cornea,

iris, conjunctiva), number of positive responses, maximum day to clear, and FHSA categories are reported in the papers for all the tested substances. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances were provided by the CTFA. The substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For Guerriero et al. (2002, 2004) studies, data were obtained on 30 test substances (pharmaceutical process materials), respectively, from an *in vivo* Draize rabbit eye test performed approximately at the same time as the *in vitro* IRE test method. The *in vivo* data on the remaining 14 substances was obtained from historical data (ECETOC 1998). Since individual rabbit eye test data was kindly provided by Frederick Guerriero and GlaxoSmithKline, the *in vivo* data was classified according to the GHS (UN 2003), EPA (1996) and EU regulatory classification system (EU 2001) for comparison with *in vitro* data obtained using the IRE test method. The individual animal *in vivo* rabbit eye data that could be obtained for substances tested using IRE are provided in **Appendix D**.

4.3 *In Vivo* Classification Criteria Used for BRD Analysis

The *in vivo* rabbit eye database used to conduct a retrospective analysis of the accuracy of the IRE test method includes studies that were conducted using from one to six rabbits. However, some of the *in vivo* classification systems considered for the accuracy analyses are currently devised to be applied to studies using no more than three rabbits. Thus, to maximize the amount of data used for the evaluation of IRE, as well as for the three other *in vitro* test methods (ICE, BCOP, HET-CAM) being evaluated, the decision criteria for each classification system were expanded to include studies that used more than three rabbits in their evaluation.

All classification systems require the scoring of rabbits using the Draize scoring system (see **Table 4-1**). Scoring of rabbits occurs until the effect is cleared, but usually not beyond 21 days after the substance is applied to the eye of the rabbit. In order for a substance to be included in the accuracy evaluations in this BRD, four criteria must apply. These criteria were:

- At least three rabbits were tested in the study, unless a severe effect (e.g., corrosion of the cornea) was noted in a single rabbit. In such cases, substance classification could proceed based on the effects observed in less than three rabbits.
- A volume of 0.1 mL or 0.1 g was tested in each rabbit. A study in which a lower quantity was applied to the eye was accepted for substance classification, provided that a severe effect (e.g., corrosion of the cornea, lesion persistence) was observed in a rabbit.
- Observations of the eye must have been made, at minimum, at 24, 48, and 72 hours following test substance application if no severe effect was observed.
- Observations of the eye must have been made until reversibility was assessed, typically meaning that all endpoint scores were cleared. Results from a study

terminated early were not used, unless the reason for the early termination was documented.

If any of the above criteria were not fulfilled, then the data for that substance were not used for the accuracy analyses.

4.3.1 GHS Classification Rules Used for BRD Analysis

The classification of substances using the GHS classification system (UN 2003) was conducted sequentially. Initially, each rabbit tested was classified into one of four categories (Category 1, Category 2A, Category 2B, and nonirritant) based on the criteria outlined in **Table 4-3**. The criteria provided in this table are identical to those described in the GHS classification and labeling manual (UN 2003). Once all rabbits were categorized, the substance classification was determined based on the proportion of rabbits with a single irritancy category.

Table 4-3 Criteria for Classification of Rabbits According to the GHS Classification System

GHS Category	Rabbit Criteria Necessary for Classification
Category 1	Group A: - Effects in the cornea, iris, or conjunctiva that were not expected to reverse or did not fully reverse[1] within the observation period of 21 days, or - A corneal opacity score of 4 at any time during the test Group B: - Rabbit with mean scores (average of the scores on day 1, 2, and 3) for opacity ≥ 3 and/or iritis ≥ 1.5
Category 2A	- Rabbit with mean scores (rabbit values are averaged across observation days 1, 2, and 3) for one of more of the following: Iritis ≥ 1 but < 1.5 Corneal opacity ≥ 1 but < 3 Redness ≥ 2 Chemosis ≥ 2 and the effects fully reverse within 21 days
Category 2B	- Rabbit with mean scores (rabbit values are averaged across observation days 1, 2, and 3) for one of more of the following: Iritis ≥ 1 but < 1.5 Corneal opacity ≥ 1 but < 3 Redness ≥ 2 Chemosis ≥ 2 and the effect fully reversed within 7 days
Nonirritant	Rabbit mean scores fall below threshold values for Category 1, 2A, and 2B

Abbreviations: GHS = United Nations (UN) Globally Harmonized System.
[1]Full reversal of the effects was defined as corneal opacity, iritis, redness, and chemosis = 0.

After each rabbit was categorized, the ocular irritancy potential of the substance was determined. As shown in **Table 4-4**, substance classification depended on the proportion of

rabbits that produced the same response. As noted above, if a substance was tested in more than three rabbits, decision criteria were expanded. Generally, the proportionality needed for classification was maintained (e.g., one out of three or two out of six rabbits were required for classification for most categories). However, in some cases, additional classification rules were necessary to include the available data. These additional rules are distinguished by italicized text in **Table 4-4**.

Table 4-4 Criteria for Classification of Substances According to the GHS Classification System (Modified from UN 2003)

GHS Category	Criteria Necessary for Substance Classification
Category 1	1. At least 1 of 3 rabbits or 2 of 6 rabbits classified as Category 1, Group A 2. *One of 6 rabbits classified as Category 1, Group A and at least 1 of 6 rabbits classified as Category 1, Group B* 3. At least 2 of 3 rabbits or 4 of 6 rabbits classified as Category 1, Group B
Category 2A	1. At least 2 of 3 rabbits or 4 of 6 rabbits classified as Category 2A 2. *One of 3 (2 of 6) rabbits classified as Category 2A and 1 of 3 (2 of 6) rabbits classified as Category 2B*
Category 2B	At least 2 of 3 rabbits or 4 of 6 rabbits classified as Category 2B
Nonirritant	At least 2 of 3 rabbits or 4 of 6 rabbits classified as nonirritant

Abbreviations: GHS = United Nations (UN) Globally Harmonized System.
Italicized text indicates rules that were developed to include additional data.

If an unequivocal substance classification could not be made due to the response pattern of the tested rabbits for a substance (e.g., one rabbit classified as Category 1, Group B; two rabbits classified as Category 2B; three rabbits classified as nonirritant), the data were not used in the analysis.

4.3.2 EPA Classification Rules Used for BRD Analysis

The classification of substances using the EPA classification system (EPA 1996) was conducted sequentially. Initially, each rabbit was classified into one of four categories (Category I to Category IV) (**Table 4-5**.)

Substance classification was dependent upon the most severe category observed among the tested rabbits. Thus, a single rabbit in a more severe category than the remaining animals would lead to classification of the substance into that category (i.e., classification of a substance was not based on the majority classification among rabbits tested).

Table 4-5 Criteria for Classification of Rabbits According to the EPA Classification System (EPA 1996)

EPA Category	Criteria for Rabbit Classification
Category I	- Corrosive, corneal involvement or irritation (iris or cornea score ≥ 1 or redness or chemosis ≥ 2) persisting more than 21 days or - Corneal effects that are not expected to reverse by 21 days
Category II	- Corneal involvement of irritation clearing[1] in 8 to 21 days
Category III	- Corneal involvement of irritation clearing in 7 days or less
Category IV	- Minimal or no effects clearing in less than 24 hours

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]For the purposes of this analysis, clearing was defined as iritis or corneal opacity score < 1 and redness or chemosis score < 2.

4.3.3 EU Classification Rules Used for BRD Analysis

Substance classification using the EU classification system was conducted sequentially (EU 2001). While average Draize scores are used for classification, the calculation of average scores for the EU system depends on the number of rabbits tested in a study (see **Section 4.1.3** for additional details). Depending on the number of rabbits tested, the appropriate average scores were calculated, then the substance was classified based on the number of rabbits with a minimal positive average (for studies that used three rabbits) or the overall average (for studies that used more than three rabbits). The criteria used for substance classification are in **Table 4-6**.

Table 4-6 Criteria for Classification of Substances According to the EU Classification System (EU 2004)

EU Category	Three Rabbits Tested	Greater than Three Rabbits Tested
R41	Two or more rabbits where the average rabbit Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 Iritis = 2 Or At least one rabbit (at end of observation period), where the effect has not reversed[1]	Overall mean rabbit Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 or Iritis > 1.5 Or At least one rabbit (at end of observation period), where the effect has not reversed
R36	Two or more rabbits where the average rabbit Draize scores over Days 1, 2, and 3 were: 2 ≤ Opacity < 3 1 ≤ Iritis < 2 Redness ≥ 2.5 Chemosis ≥ 2	Overall mean rabbit Draize scores over Days 1, 2, and 3 were: 2 ≤ Opacity < 3 1 ≤ Iritis < 1.5 Redness ≥ 2.5 Chemosis ≥ 2

Abbreviation: EU = European Union.
[1]Full reversal of the effects was defined as corneal opacity, chemosis, redness, or iritis = 0.

4.4 Availability of Original Records for the *In Vivo* Reference Data

Much of the published data on the prediction of ocular irritancy potential for test chemicals using the *in vivo* test method was limited to average score data, average animal data, or irritancy classification. An attempt to obtain the original records and/or compiled reports for the *in vivo* reference data was made. However, much of the information and data was either not readily available or not provided in the requested format.

4.5 *In Vivo* Data Quality

Ideally, all data supporting the validity of a test method should be obtained and reported from studies conducted in accordance with GLP guidelines, which are nationally and internationally recognized rules designed to produce high-quality laboratory records (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines provide an internationally standardized approach for the conduct of studies, reporting requirements, archival of study data and records, and information about the test protocol, in order to ensure the integrity, reliability, and accountability of a study.

The extent to which the *in vivo* rabbit eye studies, which were used to provide the comparative data in the published IRE validation studies, were compliant with GLP guidelines is based on the information provided in the published reports. Although an attempt was made to obtain the original study records, such records could not be obtained. Based on the available information, the *in vivo* rabbit data used in three of these four reports (Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004) were obtained in compliance with GLP guidelines.

4.6 Availability and Use of Toxicity Information from the Species of Interest

Due to the possibility of irreversible eye injury that could impair vision or cause blindness, human ocular irritancy studies are not routinely conducted. The only exceptions are for products intended for actual human eye use (e.g., contact lens solutions, ophthalmic pharmaceuticals) or cosmetic/personal care products that are known not to cause more than minimal to mild responses in rabbits. Bruner et al. (1998) and Cater et al. (2004) reported on studies conducted in humans of cosmetic and surfactant-based personal care formulations. However, all of the substances tested were classified as mild irritants or nonirritants and corresponding IRE tests were not conducted. Procter & Gamble provided information from human exposures to three consumer-product formulations as a comparison to the EU ocular toxicity classifications (EU 2001), assigned based on results from the low volume eye test (LVET). However, because all three of these formulations were classified as nonirritants or mild irritants, based on results obtained in LVET, evaluation of the accuracy of the IRE test method for identifying ocular corrosives and severe irritants in humans is not possible.

It may be possible to consider accidental human exposure injury data to identify substances or products capable of producing severe or irreversible eye injuries in humans. These data could then be compared with available rabbit data and hazard classifications to determine if the potential for severe human effects was not predicted by the rabbit test. A query to all ICCVAM regulatory agencies did not yield any substances or products known to produce

severe or irreversible human eye injury not predicted by the rabbit test. However, this lack of such substances or products must be considered in light of the surveillance and reporting systems for such injuries.

Several U.S. Federal agencies (OSHA, CPSC, and the National Institute for Occupational Safety and Health [NIOSH]) were contacted for data resulting from accidental human exposures. Based on emergency department reports for work related eye-injuries, NIOSH estimated that approximately 39,200 chemical-related eye injuries occurred in 1998, (NIOSH 2004). Approximately 10,000 of these cases were attributed to an unidentified or unspecified chemical. Additional cases (< 2500 each) were reported for injuries related to specific chemicals or chemical/product classes, which included[2]:

- acids (unspecified)
- adhesives/glues
- cement/mortar mix
- chlorine/chlorine bleach
- cleaning/polishing agents
- detergents/shampoos
- disinfectants
- drain/oven cleaners
- gasoline/jet fuels/diesel fuel
- hydrochloric acid
- nonchlorine bleach
- paint removers/thinners
- paints
- soaps
- sodium hydroxide, potassium hydroxide, and potassium carbonate
- solvents/degreasers
- sulfuric acid

However, for the product classes listed above, specific information on which products were involved are not available. No human data were provided for any of these substances, nor were details of the types of ocular injuries sustained described.

In addition, according to U.S. Bureau of Labor Statistics (BLS), 6303 lost workdays attributable to occupational eye injuries from chemical exposures were reported in 2002 (BLS 2004). These numbers may be underestimates of the actual incidence, since not all employers are required to report such injuries. The specifics of the exposures are not provided.

Without more detail about the specific nature of the substances and exposure conditions, these types of accidental human exposure injury data are not useful for evaluating the accuracy of the IRE test method for predicting human ocular hazard.

[2] These specific chemicals or chemical/product classes are listed in alphabetic order; actual numbers of cases for each specific chemical or chemical/product class are not provided.

4.7 Information About Accuracy and Reliability of the *In Vivo* Test Method

4.7.1 Information About the Accuracy of the *In Vivo* Test Method

Accuracy of the *in vivo* test method would ideally be assessed by comparison of ocular effects observed in the rabbit to those effects produced in humans. A review of the literature indicates that there are few studies in which rabbit and human responses have been carefully compared under controlled conditions to assess the accuracy of the *in vivo* test method. Therefore, most studies conduct retrospective evaluations and comparisons of responses between humans and rabbits. A review indicates that a number of studies show that responses to mild to moderate irritants were generally similar between rabbits and humans (Lewin and Guillery 1913; Suker 1913; Leopold 1945; Carpenter and Smyth 1946; McLaughlin 1946; Nakano 1958; Barkman 1969; Grant 1974). A review of these studies can be found in McDonald et al. (1987). For a severe irritant, Grant (1974) and Butscher (1953) showed that accidental exposure to neat thioglycolic acid produced similar responses in humans and rabbits.

In comparison, there have been studies where the responses to ocular irritants differ between humans and rabbits. In some cases, test substances produced more severe responses in humans than in rabbits (Lewin and Guillery 1913; Gartner 1944; Estable 1948; Marsh and Maurice 1971; Grant 1974). For example, Marsh and Maurice (1971) evaluated the effects of a 1% concentration of nonionic detergents in humans. The most severe symptoms (e.g., blurred vision and halos with corneal epithelial bedewing; most effects disappearing with 24 hours) were associated with 1% Brij 58. Comparatively, Grant (1974) showed that, in general, nonionic detergents did not damage the rabbit eye, even when tested at higher concentrations. Additional examples of disparate effects between humans and rabbits are summarized in McDonald et al. (1987). Studies with some soaps and surfactants indicated that more severe responses were produced in rabbits than in humans (Calabrese 1983). Differences between humans and rabbits with respect to anatomy and physiology, pain thresholds, exposure parameters (e.g., volume administered, length of exposure period), and potential differences in mechanism of action of test substances have been proposed as reasons for the discordant responses.

4.7.2 Information About the Reliability of the *In Vivo* Test Method

Based largely on the protocol of Draize et al. (1944), the original regulatory requirements for eye irritation testing mandated the use of at least six rabbits. In recognition of animal welfare concerns, several evaluations were conducted to assess the reliability of the test method and the consequences of reducing the number of rabbits per test from six to as few as two (DeSousa et al. 1984; Solti and Freeman 1988; Talsma et al. 1988; Springer et al. 1993; Dalbey et al. 1993; Berdasco et al. 1996). With the exception of Dalbey et al. (1993), each study concluded that reducing the number of rabbits from six to three would not have an unacceptable reduction on the predictivity of ocular irritancy classification/categorization. Analyses were performed using MAS, internal irritancy classification schemes, and/or regulatory classification schemes as endpoints for comparison. Several of these studies (DeSousa et al. 1984; Talsma et al. 1988; Dalbey et al. 1993) revealed that correlations between three-rabbit and six-rabbit classifications were the highest among substances classified on the extreme ends of the irritancy range (i.e., nonirritants and severe irritants).

These studies noted that the majority of variability among rabbit responses was observed among substances classified in the middle range of irritation (i.e., mild and moderate irritants). Accordingly, Dalbey et al. (1993) concluded that the observed variability in the middle range of irritation justified the continued routine use of six rabbits. However, based primarily on the results of these evaluations, the EPA (EPA 1998), EU (EU 2001), and the OECD (in revised TG 405), recommended the use of a maximum of three rabbits, although additional rabbits could be tested under certain circumstances (e.g., to confirm weak or moderate responses).

To further address the reliability of the rabbit eye test, ICCVAM and NICEATM used the available *in vivo* data to estimate the likelihood of underclassifying a positive substance or overclassifying a negative substance in the current one to three rabbit sequential test. Data from Draize eye testing using three to six rabbits was obtained for approximately 900 substances from U.S. Federal regulatory agencies, published studies, and scientists and organizations. Ocular irritation categories were assigned for each substance based on the GHS classification system (UN 2003). Using the available *in vivo* rabbit eye test database of 181 severe irritant studies, the distribution of individual rabbit responses within each severity class was used to estimate the likelihood of under- and over-classification rates for a sequential one to three rabbits testing strategy. Based on three different assumptions about the variability in response among substances within each classification category, the estimated underclassification rate for corrosives/severe irritants (GHS Category 1) as nonsevere irritants (GHS Category 2) or nonirritants ranged from 4% to 13%. Analyses based on physical form of the test substance suggested that underclassification rates for solids were lower than liquids (2.9% to 8.3% vs. 5.4% to 15.8%, respectively), although these differences are not statistically significant. Estimated underclassification rates were higher when a corrosive/severe irritant classification was based solely on persistent lesions present at observation day 21. By chemical class, carboxylic acids had the highest underclassification rate (16.64%). Overclassification rates of substances as corrosive/severe irritants, based on 596 studies, were estimated to be 7% to 8% for Category 2A substances, 1% for Category 2B substances, and 0% for nonirritants.

[This Page Intentionally Left Blank]

5.0 IRE TEST METHOD DATA AND RESULTS

5.1 Description of the IRE Test Method Protocols Used to Generate Data

The ocular irritancy of a wide variety of test substances was determined in four reports using the IRE test method and compared to results obtained in the *in vivo* Draize rabbit eye irritation test (CEC 1991; Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). Individual eye test data was obtained for all of these reports. In these reports, the protocols used to generate the *in vitro* IRE test method data varied in the number of endpoints measured and in the time the measurements were taken. For example, Gettings et al. (1996) only evaluated corneal swelling, whereas Balls et al. (1995) evaluated corneal opacity and corneal swelling. In the CEC (1991) study, corneal opacity, corneal swelling, and fluorescein retention were reported. In Guerriero et al. (2004), corneal opacity, corneal swelling, and fluorescein penetration with an assessment of epithelial integrity were evaluated. Variations in the protocols used to generate IRE test data in these studies for specific endpoints are shown in **Appendix A**.

5.1.1 CEC Collaborative Study (1991)

The "Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test" was sponsored by the Commission of the European Communities (CEC) and published in 1991. One study in this report evaluated 21 chemically diverse test substances using the IRE test method conducted by three independent laboratories. The IRE data was obtained using a standard IRE protocol (Burton et al. 1981) for the measurement of corneal swelling (percent increase in corneal thickness relative to pretreatment value) and assessment of corneal opacity (score of 0 to 4) over a period of 30 minutes to four hours, but fluorescein retention (score of 0 to 4) at 30 and/or 240 minutes was added. However, each laboratory used an independent irritancy prediction model to evaluate overall *in vitro* severity based on a scale of I (non-irritating or no EU label) to IV (comparable to EU R41 label) (EU 2001). The final *in vitro* irritancy rating (A; mild to D; most severe) was then compared to published *in vivo* ocular data ranked according to severity of injury using the EU (EU 2001) classification system.

5.1.2 Balls et al. (1995)

In the EC/HO International validation study, 59 test substances were evaluated using the IRE test method. In this study, data from the IRE test method was generated by four separate laboratories and consisted of measurement of corneal opacity and corneal swelling at 1 and 4 hours. Corneal opacity was evaluated using the Draize scoring system (scale of 0-4). Corneal swelling was determined by measurement of pretreatment corneal thickness and calculation of the relative increase in corneal thickness at each time point, but only the one and four hour values were used in the analysis. The effects of the test substance on epithelial integrity and the degree of fluorescein penetration were not reported in this validation study. However, while the EC/HO study only acquired data from two endpoints (corneal opacity and swelling), two authors recognized that the use of additional endpoints (i.e., fluorescein penetration and assessment of endothelial integrity) might significantly increase the accuracy of the irritancy potential assigned to a test substance. A negative, untreated control was used in each experiment, but a positive control was not included.

5.1.3 Gettings et al. (1996)

The CTFA conducted an evaluation of *in vitro* alternatives comparing 41 *in vitro* test endpoints including the IRE test method to the Draize rabbit eye irritation test. In this study, 25 surfactant-based formulations were tested representing cosmetic and personal care products. The endpoints evaluated were corneal opacity and corneal swelling (0.5, 1, 2, 3, and 4 hours following application of the test substance). Fluorescein penetration was evaluated one hour after application of the test substance. Histological assessment of the corneal tissue was determined on tissue obtained four hours after application. Corneal opacity was measured macroscopically and by slit-lamp examination. Corneal thickness was measured before application of the test substance and at each time point afterwards. The percentage increase in corneal thickness was calculated relative to the pretreatment value and expressed as corneal swelling. The mean swelling of each treatment group was then calculated and compared to *in vivo* data. The data were transformed into four groups based on "no significant swelling" to "maximal response" and compared also to data obtained in the *in vivo* rabbit eye test.

5.1.4 Guerriero et al. (2004)

The IRE test method protocol used by Guerriero et al. (2002, 2004) to compare *in vitro* and *in vivo* test results is outlined in **Appendix A** of this BRD. Corneal opacity and area of opacity were assessed macroscopically and by slit-lamp examination at 0.5, 1, 2, 3 and 4 hours after application of the test substance. Corneal swelling was calculated as the percentage increase in corneal thickness at each time point relative to a pre-application value. Fluorescein uptake and assessment of epithelial integrity were determined at four hours. The decision criteria described in **Appendix A** were used to classify a test substance as either a severe irritant or a nonsevere irritant.

5.2 Data Obtained to Evaluate the Accuracy and Reliability of the IRE Test Method

NICEATM staff requested original data from IRE test method studies. A *Federal Register* notice (Vol. 69, No. 57, pp. 13859-13861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original IRE test method data, was published on March 24, 2004. In addition, authors of published IRE studies were contacted to request original IRE test method data for the respective publications. As a result of these efforts, data for the 59 substances evaluated in the Balls et al. (1995) study using the IRE test method were obtained in electronic format. In addition, Frederick Guerriero at GlaxoSmithKline kindly provided the IRE test method data and the individual *in vivo* rabbit eye test data from the Guerriero et al. (2004) study. In addition, GlaxoSmithKline generously provided the actual chemical names and MSDS sheets of the 30 GSK test substances used in the study that were originally reported as numbered (i.e., 1 to 30) generic substitutions (e.g., substituted aniline, substituted pyridine, aromatic acetanilide). No other IRE test method data was received. Therefore, this evaluation is based, in part, on data obtained from the published literature. Furthermore, given the lack of availability of the original records for these studies, the testing laboratory's summary judgment regarding the outcome of each study cannot be evaluated. The availability of notebooks or other material containing original data is unknown.

5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data

Statistical approaches used to analyze the *in vitro* data obtained in the IRE test method protocol were discussed in **Section 2.2.12**. Evaluation of the accuracy and reliability of the *in vitro* data obtained using the IRE test method in comparison to *in vivo* data has been approached differently by various authors. In all of these reports, none of the data were compared to the GHS (UN 2003), EPA (1996), or EU (EU 2001) regulatory classification systems. Therefore, as discussed in **Sections 6.0** and **7.0**, the results of these studies were reanalyzed to assess the accuracy and reliability of the test method, when compared to these classification schemes.

5.3.1 CEC Collaborative Study (1991)

In this study of 21 test substances, *in vitro* irritation was correlated only to the EU-based irritation classification (EU 2001). It is not clear what, if any, statistical methods were used to compare these scores. No correlation coefficients were provided in the report. However, raw data for the IRE studies were provided for each test substance with an *in vivo* EU classification based on literature evaluation. This data was sufficient to permit an analysis of performance based on EU classification only.

5.3.2 Balls et al. (1995)

In this study, 59 test substances were ranked by degree of irritancy according to the mean MMAS values and the standard deviations from the mean according to ECETOC (1992). The data was arranged in a matrix containing the identity of the test substance, the *in vivo* eye irritation data and each alternative method endpoint. Data analysis was carried out on the full data set and on subsets based on solubility criteria. Correlation coefficients were determined to assess interlaboratory reproducibility of alternative data (27 test index scores) and the relationship between *in vivo* eye irritation potential versus the endpoint for each alternative method. To conduct the proposed accuracy analysis, the *in vivo* data were reclassified using the GHS (UN 2003), EPA (EPA 1996) and EU (EU 2001) irritancy classification systems.

5.3.3 Gettings et al. (1996)

In this study of 25 surfactant-based formulations, statistical analysis was divided into three components. The distributional characteristics of the Draize test and *in vitro* test results were determined, concordance analysis was used to assess the extent of association of the Draize test and *in vitro* scores, and regression analysis was used to predict Draize scores (MAS values) from *in vitro* test results. The IRE results were compared to an *in vivo* irritation score based on the FHSA regulatory classification system in which test substances are classified as irritants or nonirritants without further separation. To conduct the proposed analysis, the CTFA *in vivo* data were reclassified according to the GHS (UN 2003), EPA (1996) and EU (EU 2001) regulatory classification systems.

5.3.4 Guerriero et al. (2004)

In this study of 44 test substances, *in vitro* irritation was correlated to EU-based irritation classification (EU 2001). It is not clear what, if any, statistical methods were used to compare these scores. No correlation coefficients were given. However, adequate information to assign at least one regulatory irritancy classification (EU [EU 2001]) to each

test substance was given in the publication; these classifications were used in an analysis of performance. Additional *in vivo* data from these studies were provided on request that permitted assignment of GHS (UN 2003) and EPA (EPA 1996) regulatory classifications and additional retrospective accuracy analyses.

5.4 Summary of Results

A summary of results used to evaluate test method accuracy is shown in **Appendix C**. This appendix, sorted by substance tested, provides the name of the substance tested, the Chemical Abstract Service Registry Number (CASRN), the concentration tested, the calculated score, the irritation classification of the test substance, and the literature source. No attempt was made to identify the source and purity of a test substance if the authors did not provide such information. If available, a CASRN was entered for each test substance. This identifier was obtained from various sources, including the publication and the National Library of Medicine's ChemID database. All substances with the same CASRN were listed under the same name, regardless of the synonym that was used in the original publication.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines and with the use of coded chemicals (OECD 1998; EPA 2003a; 2003b; FDA 2003). The data quality was evaluated by a review of the methods section in literature references only and in the submitted reports. The data quality presented in the reviewed literature references can only be evaluated by what was provided in the published reports. Based on the available information, the Balls et al. (1995) study, the Gettings et al. (1996) study, and the Guerriero et al. (2004) study were reportedly conducted according to GLP guidelines (see **Section 8.0**). As described in **Section 3.4**, based on the available information, coded chemicals were employed in the Balls et al. (1995) and Gettings et al. (1996) studies. The Guerriero et al. (2004) study was performed in a GLP-compliant laboratory in which the test substances are individually coded. In the CEC pilot collaborative study, there was no indication that the studies were conducted under GLP compliance, although at least two of the three testing laboratories are known to run GLP-compliant studies. Although test substances were numbered in these studies as indicated in **Section 3.4**, it is not known if the testing laboratories were blinded with respect to the numbering.

5.6 Description of "Lot-to-Lot" Consistency, Time Frame of Studies and Testing Laboratories

Ideally, the lot-to-lot consistency of test substances is evaluated to ensure that the same substance, with the same physicochemical properties, is being evaluated over the duration of the study. A description of the procedures used to evaluate and control the lot-to-lot consistency was described in three of the published reports (CEC 1991; Balls et al. 1995; Gettings et al. 1996), but was applicable to interlaboratory studies using the same test substance. For the studies described in this BRD, substances were only tested once in each study, and therefore, lot-to-lot consistency within a study was not applicable. No attempt

was made to review the original records and assess the procedures used to evaluate different batches of tested substances.

5.6.1 CEC Collaborative Study (1991)

In the CEC collaborative IRE study, each of the 21 test substances came from a single batch with purity greater than 97% and was supplied by the Aldrich Chemical Company Limited (UK) via the FRAME (Nottingham, UK).

The testing laboratories were TNO-CIVO Institutes (Zeist, Netherlands), Shell Research Ltd. (Sittingbourne, UK), and the Instituut voor Hygeine en Epidemiologie (IHE) (Brussels, Belgium).

5.6.2 Balls et al. (1995)

The *in vivo* reference test substances used in the EC/HO Study were readily available chemicals of high consistency and purity (90 to >99.5%) with known stability in storage over time. The substances were obtained from the same source when feasible. Otherwise, substances with specifications as close as possible to the original were used. Coding, labeling, dispensing, and storage procedures were tightly controlled to avoid confusion of test substances during preparation, storage and delivery.

Four laboratories were selected for each of the nine test methods. The laboratories chosen for the IRE test method were:

- ESL, Unilever Research, Sharnbrook, Bedford, United Kingdom (Dr. Lesley Earl) – *Lead Laboratory*
- Shell Research Ltd, Sittingbourne, Kent, United Kingdom (Dr. John Gardner)
- Zeneca CTL, Macclesfield, Cheshire, United Kingdom (Dr. Richard Lewis)
- IHE, Ministry of Public Health of Environment, Brussels, Belgium (Dr. GA Jacobs)

5.6.3 Gettings et al. (1996)

The test substances used in the CTFA validation study were individual generic formulations designed to represent a variety of product types. Test substances used in the CTFA validation study were carefully controlled with respect to documentation, coding, labeling, dispensing and transfer. Documentation included manufacturer, lot and/or batch number, amount of test substance, and condition upon receipt. The substances were prepared in bulk and dispensed using stringent requirements to maintain consistency for delivery to the individual testing laboratories.

5.6.4 Guerriero et al. (2004)

In the GSK/Safe-Pharm Laboratories study, information on the individual test substances with respect to lot-to-lot variation was not given. However, for 30 of the 44 pharmaceutical process materials used as test substances, the *in vivo* data and the *in vitro* data were obtained on the same chemicals within the same time frame. Substances were analyzed for quality control purposes and reasonable purity (> 90%) was attained, but the analytical data were not immediately available for all substances at the time the study was performed (Guerriero F, personal communication). With the exception of the ECETOC (1998) test substances,

identical lots and batch numbers of test substances were used for both the *in vitro* and *in vivo* studies, performed in the same laboratory.

The testing laboratories were GSK (United States and United Kingdom) and SafePharm Laboratories (Derbyshire, United Kingdom).

5.7 Availability of Data not Submitted for External Audit, If Requested

An attempt to obtain the original study records for the IRE data was made by NICEATM. The original study records were not readily available for any of the studies; thus, it appears unlikely that such data are available for an external audit.

6.0 IRE TEST METHOD ACCURACY

6.1 Accuracy of the IRE Test Method

A critical component of an ICCVAM evaluation of the validation status of a test method is an assessment of the accuracy of the proposed test method when compared to the current reference test method (ICCVAM 2003). This aspect of assay performance is typically evaluated by calculating:

- accuracy (concordance): the proportion of correct outcomes (positive and negative) of a test method
- sensitivity: the proportion of all positive substances that are classified as positive
- specificity: the proportion of all negative substances that are classified as negative
- positive predictivity: the proportion of correct positive responses among substances testing positive
- negative predictivity: the proportion of correct negative responses among substances testing negative
- false positive rate: the proportion of all negative substances that are falsely identified as positive
- false negative rate: the proportion of all positive substances that are falsely identified as negative.

The ability of the IRE test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS, EPA, and EU[1], was evaluated separately for each *in vitro-in vivo* comparative study (i.e., publication) reviewed in **Sections 4.0** and **5.0**. The three ocular hazard classification systems considered during this analysis use different classification systems and decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results (see **Section 4.0**). All three regulatory classification systems are based on individual animal data in terms of the magnitude of the response and, for the EPA (1996) and GHS (UN 2003), on the extent to which induced ocular lesions fail to reverse by day 21. Thus, to evaluate the accuracy of the IRE test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at different observation times was needed for each substance. However, these data were not consistently available in the reports considered, which limited the number of test results that could be used for assessment of test method accuracy. For example, the CEC (1991) collaborative study provided a EU ocular irritancy classification (i.e., R41, R36, nonirritant [EU 2001]) for the 21 substances tested but did not provide individual *in vivo* rabbit eye data, which precludes an accuracy analysis based on the GHS and EPA classification systems. Furthermore, most of the *in vivo* classifications used for the analyses presented in this section are based on the results of a single study. Unless otherwise indicated, variability in the *in vivo* classification is unknown.

[1] 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), Category I according to the EPA classification system (EPA 1996), or as R41 according to the EU classification system (EU 2001) (see **Section 1.0**).

In addition, the accuracy assessments conducted were based on IRE test data that were evaluated by investigators using different endpoints for evaluation and different decision criteria to classify the irritancy potential of test substances. As discussed in **Section 2.2.12**, some IRE studies were conducted using the Draize scoring system to evaluate corneal opacity with or without area of opacity evaluated. Some studies also included other endpoints, such as fluorescein retention or penetration and epithelial integrity (based on slit-lamp observations and/or histology). Furthermore, not all studies evaluated or reported data for all of the time points typically measured in IRE (i.e., 0.5, 1, 2, 3, 4 hours). For example, the CEC (1991) collaborative study reported corneal opacity, corneal swelling, and fluorescein retention at one and four hours. In the Balls et al. (1995) validation study, corneal opacity and corneal swelling were the only endpoints reported and were evaluated at one and four hours. In contrast, Gettings et al. (1996; REET 1) only reported the mean extent of corneal swelling across time (1 to 4 hours). Guerriero et al. (2004) reported maximal corneal opacity (opacity x area), maximal corneal swelling, fluorescein penetration (intensity x area) and assessment of epithelial integrity (i.e., 0.5, 1, 2, 3, and 4 hours). In this study, the decision criteria (Prediction Model) for identification of a severe irritant were based on exceeding cut-off values for any of these parameters (maximum corneal opacity ≥ 3; maximum corneal swelling ≥ 25%; maximum corneal fluorescein penetration ≥ 4; any pitting, stippling, mottling, sloughing, or ulceration of epithelium) (see **Appendix A)**. IRE data from each of these four studies were converted into an irritancy classification using the decision criteria outlined in Guerriero et al. (2004), since these were the only decision criteria that specifically were designed to detect severe ocular irritants (see **Appendix A)**.

A limitation of the available IRE data is that the numbers of endpoints used by various investigators differed. In the studies by Guerriero et al (2004), four different ocular endpoints were used. Comparatively, data from the other studies (e.g., Balls et al. 1995) were conducted with between one and three endpoints. In order to make use of all the available data, an "Expanded Data Set" was developed and evaluated. In this data set, any substance evaluated by any of the studies that would be classified as an *in vitro* severe irritant based on a positive result using any of the four ocular endpoints was identified as a corrosive or severe irritant and included into the database (CEC 1991; Balls et al. 1995; Gettings et al. 1996). Substances in these other studies that were not identified as ocular corrosives or severe irritants could not be used, since a positive result in any of the omitted endpoints might have resulted in a severe irritant classification. For example, in Gettings et al. (1996), only corneal swelling was measured. Substances that produced corneal swelling ≥ 25% were classified as severe irritants and were included in the "Expanded Data Set." However, a substance that did not produce ≥ 25% corneal swelling, might have produced a corneal opacity score, fluorescein penetration score, or damage to the epithelium that would have classified it as a severe irritant had those endpoints been evaluated.

In addition to the analysis for the "Expanded Data Set", an analysis based on a "Pooled Data Set" was conducted. Both performance analyses were included to increase the number of test substances evaluated given the limitations of each data set. The Expanded Data Set was used to identify ocular corrosives or severe irritants based on a positive response in any of the four ocular endpoints used by Guerriero et al. (2004); the decision criteria for the IRE test method used in the BRD performance analyses. However, there is bias associated with this data set,

because negative responses could not be included in studies where there were less than four ocular endpoints evaluated (e.g., Guerriero et al. 2004), because any omitted endpoint could have resulted in a positive response had it been tested and, therefore, only positive outcomes that met the decision criteria for any single endpoint could be included. A Pooled Data Set was included in the analysis that used all available data from the four studies and included negative responses. However, this data set is also limited in that it includes data with positive or negative outcomes from studies in which less than four ocular parameters were evaluated.

Using the classification systems discussed in **Section 5.0**, the *in vitro* irritancy potential of each substance was determined using data supplied in the published report or submitted in response to the *FR* Notice request for data (**Section 5.2**). For the "per study" accuracy analysis, two different types of analyses were used. In the first analysis, the IRE ocular irritancy potential of each substance in each report was determined (**Appendix C**). When the same substance was evaluated in multiple laboratories (see Balls et al. 1995 in **Appendix C**), the IRE ocular irritancy potential for each test was determined. Subsequently, based on the majority of ocular irritancy classification calls, an overall IRE ocular irritancy classification was assigned (e.g., if two tests classified a substance as a nonsevere irritant and three tests classified a substance as a severe irritant; the overall *in vitro* irritancy classification for the substance would be severe irritant). When there was an even number of different irritancy classifications for substances (e.g., two tests classified a substance as a nonsevere irritant and two tests classified a substance as a severe irritant), the more severe irritancy classification was used for the overall classification for the substance (severe irritant, in this case). Once the ocular irritancy potential classification was determined for each substance in a report, the ability of the IRE test method to identify ocular corrosives and severe irritants, as defined by the three different regulatory classification systems (EPA 1996, EU 2001, UN 2003), was determined for each report (**Appendix D**).

In the second analysis used in the "per study" evaluation, each irritancy classification obtained on the same substance tested in multiple laboratories was used separately to assess test method accuracy (i.e., results were not combined across multiple laboratories to develop an overall IRE ocular irritancy classification). The ability of the IRE test method to identify ocular corrosives and severe irritants, as defined by the three different classification systems, was then determined for reports where multiple results were available for the substances tested. This approach was applied to the CEC (1991) and the Balls et al. (1995) studies, the only reports that included multiple laboratory study data.

6.1.1 GHS Classification System: IRE Test Method Accuracy

Accuracy analyses using the GHS regulatory classification system (UN 2003)[2] were conducted on data obtained from three reports (Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). To the extent possible, severe ocular irritants were identified from the *in vitro* data in these reports using the Guerriero et al. (2004) IRE test method scoring system

[2] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify GHS (UN 2003) Category 1 irritants (i.e., severe irritants); substances classified as GHS Category 2A and 2B irritants were identified as nonsevere irritants.

described in **Section 6.1**. For example, two ocular parameters included in the recommended protocol, fluorescein penetration and assessment of epithelial integrity, were not assessed in the IRE studies by Balls et al. (1995) and Gettings et al. (1996). The GHS classification assigned to each test substance is shown in **Appendix D**. The performance characteristics (i.e., accuracy, sensitivity, specificity, positive predictivity, negative predictivity, false positive rate, and false negative rate) were determined for each of the three studies based on the available *in vivo* reference data for the substances tested in these studies (**Table 6-1**). Of the three studies, Balls et al. (1995) provided IRE data for substances tested in multiple laboratories; the first set of accuracy calculations for these studies in **Table 6-1** (n = 59) represents the results obtained using the consensus call for each test substance, while the second set of accuracy calculations for each study represents the results obtained when each independent test result from each laboratory was considered separately (n = 236).

6.1.1.1 *Balls et al. (1995)*

Based on the available *in vivo* rabbit eye data, 54 of the 59 test substances could be assigned a classification using the GHS system (UN 2003) (**Table 6-1**). The remaining five substances had inadequate *in vivo* data for assigning a classification according to the GHS system (UN 2003). Based on these 54 substances, the IRE test method had an accuracy of 54% (29/54), a sensitivity of 68% (15/22), a specificity of 44% (14/32), a false positive rate of 56% (18/32), and a false negative rate of 32% (7/22) (**Table 6-1**).

For Balls et al. (1995), using the second approach in which the result of each IRE test is considered separately and GHS classification was possible (n = 216/236), the IRE test method has an accuracy of 60% (130/216), a sensitivity of 72% (63/88), a specificity of 52% (67/128), a false positive rate of 48% (61/128) and a false negative rate of 28% (25/88) for identifying ocular corrosives and severe irritants according to the GHS system (UN 2003) (**Table 6-1**).

6.1.1.2 *Gettings et al. (1996)*

In this study, based on the provided *in vivo* rabbit eye test data, 24 of the 25 test substances could be classified according to the GHS system (UN 2003). Using these data, the IRE test method has an accuracy of 67% (16/24), a sensitivity of 63% (10/16), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 38% (6/16) (**Table 6-1**).

6.1.1.3 *Guerriero et al. (2004)*

Based on the available *in vivo* rabbit eye data, 38 of 44 substances could be classified according to the GHS system (UN 2003). Five excluded substances (including two glycols) were classified in the report as severe irritants based on *in vitro* data only (e.g., pH > 11 or < 2) and could not be used for this analysis. In addition, *in vivo* data was not provided for the sixth excluded substance. For the 38 substances, the IRE test method has an accuracy of 79% (30/38), a sensitivity of 100% (11/11), a specificity of 70% (19/27), a false positive rate of 30% (8/27), and a false negative rate of 0% (0/11) (**Table 6-1**).

Table 6-1 Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by Study

Data Source	Anal.[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al. 1995[4]	A	54/59	54	29/54	68	15/22	44	14/32	45	15/23	67	14/21	56	18/32	32	7/22
Balls et al. 1995[5]	A	216/236	60	130/216	72	63/88	52	67/128	51	63/124	73	67/92	48	61/128	25	28/88
Gettings et al. 1996	B	24/25	67	16/24	63	10/16	75	6/8	83	10/12	50	6/12	25	2/8	38	6/16
Guerriero et al. 2004	C	38/44	79	30/38	100	11/11	70	19/27	58	11/19	100	19/19	30	8/27	0	0/11
Expanded Data Set[6]	-	76/91	68	52/76	100	33/33	44	19/43	58	33/57	100	19/19	56	24/43	0	0/33
Pooled Data Set[7]	-	107/149	65	70/107	70	33/47	62	37/60	59	33/56	73	37/51	38	23/60	30	14/47

A = 1 and 4 hour corneal opacity and swelling.
B = Mean corneal swelling at 0.5, 1, 2, 3 and 4 hours and a numerical irritation scale ranging from 0-4 based on the swelling measurements.
C = Maximum corneal opacity, mean corneal swelling, maximum fluorescein uptake and evaluation of epithelial integrity (1, 2, 3, 4 hours).
[1]Anal. = Analytical method used to transform the sample data into IRE classification.
[2]N = Number of substances included in this analysis/number of substances in the study.
[3]The data on which the percentage calculation is based.
[4]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[5]Performance calculated using each individual *in vitro* classification from each of the four laboratories.
[6]Expanded Data Set includes substances classified as corrosives/severe irritants based on *in vitro* results by any single endpoint.
[7]Pooled Data Set includes data from Balls et al. (1995), Gettings et al. (1996), Guerriero et al. (2004). Consensus calls were used for substances tested more than once.

6.1.1.4 *Expanded Data Set*

Based on the available *in vivo* rabbit eye data and using the Expanded Data Set described in **Section 6.1,** 76 of 91 could be classified according to the GHS system (UN 2003). For the 76 substances classified, the IRE test method has an accuracy of 68% (52/76), a sensitivity of 100% (33/33), a specificity of 44% (19/43), a false positive rate of 56% (24/43), and a false negative rate of 0% (0/33).

6.1.1.5 *Pooled Data Set*

An additional analysis using pooled data from the Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004) studies is shown in **Table 6-1**. In this pooled data set, consensus calls were used for *in vitro* results on nine substances that were tested in more than one laboratory. In this pooled data set, an accuracy of 65% (70/107), a sensitivity of 70% (33/47), a specificity of 62% (37/60), a false positive rate of 38% (23/60), and a false negative rate of 30% (14/47) was obtained.

6.1.1.6 *Discordant Results According to the GHS Classification System*

In order to evaluate discordant responses of the IRE test method relative to the *in vivo* hazard classification, several subanalyses were performed. The subgroup analyses were conducted for both the Expanded Data Set (n = 76) and the Pooled Data Set (n = 107). These analyses included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, pH, physical form).

6.1.1.7 Expanded Data Set

As shown in **Table 6-2,** various subgroups of test substances impacted the performance of the IRE test method in the Expanded Data Set. For example, when substances were divided according to chemical class and there were at least five test substances included, false positive rates were greatest for alcohols (60% [6/10]), amines (60% [3/5]), esters (67% [4/6]), heterocycles (50% [4/8]), and ketones (67% [4/6]). There were no false negatives observed for any chemical class.

When physical properties were considered, liquids had a false positive rate of (83% [19/23]) and solids had a false positive rate of (25% [5/20]).

Of 10 surfactants that were assigned a GHS classification (UN 2003), 67% (2/3) were overpredicted. Three nonionic surfactants produced a false positive rate of 50% (1/2) and a 100% false negative rate (1/1). There were no anionic surfactants identified. Of 12 surfactant-based formulations tested, a 100% (2/2) false positive response was produced and none produced false negative responses (0/10).

Table 6-2 False Positive and False Negative 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[4]	
		%	No.[5]	%	No.
Overall	76	56	24/43	0	0/33
Chemical Class[6]					
Alcohol	11	60	6/10	0	0/1
Amide	5	0	0/3	0	0/2
Amine	9	60	3/5	0	0/4
Carboxylic acid	5	67	2/3	0	0/2
Ester	6	67	4/6	-	0/0
Ether	8	40	2/5	0	0/3
Formulation	12	100	2/2	0	0/10
Heterocycle	16	50	4/8	0	0/8
Ketone	6	67	4/6	-	0/0
Onium compound	9	33	1/3	0	0/6
Sulfur compound	7	20	1/5	0	0/2
Properties of Interest					
Liquid/Solution	43	83	19/23	0	0/20
Solids	33	25	5/20	0	0/13
Surfactants[7] - Total	10	67	2/3	0	0/7
-nonionic	3	50	1/2	0	0/1
-anionic	-	-	-	-	-
-cationic	7	100	1/1	0	0/6
Surfactant-based formulations	12	100	2/2	0	0/10
pH - Total[8]	27	24	4/17	0	0/10
-acidic (pH < 7.0)	18	20	2/10	0	0/8
-basic (pH > 7.0)	7	33	2/6	0	0/1
-neutral (pH = 7.0)	2	0	0/1	0	0/1
Category 1 Subgroup[9] - Total	25[11]	-	-	0	0/25
- 4 (CO=4 at any time)	8	-	-	0	0/8
- 3 (severity/persistence)	3	-	-	0	0/3
- 2 (severity)	2	-	-	0	0/2
- 2-4 combined[10]	13	-	-	0	0/13
- 1 (persistence)	12	-	-	0	0/12

[1]GHS = Globally Harmonized System (UN 2003).
[2]N = Number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.
[8]Total number of GHS Category 1 substances for which pH information was obtained.
[9]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 Corneal Opacity [CO] = 4); 3: based on lesions that are severe (not including CO = 4) and persistent; 4: CO = 4 at any time.
[10]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.
[11]The number of substances evaluated in the Category 1 subgroup analysis may be less than the number of *in vivo* Category 1 subsstances evaluated, since some substances could not be classified into the subgroups used in the evaluation.

Overall, the false positive rate of 27 substances with pH information that assigned a classification according to the GHS system (UN 2003) was 24% (4/17) with no false negative responses (0% [0/10]). Of the eighteen acidic substances (pH < 7.0) tested, the false positive rate was 20% (2/10) with no false negatives produced (0% [0/8]). Of the seven basic substances (pH > 7.0) evaluated, a higher false positive rate of 33% (2/6) was observed with no false negative outcomes (0% [0/1]). The two neutral substances (pH = 7.0) did not produce any false positive or false negative responses.

Finally, for 25 substances that were assigned a GHS classification (UN 2003), there were no incorrect *in vitro* classifications (false positive or false negative) based on whether the ocular lesions were based on either severity (n = 13) or persistence (n = 12).

6.1.1.8 Pooled Data Set

As shown in **Table 6-3,** various subgroups of test substances impacted the performance of the IRE test method in the Pooled Data Set. For example, when substances were divided according to chemical class and there were at least 5 test substances included, false positive rates were greatest for alcohols (55% [6/11]), amines (50% [3/6]), and ketones (67% [4/6]). The false negative rates were greatest for carboxylic acids (67% [4/6]) and organic compounds (50% [3/6]).

When physical properties were considered, liquids had higher false positive rate (49% [18/37]) when compared to solids (22% [5/23]). Liquids had a 29% (8/28) false negative rate compared to a 32% (6/19) false negative rate for solids.

Of 13 surfactants that were assigned a GHS classification (UN 2003), 40% (2/5) were overpredicted and 12% (1/8) were underpredicted. Four nonionic surfactants produced a false positive rate of 33% (1/3) with no false positive responses (0% [0/1]). Of two anionic surfactants identified, no false positives were produced (0% [0/1]), but there was one false negative outcome (100% [1/1]). Seven cationic surfactants were available with one false positive (100% [1/1]) and no false negative outcomes (0% [0/6]). Of 24 surfactant-based formulations, 25% (2/8) were overpredicted and 38% (6/16) were underpredicted.

Overall, the false positive rate of 27 substances with pH information that assigned a classification according to the GHS system (UN 2003) was 24% (4/17) with no false negative responses (0% [0/10]). Eighteen acidic substances (pH < 7.0) produced a false positive rate of 20% (2/10) with no false negative outcomes (0% [0/8]). Seven basic substances (pH > 7.0) produced a higher false positive rate of 33% (2/6) with no false positive outcomes (0% [0/1]). Two neutral substances (pH =7.0) did not produce any false positive or false negative responses.

Finally, for 37 substances that were assigned a GHS Category 1 classification (UN 2003), the false negative rate was 32% (12/37). False negative rates were greater for substances classified *in vivo* (according to the GHS classification system) based on persistent lesions (37% [7/19]), rather than severe lesions (28% [5/18]). However, three substances that caused severe lesions *in vivo* (corneal opacity = 4) were false negatives.

Table 6-3 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS[1] Classification System (Analysis Based on the Pooled Data Set)

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	107	38	23/60	30	14/47
Chemical Class[6]					
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
Surfactants[7] - Total	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
Surfactant-based formulations	24	25	2/8	38	6/16
pH - Total[8]	27	24	4/17	0	0/10
-acidic (pH < 7.0)	18	20	2/10	0	0/8
-basic (pH > 7.0)	7	33	2/6	0	0/1
-neutral (pH = 7.0)	2	0	0/1	0	0/1
Category 1 Subgroup[9] - Total	37[11]	-	-	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[10]	18	-	-	28	5/18
- 1 (persistence)	19	-	-	37	7/19

[1]GHS = Globally Harmonized System (UN 2003).
[2]N = Number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.
[8]Total number of GHS Category 1 substances for which pH information was obtained.
[9]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 Corneal Opacity [CO]=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO = 4 at any time.
[10]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.
[11]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.

6.1.2 EPA Classification System: IRE Test Method Accuracy

Accuracy analyses for ocular corrosives and severe irritancy, as defined by the EPA regulatory classification system[3] were conducted on data obtained from Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004). The EPA classification assigned to each test substance is presented in **Appendix D**. To the extent possible, severe ocular irritants were identified from the *in vitro* data in these reports using the Guerriero et al. (2004) IRE test method scoring system described in **Section 6.1**. The performance characteristics of the three studies are shown in **Table 6-4** and are based on the available *in vivo* reference data for each study. Of the three studies, Balls provided IRE data for substances tested in multiple laboratories; the first set of accuracy calculations for these studies in **Table 6-4** (n = 59) represents the results obtained using the consensus call for each test substance, while the second set of accuracy calculations for each study represents the results obtained when each independent test result from each laboratory was considered separately (n = 236).

6.1.2.1 *Balls et al. (1995)*

Based on the available *in vivo* rabbit eye data, 53 of the 59 substances tested in this study could be assigned an EPA classification (**Table 6-4**) (EPA 1996). The remaining six substances had inadequate *in vivo* data for assigning a classification according to the EPA system (EPA 1996). For the 53 substances that could be evaluated, the IRE test method has an accuracy of 51% (27/53), a sensitivity of 65% (13/20), a specificity of 42% (14/33), a false positive rate of 58% (19/33), and a false negative rate of 35% (7/20) (**Table 6-4**).

For Balls et al. (1995), using the second approach in which the result of each IRE test result is considered separately and test substances could be classified according to the EPA system (EPA 1996) (n = 208/236), the IRE test method has an accuracy of 56% (116/208), a sensitivity of 65% (47/72), a specificity of 51% (69/136), a false positive rate of 49% (67/136), and a false negative rate of 35% (25/72), for identifying ocular corrosives and severe irritants as classified by the EPA (**Table 6-4**).

6.1.2.2 *Gettings et al. (1996)*

Based on the available *in vivo* rabbit eye test data, all 25 test substances could be assigned an EPA ocular hazard classification (EPA 1996). Using these data, the IRE test method has an accuracy of 64% (16/25), a sensitivity of 59% (10/17), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 41% (7/17) (**Table 6-4**).

6.1.2.3 *Guerriero et al. (2004)*

Based on the available *in vivo* rabbit eye test data, 38 of the 44 substances could be assigned an EPA hazard classification (EPA 1996) (**Table 6-4**). The remaining six substances had inadequate *in vivo* data for assigning a classification according to the EPA system (EPA 1996). For the 38 substances that could be evaluated, the IRE test method has an accuracy of 79% (30/38), a sensitivity of 100% (11/11), a specificity of 70% (19/27), a false positive rate of 30% (8/27), and a false negative rate of 0% (0/11) (**Table 6-4**).

---

[3] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify EPA (EPA 1996) Category I irritants (i.e., severe irritants); substances classified as EPA Category II, III, or IV irritants were defined as nonsevere irritants.

Table 6-4 Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by Study

Data Source	Anal.[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al. 1995[4]	A	53/59	51	27/53	65	13/20	42	14/33	41	13/32	67	14/21	58	19/33	35	7/20
Balls et al. 1995[5]	A	208/236	56	116/208	65	47/72	51	69/136	41	47/114	73	69/94	49	67/136	35	25/72
Gettings et al. 1996	B	25/25	64	16/25	59	10/17	75	6/8	83	10/12	46	6/13	25	2/8	41	7/17
Guerriero et al. 2004	C	38/44	79	30/38	100	11/11	70	19/27	58	11/19	100	19/19	30	8/27	0	0/11
Expanded Data Set[6]	-	76/91	66	50/76	100	31/31	42	19/45	54	31/57	100	19/19	58	26/45	0	0/31
Pooled Data Set[7]	-	107/149	64	68/107	69	31/45	60	37/62	55	31/56	73	37/51	40	25/62	31	14/45

A = 1 and 4 hour corneal opacity and swelling.
B = Mean corneal swelling at 0.5, 1, 2, 3 and 4 hours and a numerical irritation scale ranging from 0-4 based on the swelling measurements.
C = Maximum corneal opacity, mean corneal swelling, maximum fluorescein uptake and evaluation of epithelial integrity (1, 2, 3, 4 hours).
[1]Anal. = Analytical method used to transform the sample data into IRE classification.
[2]N = Number of substances included in this analysis/number of substances in the study.
[3]The data on which the percentage calculation is based.
[4]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[5]Performance calculated using each individual *in vitro* classification from each of the four laboratories.
[6]Expanded Data Set includes substances classified as corrosives/severe irritants based on *in vitro* results by any single endpoint.
[7]Pooled Data Set includes data from Balls et al. (1995), Gettings et al. (1996), Guerriero et al. (2004). Consensus calls were used for substances tested more than once.

6.1.2.4 *Expanded Data Set*

Based on the available *in vivo* rabbit eye data and using the Expanded Data Set described in **Section 6.1,** 76 of 91 could be classified according to the EPA system (EPA 1996). For the 76 substances classified, the IRE test method has an accuracy of 66% (50/76), a sensitivity of 100% (31/31), a specificity of 42% (19/45), a false positive rate of 58% (26/45), and a false negative rate of 0% (0/31).

6.1.2.5 *Pooled Data Set*

An additional analysis using pooled data from the Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004) studies is shown in **Table 6-4**. In this pooled data set, consensus calls were used for *in vitro* results on two substances that were tested in more than one laboratory. In this pooled data set, an accuracy of 64% (68/107), a sensitivity of 69% (31/45), a specificity of 60% (37/62), a false positive rate of 40% (25/62), and a false negative rate of 31% (14/45) was obtained.

6.1.2.6 *Discordant Results According to the EPA Classification System*

In order to evaluate discordant responses of the IRE test method relative to the *in vivo* hazard classification, several accuracy subanalyses were performed. Due to a limited number of available substances using the decision criteria outlined in the IRE BRD, the subgroup analyses were based on both the Expanded Data Set (n = 76) shown in **Table 6-5** and the Pooled Data Set (n = 107) shown in **Table 6-6**. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, pH, physical form).

As indicated in **Table 6-5,** using 76 substances in the Expanded Data Set, various subgroups of test substances impacted the performance of the IRE test method. For example, when substances were divided according to chemical class and there were at least 5 test substances included, false positive rates were greatest for alcohols (75% [6/8]), amines (67% [4/6]), esters (67% [4/6]), ethers (50% [3/6]), heterocycles (50% [4/8]), and ketones (67% [4/6]). There were no false negatives observed for any chemical class.

When physical properties were considered, liquids had a higher false positive rate (83% [20/24]) when compared to solids (29% [6/21]).

Of nine surfactants that were classified using the EPA classification system (EPA 1996), the false positive rate was 100% (3/3) with no false negative responses (0% [0/6]). Three nonionic surfactants produced a false positive rate of 100% (2/2) and a false negative rate of 0% (0/1). Six cationic surfactants produced a false positive rate of 100% (1/1) with no false negative responses 0% [0/5]). There were no anionic surfactants identified. Of 12 surfactant-based formulations, none (0/12) were overpredicted and none were underpredicted.

Overall, the false positive rate for 27 substances with pH information that were assigned a classification according to the EPA system (EPA 1996) was 24% (4/17) with no false negatives (0% [0/10]). Eighteen acidic substances (pH < 7.0) produced a false positive rate of 20% (2/10) with no false negative responses (0% [0/8]). Seven basic substances (pH >

7.0) produced a higher false positive rate (33% [2/6]) than the acidic substances with no false negative responses (0% [0/1]). Two neutral substances (pH = 7.0) did not produce any false positive or false negative responses.

Table 6-5. False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EPA[1] Classification System (Analysis Based on the Expanded Data Set)

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	76	58	26/45	0	0/31
Chemical Class					
Alcohol	10	75	6/8	0	0/2
Amide	5	0	0/3	0	0/2
Amine	10	67	4/6	0	0/4
Carboxylic acid	6	67	2/3	0	0/3
Ester	6	67	4/6	-	0
Ether	8	50	3/6	0	0/2
Formulation	12	100	2/2	0	0/10
Heterocycle	15	50	4/8	0	0/7
Ketone	6	67	4/6	-	0
Onium compound	11	67	4/6	0	0/5
Sulfur compound	7	20	1/5	0	0/2
Properties of Interest					
Liquid/Solution	43	83	20/24	0	0/19
Solid	33	29	6/21	0	0/12
Surfactants – Total	9	100	3/3	0	0/6
-nonionic	3	100	2/2	0	0/1
-anionic	-	-	-	-	-
-cationic	6	100	1/1	0	0/5
Surfactant-based formulations	12	0	0/12	-	-
pH – Total[7]	27	24	4/17	0	0/10
- acidic (pH < 7.0)	18	20	2/10	0	0/8
- basic (pH > 7.0)	7	33	2/6	0	0/1
- neutral (pH = 7.0)	2	0	0/1	0	0/1

[1]EPA = U.S. Environmental Protection Agency (EPA 1996).
[2]N = Number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]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). See **Appendix B**.
[7]Total number of EPA Category I substances for which pH information was available.

Table 6-6 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EPA[1] Classification System (Analysis Based on the Pooled Data Set)

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	107	40	25/62	31	14/45
Chemical Class[6]					
Alcohol	13	55	6/11	50	1/2
Amide	5	0	0/3	0	0/2
Amine	12	57	4/7	20	1/5
Carboxylic acid	12	50	3/6	50	3/6
Ester	10	30	3/10	-	0/0
Ether	9	43	3/7	0	0/2
Formulation	25	25	2/8	41	7/17
Heterocycle	17	44	4/9	13	1/8
Ketone	6	67	4/6	-	0/0
Onium compound	9	33	1/3	0	0/6
Organic	13	29	2/7	50	3/6
Sulfur compound	8	20	1/5	33	1/3
Properties of Interest					
Liquid/Solution	66	50	19/38	32	9/28
Solids	41	25	6/24	29	5/17
Surfactants[7] - Total	12	50	3/6	17	1/6
-nonionic	4	50	2/4	-	0/0
-anionic	2	0	0/1	100	1/1
-cationic	6	100	1/1	0	0/5
Surfactant-based formulations	25	25	2/8	41	7/17
pH - Total[8]	27	24	4/17	0	0/10
-acidic (pH < 7.0)	18	20	2/10	0	0/8
-basic (pH > 7.0)	7	33	2/6	0	0/1
-neutral (pH = 7.0)	2	0	0/1	0	0/1

[1]EPA = U.S. Environmental Protection Agency (EPA 1996).
[2]N = Number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.
[8]Total number of EPA Category I substances for which pH information was obtained.

As indicated in **Table 6-6,** using 107 substances in the Pooled Data Set, various subgroups of test substances impacted the performance of the IRE test method. For example, when substances were divided according to chemical class and there were at least 5 test substances included, false positive rates were greatest for alcohols (55% [6/11]), amines (57% [4/7]), carboxylic acids (50% [3/6]), and ketones (67% [4/6]). False negative rates were greatest for carboxylic acids (50% [3/6]) and organic compounds (50% [3/6]).
When physical properties were considered, liquids had a higher false positive rate (50% [19/38]) when compared to solids (25% [6/24]). The false negative rate of liquids was 32% (9/28) and 29% (5/17) for solids.

Of 12 surfactants that were classified using the EPA classification system (EPA 1996), the false positive rate was 50% (3/6) and the false negative rate was 17% (1/6). Four nonionic surfactants produced a false positive rate of 50% (2/4) with no false negative responses. Of two anionic surfactants identified, there were no false positives (0% [0/1]) and one false negative outcome (100% [1/1]). Six cationic surfactants produced a false positive rate of 100% (1/1) and a false negative rate of 0% (0/5). Of 25 surfactant-based formulations, 25% (2/8) were overpredicted and 41% (7/17) were underpredicted.

Overall, the false positive rate for 27 substances with pH information that were assigned a classification according to the EPA system (EPA (1996) was 24% (4/17) with no false negatives (0% [0/10]). Eighteen acidic substances (pH < 7.0) produced a false positive rate of 20% (2/10) with no false negative responses (0% [0/8]). Seven basic substances (pH > 7.0) produced a higher false positive rate (33% [2/6]) than the acidic substances with no false negative outcomes (0% [0/1]).

6.1.3 EU Classification System: IRE Test Method Accuracy

Accuracy analyses using the EU regulatory classification system[4] (EU 2001) were conducted on data obtained from CEC (1991), Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004). To the extent possible, severe ocular irritants were identified from the *in vitro* data in these reports using the Guerriero et al. (2004) IRE test method scoring system described in **Section 6.1**. The EU classification (EU 2001) assigned to each test substance is presented in **Appendix D**. To the extent possible, severe ocular irritants were identified from the *in vitro* data in these reports using the Guerriero et al. (2004) IRE test method scoring system described in **Section 6.1**. The performance characteristics of the four studies are shown in **Table 6-7** and are based on the available *in vivo* reference data for each study. Of the four studies, CEC (1991) and Balls et al. (1996) provided IRE data for substances tested in multiple laboratories; the first set of accuracy calculations for these studies in **Table 6-7** (n = 21 and n = 59, respectively) represents the results obtained using the consensus call for each test substance, while the second set of accuracy calculations for each study represents the results obtained when each independent test result from each laboratory was considered separately (n = 63 and n = 236, respectively).

6.1.3.1 *CEC Collaborative Study (1991)*

In this collaborative study, 15 of 21 substances tested had sufficient information to assign a EU classification (EU 2001). Of the 15 substances that could be evaluated, the IRE test method had an accuracy of 87% (13/15), a sensitivity of 100% (5/5), a specificity of 80% (8/10), a false positive rate of 20% (2/10), and a false negative rate of 0% (0/5) (**Table 6-7**).

When the performance was calculated on each individual test substance based on availability of *in vivo* rabbit eye test data (n = 44/63), the IRE test method had an accuracy of 77% (34/44), a sensitivity of 86% (12/14), a specificity of 73% (22/30), a false positive rate of 27% (8/30), and a false negative rate of 14% (2/14) (**Table 6-7**).

---

[4] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify R41 irritants (i.e., severe irritants); substances classified as R36 were defined as nonsevere irritants.

Table 6-7 Evaluation of the Performance of the IRE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to *In Vivo* Findings, as Defined by the EU Classification System, by Study

Data Source	Anal.[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
CEC 1991[4]	A	15/21	87	13/15	100	5/5	80	8/10	71	5/7	100	8/8	20	2/10	0	0/5
CEC 1991[5]	A	44/63	77	34/44	86	12/14	73	22/30	60	12/20	92	22/24	27	8/30	14	2/14
Balls et al. 1995	B	49/59	55	27/49	74	14/19	43	13/30	45	14/31	72	13/18	57	17/30	26	5/19
Balls et al. 1995[e]	B	196/236	62	121/196	76	58/76	53	63/120	50	58/115	78	63/81	48	57/120	24	18/76
Gettings et al. 1996	C	24/25	67	16/24	63	10/16	75	6/8	83	10/12	50	6/12	25	2/8	38	6/16
Guerriero et al. 2004	D[f]	38/44	79	30/38	100	11/11	70	19/27	58	11/19	100	19/19	30	8/27	0	0/11
Expanded Data Set[6]	-	80/90	70	56/80	100	37/37	44	19/43	61	37/61	100	19/19	56	24/43	0	0/37
Pooled Data Set[7]	-	114/149	69	79/114	76	37/49	65	42/65	62	37/60	78	42/54	35	23/65	24	12/49

A = Corneal opacity; corneal swelling, fluorescein retention at 0.5, 1, 1.25, 2, 3 and 4 hours.
B = 1 and 4 hour corneal opacity and swelling.
C = Mean corneal swelling at 0.5, 1, 2, 3 and 4 hours and a numerical irritation scale ranging from 0-4 based on the swelling measurements.
D = Corneal opacity/area; fluorescein penetration, corneal swelling, epithelial integrity at 0.5, 1, 2, 3, and 4 hours using assigned cut-off values.
[1]Anal. = Analytical method used to transform the sample data into IRE classification.
[2]N = Number of substances included in the study/number of substances in the study.
[3]The data on which the percentage calculation is based.
[4]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the three or four testing laboratories.
[5]Performance calculated using each individual *in vitro* classification from each of the four laboratories.
[6]Expanded Data Set includes substances classified as corrosives/severe irritants based on *in vitro* results by any single endpoint.
[7]Pooled Data Set includes data from CEC (1991), Balls et al. (1995), Gettings et al. (1996), Guerriero et al. (2004). Consensus calls were used for substances tested more than once.

6.1.3.2 *Balls et al. (1995)*

In this validation study, 49 of 59 substances tested could be classified according to the EU system (EU 2001). Using these data, the IRE test method had an accuracy of 55% (27/49), a sensitivity of 74% (14/19), a specificity of 43% (13/30), a false positive rate of 57% (17/30), and a false negative rate of 26% (5/19) (**Table 6-7**).

Using the second approach, in which the result of each IRE test method experiment was considered separately (n = 196/236), the IRE test method had an accuracy of 62% (121/196), a sensitivity of 76% (58/76), a specificity of 53% (63/120), a false positive rate of 48% (57/120) and a false negative rate of 24% (18/76), for identifying ocular corrosives and severe irritants as classified by the EU (EU 2001).

6.1.3.3 *Gettings et al. (1996)*

Based on the available *in vivo* rabbit eye data, 24 of the 25 substances tested could be assigned a classification according to the EU system (EU 2001). For these substances, the IRE test method had an accuracy of 67% (16/24), a sensitivity of 63% (10/16), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 38% (6/16) (**Table 6-7**).

6.1.3.4 *Guerriero et al. (2004)*

In this study, 38 of 44 substances tested could be classified using the EU classification system (EU 2001). Five substances were excluded from EU classification based on the use of *in vitro* data only (e.g., pH extremes) to classify the ocular irritancy of these substances according to the EU system (EU 2001). Furthermore, although the EU classification (EU 2001) of one substance (allyl alcohol) was based on *in vivo* rabbit eye data, the raw *in vivo* scores were not available for subsequent analysis. For these 38 substances, the IRE test method had an accuracy of 79% (30/38), a sensitivity of 100% (11/11), a specificity of 70% (19/27), a false positive rate of 30% (8/27), and a false negative rate of 0% (0/11) (**Table 6-7**).

6.1.3.5 *Expanded Data Set*

Based on the available *in vivo* rabbit eye data and using the Expanded Data Set described in **Section 6.1,** 80 of 90 substances could be classified according to the EU system (EU 2001). For the 80 substances classified, the IRE test method has an accuracy of 70% (56/80), a sensitivity of 100% (37/37), a specificity of 44% (19/43), a false positive rate of 56% (24/43), and a false negative rate of 0% (0/37).

6.1.3.6 *Pooled Data Set*

An additional analysis using pooled data from the CEC (1991), Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004) studies is shown in **Table 6-7**. In this pooled data set, consensus calls were used for *in vitro* results on eight substances that were tested in more than one laboratory. In this pooled data set, an accuracy of 69% (79/114), a sensitivity of 76% (37/49), a specificity of 65% (42/65), a false positive rate of 35% (23/65), and a false negative rate of 24% (12/49) was obtained.

6.1.3.7 *Discordant Results According to the EU Classification System*

In order to evaluate discordant responses of the IRE test method relative to the *in vivo* hazard classification, several accuracy subanalyses were performed using the Expanded Data Set (n = 80 substances) and the Pooled Data Set (n = 114 substances). These included specific classes of chemicals with sufficiently robust numbers of substances (n ≥ 5), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, pH, physical form).

As indicated in **Table 6-8** using the 80 substances in the Expanded Data Set, various subgroups of test substances impacted the performance of the IRE test method. For example, when substances were divided according to chemical class and there were at least 5 test substances included, false positive rates were greatest for alcohols (60% [6/10]), amines (60% [3/5]), carboxylic acids (60% [3/5]), esters (67% [4/6]), heterocycles (50% [4/8]), and ketones (67% [4/6]). There were no false negatives observed for any chemical class.

When physical properties were considered, liquids had a high false positive rate (82% [18/22]) when compared to solids (25% [5/20]).

Of nine surfactants that were classified using the EU system (EU 2001), the false positive rate was 100% (3/3) with no false negatives (0% [0/6]). For three nonionic surfactants, the false positive rate was 100% (2/2), while the false positive rate was 0% (0/1). For six cationic surfactants the false positive rate was 100% (1/1) with no false negatives identified (0% [0/5]). There were no anionic surfactants identified. None of 12 surfactant-based formulations (0/12) produced false positive responses and none produced false negative responses.

Overall, the false positive rate of 27 substances with pH information that could be classified according to the EU (EU 2001) system was 24% (4/17) with a false negative rate of 0% (0/8). Eighteen acidic substances (pH < 7.0) produced a false positive rate of 20% (2/10) and a false negative rate of 0% (0/8). Seven basic substances (pH > 7.0) produced a higher false positive rate of 33% (2/6) than the acidic substances with no false negatives (0/1). Neutral substances (pH = 7.0; n = 2) did not produce any false positive (0% [0/1]) or false negative responses (0% [0/1]).

As indicated in **Table 6-9** using the 114 substances in the Pooled Data Set, various subgroups of test substances impacted the performance of the IRE test method. For example, when substances were divided according to chemical class and there were at least 5 test substances included, false positive rates were greatest for alcohols (46% [6/13]), heterocycles (44% [4/9]), and ketones (67% [4/6]). False negative rates were greatest for formulations (38% [6/16]).

When physical properties were considered, liquids had a high false positive rate (43% [18/42]) when compared to solids (22% [5/23]). The false negative rate for liquids was 22% (7/32) and 29% (5/17) for solids.

Table 6-8 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EU[1] Classification System (Analysis Based on the Expanded Data Set)

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	80	56	24/43	0	0/37
Chemical Class[6]					
Alcohol	11	60	6/10	0	0/1
Amide	5	0	0/3	0	0/2
Amine	9	60	3/5	0	0/4
Carboxylic acid	7	60	3/5	0	0/4
Ester	6	67	4/6	-	0
Ether	8	40	2/5	0	0/3
Formulation	12	100	2/2	0	0/10
Heterocycle	16	50	4/8	0	0/8
Ketone	6	67	4/6	-	0
Onium compound	10	33	1/3	0	0/7
Sulfur compound	7	20	1/5	0	0/2
Properties of Interest					
Liquid/Solution	48	82	18/22	0	0/26
Solid	32	25	5/20	0	0/12
Surfactants – Total	9	100	3/3	0	0/6
-nonionic	3	100	2/2	0	0/1
-anionic	-	-	-	-	-
-cationic	6	100	1/1	0	0/5
Surfactant-based formulations	12	0	0/12	-	-
pH – Total[7]	27	24	4/17	0	0/10
- acidic (pH < 7.0)	18	20	2/10	0	0/8
- basic (pH > 7.0)	7	33	2/6	0	0/1
- neutral (pH = 7.0)	2	0	0/1	0	0/1

[1]EU = European Union (EU 2001).
[2]N = Number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]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). See **Appendix B**.
[7]Total number of EU R41 substances for which pH information was available.

Table 6-9 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the EU[1] Classification System (Analysis Based on the Pooled Data Set)

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	114	35	23/65	24	12/49
Chemical Class[6]					
Alcohol	15	46	6/13	50	1/2
Amide	5	0	0/3	0	0/2
Amine	12	43	3/7	20	1/5
Carboxylic acid	12	33	2/6	33	2/6
Ester	12	25	3/12	-	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	11	33	1/3	0	0/8
Organic	12	17	1/6	33	2/6
Sulfur compound	8	20	1/5	33	1/3
Properties of Interest					
Liquid/Solution	74	43	18/42	22	7/32
Solids	40	22	5/23	29	5/17
Surfactant - Total	13	40	2/5	0	0/8
-nonionic	4	33	1/3	0	0/1
-anionic	1	0	0/1	-	0/0
-cationic	8	100	1/1	0	0/7
Surfactant-based formulations	24	25	2/8	38	6/16
pH – Total[7]	27	24	4/17	0	0/10
-acidic (pH < 7.0)	18	20	2/10	0	0/8
-basic (pH > 7.0)	7	33	2/6	0	0/1
-neutral (pH = 7.0)	2	0	0/1	0	0/1

[1]EU = European Union (EU 2001).
[2]N = Number of substances.
[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Total number of substances for which pH information was obtained.

Of 13 surfactants that were classified using the EU system (EU 2001), the false positive rate was 40% (2/5) with a false negative rate of 0% (0/8). For four nonionic surfactants, the false positive rate was 33% (1/3), while the false negative rate was 0% (0/1). One anionic surfactant was identified that produced no false positive (0% [0/1]) or false negative (0/0) responses. For eight cationic surfactants the false positive rate was 100% (1/1) with no false negatives identified (0% [0/7]). For 25 surfactant-based formulations, the overprediction rate was 25% (2/8) and the false negative rate was 38% (6/16).

Overall, the false positive rate for substances with pH information that were classified according to the EU system (EU 2001) was 24% (4/17) with no false negatives (0% [0/10]). Eighteen acidic substances (pH < 7) produced a false positive rate of 20% (2/10) and a false negative rate of 0% (0/8). Seven basic substances (pH > 7) produced a higher false positive

rate of 33% (2/6) than the acidic substances with no false negative outcomes (0/1). Two neutral substances (pH = 7) did not produce any false positive (0% [0/1]) or false negative responses (0% [0/1]).

6.2 Accuracy of the IRE Test Method for Identifying Ocular Corrosives and Severe Irritants - Summary of Results

While there were some differences in results among the three hazard classification systems evaluated (i.e., GHS [UN 2003], EPA [1996], EU [2001]), the accuracy analysis revealed that IRE test method performance was comparable among the three systems. As can be seen in **Tables 6-1**, **6-4**, and **6-7**, the overall accuracy of the IRE test method ranged from 51 to 87%, depending on the classification system used. For example, in the Balls et al. (1995) study the accuracy among the three regulatory classifications systems (GHS [UN2003]; EPA [1996], and EU [EU 2001]) ranged from 51 to 55%, sensitivity ranged from 65 to 74%, specificity ranged from 42 to 44%, the false positive rate ranged from 56 to 58%, and the false negative rate ranged from 26 to 35%. For the Guerriero et al. (2004) study, the overall accuracy, sensitivity, specificity, false positive rate, and false negative rate was 79%, 100%, 70%, 30%, and 0%, respectively, across the three regulatory systems. Given the relatively homogeneous performance of the IRE test method among the three classification systems, the discussion below encompasses the three hazard classification systems, unless otherwise indicated.

6.2.1 Discordance Among Chemical Classes

The accuracy analysis based on chemical class was performed using the Expanded Data Set (n = 76 to 80) and the Pooled Data Set (n = 107 to 114) depending on the regulatory classification system used, because each data set presented advantages and disadvantages. For the purposes of these analyses, chemical classes represented by fewer than five substances were not considered.

The results of this analysis on the Expanded Data Set indicated that alcohols are often overpredicted in the IRE test method (60% to 75% [6/8 to 6/10] false positive rate, depending on the classification system used). Amines (60 to 67% [3/5 to 4/6]), carboxylic acids (60 to 67% [2/3 to 3/5]), esters (67% [4/6]), heterocycles (50% [4/8]), ketones (67% [4/6]) and onium compounds (33% to 67% [1/3 to 4/6]) also were overpredicted.

There were no underpredicted substances in the Expanded Data Base.

The results of this analysis on the Pooled Data Set indicated that alcohols are often overpredicted in the IRE test method (46 to55%[6/11 to 6/13]) false positive rate, depending on the classification system used). Amines (43 to 57% [3/7 to 4/7]), carboxylic acids (33 to 50% [2/6 to 3/6]), heterocycles (44% [4/9 across all classifications]), and ketones (67% [4/6]) also had high false positive rates. Liquid substances produced a false positive rate of 43% (18/42), and solid substances produced a false positive rate of 22% (5/23).

There were false negative responses produced in the Pooled Data Set by alcohols (50% [1/2]), carboxylic acids (33 to 67% [2/6 to 4/6]), formulations (38 to 41% [6/16 to 7/17]), and organic compounds (33 to 50% [2/6 to 3/6]).

6.2.2 Discordance Among Physical or Chemical Properties of Interest

With regard to physical form of the substances overpredicted by the IRE test method using the Expanded Data Set, 19 to 20 were liquids or solutions and five to six were solids. Considering the proportion of the total available database, liquids (19/23 to 20/24) appear more likely than solids (5/20 to 6/21) to be overpredicted by the IRE test method.

Of nine to 13 surfactants evaluated, 40 to 100% (2/5 to 3/3) were overpredicted across the three regulatory classification systems. One or both (50 to 100%) of two surfactants that could be identified as nonionic surfactants were overpredicted depending on the classification system used. One substance identified as a cationic surfactant was overpredicted across the three regulatory classification systems. Of the 12 surfactant-based formulations evaluated across regulatory classification systems, the overprediction rate was 0% (0/12) and no substances were underpredicted.

Of 27 substances with pH information, 24% (4/17) were overpredicted across the three regulatory classification systems. Basic substances (pH > 7) appear to contribute the highest false positive rate (33% [4/6]) across the three regulatory classification systems.

Of the twenty-five substances categorized as GHS Category 1 (UN 2003) severe irritants, 12 were subgrouped as producing persistent lesions (Subgroup 1), whereas 13 were subgrouped as producing severe lesions (subgroup 2 to 4). There were no underpredicted substances in these subgroups.

With regard to physical form of the substances overpredicted by the IRE test method using the Pooled Data Set, 18 to 19 were liquids or solutions and 5 to 6 were solids. Considering the proportion of the total available database, liquids (18/42 to 19/38) appear more likely than solids (5/23 to 6/24) to be overpredicted by the IRE test method.

Of the 17 to 25 surfactants evaluated, 25 to 36% (2/8 to 4/11) were overpredicted across the three regulatory classification systems. The actual number of overpredicted substances for any specific form of surfactant (nonionic, cationic, or anionic) ranged from 0 to 2 and was not adequate to draw any significant conclusions on these subclasses from the data. Of the 25 surfactant-based formulations, 25% were overpredicted (2/8) and 38% (6/16) were underpredicted.

Of 27 substances with pH information, 24% (4/17) were overpredicted across the three regulatory classification systems. Basic substances (pH > 7) appear to contribute the highest false positive rate (33%; 4/6) across the three regulatory classification systems.

Of the 37 substances categorized as GHS Category 1 (UN 2003) severe irritants, 19 were subgrouped as producing persistent lesions (Subgroup 1), whereas 18 were subgrouped as producing severe lesions (subgroup 2 to 4), while underpredicted substances in the Pooled

Data Set (25 to 37% [1/4 to 7/19]),. However, the underprediction rate was relatively uniform across all subgroups and was independent of persistence or severity.

[This Page Intentionally Left Blank]

7.0 IRE TEST METHOD RELIABILITY

An assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM 2003). Repeatability refers to 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 (ICCVAM 1997 2003). Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Interlaboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test chemicals, and indicates the extent to which a test method can be transferred successfully among laboratories. A reliability assessment includes reviewing the rationale for selecting the substances used to evaluate test method reliability, a discussion of the extent to which the substances tested represent the range of possible test outcomes and the properties of the various substances for which the test method is proposed for use, and a quantitative and/or qualitative analysis of repeatability and intra- and inter-laboratory reproducibility. In addition, measures of central tendency and variation are summarized for historical control data (negative, vehicle, and positive), where applicable.

Due to the lack of quantitative IRE test method data for replicate experiments within an individual laboratory, an evaluation of the intralaboratory repeatability and reproducibility of the IRE test method could not be conducted. However, multilaboratory qualitative and quantitative IRE test data were available for a collaborative study by the CEC (1991) and a validation study conducted by Balls et al. (1995). Three laboratories participated in the CEC (1991) collaborative study and four laboratories participated in the Balls et al. (1995) validation effort. In the CEC (1991) study, each substance tested was assigned a EU classification (R41, R36, or nonirritant [EU 2001]) based on *in vivo* rabbit eye test results. However, due to the lack of individual rabbit *in vivo* Draize scores, a reliability assessment for the CEC (1991) study using the GHS (UN 2003) or EPA (EPA 1996) classification criteria was not possible. The Balls et al. (1995) data were used for an evaluation of the interlaboratory reproducibility of the IRE test method according to the GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) classification systems.

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the IRE Test Method

The quality of a reliability evaluation depends on the extent to which the substances tested adequately represent the range of physicochemical characteristics and response levels that the test method must be capable of evaluating. The only sources of data for conducting an assessment of interlaboratory reproducibility were the CEC (1991) collaborative study and the EC/HO validation study reported on by Balls et al. (1995).

The CEC (1991) collaborative pilot study evaluated the reproducibility of the IRE test method using 21 substances. These substances were provided by FRAME via Aldrich Chemical Company Limited and were selected to cover a full range of eye irritation

potential. A single supplier provided the substances and each chemical originated from a single batch. All of the substances were > 96% pure. The authors also intended the list of substances to be representative of a variety of chemical structures and representative of currently used industrial chemicals. Furthermore, the authors chose substances with *in vivo* data to which a EU risk phrase could be assigned and, where possible, those that had been used in previous validation studies.

The Balls et al. (1995) study evaluated the performance and reproducibility of the IRE test method using 60 "substances" (i.e., there were 52 different substances with four substances tested at two different concentrations and two substances tested at three concentrations, for a total of 60 possible ocular irritation outcomes). To be selected for inclusion in this study, the substances had to be single chemicals (no mixtures) available at high purity and stable when stored, and the reference *in vivo* rabbit eye data had to have been generated since 1981 according to OECD TG 405 following GLP guidelines. In addition, substances were selected to ensure an adequately diverse group of physicochemical characteristics and levels of irritancy severity. One substance (thiourea) was tested *in vitro* in the IRE assay but, due to its excessive toxicity *in vivo*, excluded from the comparison of *in vitro* and *in vivo* test results.

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Quantitative and Qualitative Assessments of Intralaboratory Repeatability

Generally, analyses of intralaboratory repeatability have included approaches such as:

- a coefficient of variation (CV) analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. 1996)
- analysis of variance (ANOVA) methods (e.g., Holzhütter et al. 1996; ASTM 1999)

Due to the lack of available IRE test data for replicate enucleated rabbit eyes within individual experiments and for experiments conducted on the same substance under exactly the same conditions, an evaluation of the intralaboratory repeatability of the IRE test method could not be conducted.

7.2.2 Quantitative and Qualitative Assessments of Intralaboratory Reproducibility

Generally, analyses of intralaboratory reproducibility have included approaches such as:

- a CV analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. 1996)
- ANOVA methods (e.g., Holzhütter et al.[1996; ASTM 1999)

Due to the lack of available IRE test data for experiments conducted multiple times in the same laboratory, an evaluation of IRE test method intralaboratory reproducibility could not be conducted.

7.2.3 Assessment of Interlaboratory Reproducibility

Generally, analyses of interlaboratory variability have included approaches such as:

- determination of the extent of concordance among laboratories in assigning the same regulatory classification for a particular substance (e.g., Holzhütter et al. 1996)
- a CV analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. 1996)
- ANOVA methods (e.g., Holzhütter et al. 1996; ASTM 1999)
- bivariant scatter diagrams/correlation analyses for pairs of laboratories to assess the extent possibility of divergence (e.g., Holzhütter et al. 1996)

Several of the studies discussed in **Section 6.0** included interlaboratory data for at least a subset of the substances evaluated. Using this data, the ability of the IRE test method to reproducibly identify ocular corrosives and severe irritants versus nonsevere irritants and nonirritants was evaluated using two approaches.

In the first approach, a qualitative assessment of reproducibility was conducted. In this evaluation, the individual laboratory *in vitro* ocular irritation classification for each substance was used to evaluate the extent of agreement among the participating laboratories in their ability to identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants. The reliability of the IRE test method was assessed separately for each study (i.e., publication) with multiple laboratory data reviewed in **Sections 4.0** and **5.0**.

Substances classified, based on IRE test data, as corrosive/severe irritants or nonsevere irritants/nonirritants were further classified by their *in vivo* rabbit eye test results, as determined within the GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) classification systems. Because the focus of this reliability assessment is on the interlaboratory reproducibility of IRE in identifying corrosives/severe irritants versus nonsevere irritants/nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants. For example, three laboratories could classify a chemical as a nonirritant and one laboratory could classify the same chemical as a moderate irritant. Within this analysis, this distribution of classification calls would be considered as 100% agreement between laboratories.

In the second approach, a quantitative assessment of reproducibility was determined. CVs where laboratory scores were available for substances tested were reported or determined. The reproducibility of the IRE test method was assessed for studies (i.e., publication) reviewed in **Sections 4.0** and **5.0** where individual testing laboratory data was available.

7.2.3.1 *Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System*

For this classification system (UN 2003), one study could be used to assess the interlaboratory reproducibility of the IRE test method (Balls et al. 1995). The four participating laboratories in this EC/HO validation study were in agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 35 (59%) of the 59 substances tested.

As shown in **Table 7-1:**

Table 7-1 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the GHS Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 75% Agreement Among Laboratories (%)	Substances with 50% Agreement Among Laboratories (%)
+/+	14	4	14 (100)	0 (0)	0 (0)
+/-	9	4	5 (56)	4 (44)	0 (0)
-/+	20	4	8(40)	3 (15)	9 (45)
-/-	14	4	6 (43)	8 (57)	0 (0)
?/-	1	4	1 (100)	0 (0)	0 (0)
?/+	1	4	1 (100)	0 (0)	0 (0)
TOTAL	**59**	**4**	**35 (59)**	**15 (25)**	**9 (15)**

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category 1); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category 2A, 2B) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), a GHS classification (UN 2003) could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of substances tested multiple times *in vitro.*

- All four participating laboratories agreed on the classification of 14 (100%) of the 14 substances that were GHS corrosives/severe irritants[1].
- Five (56%) of the nine substances classified according to the GHS based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by all four participating laboratories as nonsevere irritants (i.e., Category 2A and 2B irritants) or nonirritants whereas four of the nine substances (44%) had 75% agreement among the laboratories. The five substances incorrectly classified by all four laboratories were Captan 90 concentrate, dibenzoyl-L-tartaric acid, 2,5-dimethylhexanediol, 15% sodium lauryl sulfate, and sodium perborate tetrahydrate.
- Eight (40%) of the 20 substances classified according to the GHS based on *in vivo* rabbit eye data as nonsevere irritants or nonirritants were incorrectly classified by the four laboratories as corrosives or severe irritants. Of the 12 substances (60%) with discordant *in vitro* classification results among the four laboratories, three (15%) (ethyl acetate, iso-propanol, and methyl acetate) were incorrectly classified by three of the four laboratories and nine (45%) (acetone, 0.1% cetylpyridinium bromide, ethyl-2-methylacetoacetate, Fomesafen, Maneb, methylisobutylketone, n-octanol, polyethylene glycol 400, and toluene) were incorrectly classified by two of the four laboratories.

[1] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification.

- All four laboratories agreed on the classification of six (43%) of the 14 substances classified as GHS nonsevere irritants/nonirritants. Of the eight substances (57%) with discordant classification results, all eight substances (ammonium nitrate, butyl acetate, 4-carboxybenzaldehyde, dibenzyl phosphate, 2,6-dichlorobenzoyl chloride, tetra-aminopyrimidine sulfate, 3% trichloroacetic acid, and Tween 20) were correctly classified by three of the four laboratories.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), two of the 59 substances tested could not be classified according to the GHS classification scheme (UN 2003). All four laboratories were in agreement with the classification of one of these substances as a nonsevere irritant/nonirritant and of one substance as a corrosive/severe irritant.

7.2.3.2 *Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System*

The four participating laboratories in the EC/HO study (Balls et al. 1995) were in 100% agreement with the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 36 (61%) of the 59 substances tested. As shown in **Table 7-2:**

Table 7-2 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EPA Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 75% Agreement Among Laboratories (%)	Substances with 50% Agreement Among Laboratories (%)
+/+	18	4	18 (100)	0 (0)	0 (0)
+/-	7	4	4 (57)	3 (43)	0 (0)
-/+	20	4	8(40)	3 (15)	9 (45)
-/-	14	4	6 (43)	8 (57)	0 (0)
?/-	0	4	0 (0)	0 (0)	0 (0)
?/+	0	4	0 (0)	0 (0)	0 (0)
TOTAL	59	4	36 (61)	14 (24)	9 (15)

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or nonirritant (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of substances tested multiple times *in vitro*.

- All four participating laboratories[2] agreed on the classification of eighteen (100%) of the 18 substances that were EPA (EPA 1996) corrosives/severe irritants.
- Four (57%) of the seven substances classified according to the EPA (EPA 1996) based on *in vivo* rabbit eye data as corrosives or severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants (i.e., Category II or III) or nonirritants (Category IV). Three substances (43%) were shown to have discordant *in vitro* classification results among the four participating laboratories (Captan 90 concentrate, 2,5-dimethylhexanediol, and sodium lauryl sulfate [15%]). These substances were incorrectly identified by three of the four laboratories.
- Eight (40%) of the 20 substances classified according to the EPA (EPA 1996) based on *in vivo* rabbit eye data as nonsevere irritants or nonirritants were incorrectly classified by the four laboratories as corrosives or severe irritants. Of the 12 remaining substances (60%), three substances (15%) (ethyl acetate, iso-propanol, and methyl acetate) were incorrectly classified by three of the four laboratories and nine substances (45%) (acetone, cetylpyridinium bromide, ethyl-2-methylacetoacetate, Fomesafen, Maneb, methylisobutyl ketone, n-octanol, polyethylene glycol 400, and toluene) by two of the four laboratories.
- Six (43%) of the 14 substances classified according to the EPA (EPA 1996) based on *in vivo* rabbit eye data as nonsevere irritants/nonirritants were correctly classified by all four laboratories. All eight substances (57%) with discordant classification results (ammonium nitrate, butyl acetate, 4-carboxybenzaldehyde, dibenzyl phosphate, 2,6-dichlorobenzoyl chloride, tetra-aminopyrimidine sulfate, 3% trichloroacetic acid, and Tween 20) were correctly classified by three of the four laboratories.

7.2.3.3 *Interlaboratory Reproducibility Based on In Vitro Irritancy Classification Relative to the In Vivo Classification Using the EU Classification System*

A reliability analysis of the IRE test method in terms of the EU classification system could be conducted for the CEC (1991) collaborative study and the Balls et al. (1995) validation study.

In the CEC (1991) collaborative study, the participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 17 (81%) of the 21 substances tested.

As shown in **Table 7-3:**

[2] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification.

Table 7-3 Interlaboratory Variability of CEC Collaborative Study (1991) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EU Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 67%[2] Agreement Among Laboratories (%)	Substances with 33%[3] Agreement Among Laboratories (%)
+/+	5	3	3 (60)	1 (20)	1 (20)
+/-	0	3	0 (0)	0 (0)	0 (0)
-/+	2	3	2 (100)	0 (0)	0 (0)
-/-	8	3	6 (75)	2(25)	0 (0)
?/-	2	2[5]	2 (100)	0 (0)	0 (0)
?/+	4	3	4 (100)[6]	0 (0)	0 (0)
TOTAL	21	3[6]	17 (81)	3 (14)	1 (5)

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or severe irritant (Category R41); a "-" indicates that the substance was assigned an overall classification of nonirritant (Category R36); a "?" indicates that, due to the lack of appropriate *in vivo* data (i.e., insufficient dose volume), an EU classification (EU 2001) could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of substances tested multiple times *in vitro*.
[2] When two of three laboratories were concordant.
[3] When one of three laboratories was concordant.
[4] With the exception of the two (+/-) substances.
[5]Two of the three testing laboratories evaluated these two substances.
[6] One of the four substances was tested in two laboratories with severe classifications assigned.

- Three (60%) of five substances classified according to *in vivo* rabbit eye data as corrosives/severe irritants were identified correctly by all three laboratories[3]. One discordant substance (sodium dodecyl sulfate) was correctly classified by two of the three laboratories, and one (dibutyltin chloride) was correctly classified by one of three laboratories.
- Of the 21 substances evaluated, none were identified as false negative (i.e., as a corrosive/severe irritant *in vivo* and as a nonsevere irritant *in vitro*).
- Two of two substances (100%) were incorrectly classified as corrosives/severe irritants by all three laboratories (100%). There were no discordant substances.
- Six of eight (75%) substances were in complete agreement among laboratories for identification of nonsevere irritants/nonirritants. Two discordant substances (25%) (Brij 35 and 2-butoxyethylacetate) were identified as nonsevere irritants/nonirritants by two of the three testing laboratories.
- Both laboratories (only two of three laboratories tested these substances) agreed in the identification of two substances as nonsevere irritants/nonirritants (100%), although no *in vivo* classification could be assigned to these substances.

[3] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification.

- All three laboratories agreed in the identification of four substances as severe irritants (100%), although no *in vivo* classification could be assigned to these substances.

Using the Balls et al. (1995) validation data set, the participating laboratories were in 100% agreement with the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 37 (63%) of the 59 substances tested. As shown in **Table 7-4:**

Table 7-4 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EU Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 75% Agreement Among Laboratories (%)	Substances with 50% Agreement Among Laboratories (%)
+/+	12	4	12 (100)	0 (0)	0 (0)
+/-	6	4	3 (50)	3 (50)	0 (0)
-/+	18	4	7(39)	2 (11)	9 (50)
-/-	12	4	6 (50)	6 (50)	0 (0)
?/-	6	4	4 (67)	2 (33)	0 (0)
?/+	5	4	5 (100)	0 (0)	0 (0)
TOTAL	59	4	37 (63)	13 (22)	9 (15)

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or severe irritant (Category R41); a "-" indicates that the substance was assigned an overall classification of nonirritant (Category R36); a "?" indicates that, due to the lack of appropriate *in vivo* data (i.e., insufficient dose volume), an EU classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of substances tested multiple times *in vitro*.

- All four participating laboratories agreed on the classification of 12 (100%) of the 12 substances that were EU corrosives/severe irritants[4].
- Three (50%) of the six substances classified according to the EU (EU 2001) based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by all four laboratories as nonsevere irritants/nonirritants. Of the three substances (50%) with discordant *in vitro* classification results among the four participating laboratories, all three substances (Captan 90 concentrate, dibenzoyl-L-tartaric acid, and 2,5-dimethylhexanediol) were incorrectly classified by three of the four laboratories.
- Seven (39%) of the 18 substances classified according to the EU (EU 2001) based on *in vivo* rabbit eye data as nonsevere irritants/nonirritants were incorrectly classified by all four participating laboratories as corrosives/severe irritants. Of the 11 substances (61%) with discordant *in vitro* classification results among the four participating laboratories, two substances (44%), ethyl

[4] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification.

acetate and methyl acetate, were incorrectly classified by three of the four laboratories and nine (50%) were incorrectly classified by two of the four laboratories (acetone,γ-butyrolactone, 0.1% cetylpyridinium bromide, ethyl-2-methylacetaoacetate, fomesafen, methylisobutylketone, n-octanol, polyethylene glycol 400, and toluene).

- All four participating laboratories agreed on the classification of six (50%) of the 12 substances classified as EU (EU 2001) nonsevere irritants/nonirritants. Three of the four laboratories were in agreement for the six substances (50%) with discordant *in vitro* classification results (ammonium nitrate, 4-carboxybenzaldehyde, dibenzyl phosphate, tetra-aminopyrimidine sulfate, 3% trichloroacetic acid, and Tween 20).
- Four of six (67%) of substances were classified *in vitro* as nonirritants by all four laboratories, but could not be classified *in vivo* due to lack of sufficient data. Two of the six (33%) were classified as nonsevere irritants/nonirritants *in vitro* by three of the four laboratories.
- Five of five (100%) substances were classified *in vitro* as corrosives/severe irritants by all four laboratories, but could not be classified *in vivo* due to lack of appropriate data.

7.2.3.4 *Common Chemical or Product Classes Among Substances with Discordant Interlaboratory Results Using the GHS, EPA, and EU Classification Systems*

In the CEC (1991) study, GHS and EPA classifications were not available due to lack of *in vivo* rabbit eye data. Using the EU (EU 2001) classification system, four discordant substances that were incorrectly classified *in vitro* had no commonality with respect to chemical or product class. There were no false negative or false positive discordant substances in this analysis.

Twenty-four, 23, and 22, substances, respectively, for the GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) classification systems exhibited interlaboratory differences in *in vitro* classification in the Balls et al. (1995) study. Six esters, four alcohols, three carboxylic acids, and three ketones exhibited discordant results. Four substances (Captan 90 concentrate [pesticide], dibenzoyl-L-tartaric acid, 2,5-dimethylhexanediol [pesticide], and 15% sodium lauryl sulfate [surfactant]) were consistently found in at least two of three classification systems to be underpredicted with 75% concordance among the laboratories. Three substances (ethyl acetate, methyl acetate [acetates/solvents], and iso-propanol [alcohol/solvent]) were consistently overpredictive in all three classification systems with 75% concordance between laboratories. Nine substances (acetone, 0.1% cetylpyridinium bromide, ethyl-2-methylacetoaceate, Fomesafen, Maneb, methyl isobutylketone, n-octanol, polyethylene glycol 400, and toluene) were consistently found to be overpredictive with 50% concordance among the testing laboratories in at least two of the three classification systems. Solvent (nonaqueous water miscible and nonmiscible) was the product class appearing most frequently among all of these discordant substances. Eight of the discordant substances belonged to this product class. Surfactants/soaps (3) and pesticides (4) were other product classes for which discordant results were observed.

7.2.4 Coefficient of Variation Analysis

7.2.4.1 *CEC Collaborative Study (CEC 1991)*

Mean endpoint values (i.e., corneal opacity, corneal swelling, and fluorescein retention at one and four hours; one laboratory used a 1.25 hour endpoint) for each substance tested were provided from each of the three laboratories participating in the CEC (1991) study. These values were used to calculate the SD and %CV values for each IRE test method endpoint for each substance to provide a quantitative assessment of interlaboratory variability (**Table 7-5**)

Mean and median %CV values for the entire dataset were also calculated to provide an assessment of overall variability. Traditionally, mean/median %CV values of less than 35% have been considered satisfactory for biologically based test methods (Fentem et al. 1998). For IRE, there is moderate interlaboratory variability for each of the four endpoints, with fluorescein retention at four hours representing the largest %CV (59%) and corneal swelling at four hours representing the lowest %CV value (33%). When only severe irritants (EU Category R41[5] [EU 2001], based on *in vivo* data) are considered, the interlaboratory variability is lower for all endpoints. Corneal swelling at 1.25 hours retains the highest variability (CV of 37%) and 4-hour corneal opacity the lowest (CV of 16%). It should be noted that this analysis was performed without using a correction factor to normalize corneal swelling values, a practice that has been suggested if different depth measuring devices were used among the different laboratories (Prinsen M, personal communication). The overall median CV of the 4-hour corneal swelling values was 70% (40/57.3) of the mean, whereas all other parameters ranged from 47% (28/58.9) to 81% (43.0/53.3) of their respective means. The overall median CV of the 4-hour corneal swelling for severe irritants was 100% (35.5/35.4) of the mean with the other parameters ranging from 83% (30.5/36.6) to 99% (15.4/15.5) of their respective means. These values suggest that efforts to increase the interlaboratory reproducibility of the test method might be warranted.

There do not appear to be physicochemical characteristics that are common to most of the substances with the most variable responses (defined as > 100% CV in any of the endpoints). All of the substances in the CEC study were tested as liquids (some were diluted to concentrations used in the *in vivo* studies). Of nine substances with significant variability in at least one endpoint, there were no obvious chemical or product classes that appear to be responsible for the variability. Four substances (sodium fluorescein, glycerol, triethanolamine, and n-hexane) had variability in more than one endpoint.

[5] GHS classification (UN 2003) was not available for this dataset.

Table 7-5 Quantitative Evaluation of the Interlaboratory Variability of the IRE Test Method (CEC 1991)

Substance	CS(1)				FR (1)		CO(2)	
	1.25 Hour	1.25 Hour	4 Hour	4 Hour	4 Hour	4 Hour	4 Hour	4 Hour
	Mean	(%CV)	Mean	(%CV)	Mean	(%CV)	Mean	(%CV)
Acetic acid	20.7	56	40.7	30	2.33	25	1.70	25
Brij 35	7.67	38	12.3	25	0.87	93	0.50	141
Benzalkonium chloride	40.3	31	82.7	38	2.67	22	3.00	0.00
Dimethylsulfoxide	8.00	66	11.7	95	1.33	87	0.50	141
Sodium fluorescein	2.33	138	4.70	173	0.67	172	0.00	0.00
Glycerol	3.33	92	5.33	43	0.33	175	0.40	141
Triacetin	2.67	43	0.67	172	0.00	0.00	0.00	0.00
Mercury chloride	19.0	30	76.0	35	2.50	28	2.40	24
Silver nitrate	14.0	7.1	16.7	40	1.00	100	1.75	20
Sodium hydroxide	38.7	19	67.3	22	3.00	0.00	3.00	0.00
Toluene	9.00	22	10.7	30	1.73	37	0.60	141
Triethanolamine	3.33	148	6.33	97	0.07	165	0.00	0.00
n-Hexane	4.00	132	7.00	108	0.33	175	0.00	0.00
Chloroform	17.0	60	30.7	47	3.00	0.00	1.70	25
2-Methoxy ethanol	12.7	54	42.7	7.2	2.67	22	2.40	24
n-Butanol	31.3	6.7	60.3	16	3.00	0.00	2.50	28
Acetaldehyde	12.3	21	34.7	20	2.93	3.9	1.25	28
2-Butoxy ethylacetate	10.0	20	23.0	68	1.67	35	0.95	7.4
Sodium dodecylsulfate	15.7	47	24.0	61	2.07	78	1.40	40
Dibutyltin chloride	11.0	26	29.5	41	2.00	0.00	1.00	0.00
Tributyltin chloride	22.3	63	97.0	36	2.53	20	2.10	6.7
Mean for All Substances	14.5	53.3	32.6	57.3	1.7	58.9	1.3	37.7
Median for All Substances		43.0		40.0		28.0		24.0
Range for All Substances	2.3-40	6.7-148	0.7-97	7.2-173	0-3.0	0.0-175	0-3.0	0-141
Mean for Severe Irritants (EU)	22.5	36.6	56.5	35.4	2.5	22.1	2.0	15.5
Median for Severe Irritants (EU)		30.5		35.5		21.0		15.4
Range for Severe Irritants (EU)	11-40	19-63	24-97	20-61	2.0-3.0	0-78	1.0-3.0	0 -40

CO = Corneal opacity; CS = Corneal swelling; FR = Fluorescein retention, SD = Standard deviation; %CV = Percent coefficient of variation

[1]Substances listed in bolded italics are classified *in vivo* as severe irritants (Category 1) according to GHS (UN 2003).

7.2.4.2 *Balls et al. (1995)*

Mean endpoint values (i.e., corneal opacity and corneal swelling at 1 and 4 hours) for each substance tested were provided from each of the four laboratories participating in the EC/HO study. These values were used to calculate the standard deviation and CV for each IRE test method endpoint for each substance to provide a quantitative assessment of interlaboratory variability (**Table 7-6**). Mean and median %CV values for the entire dataset were also calculated to provide an assessment of overall variability. Traditionally, mean/median %CV values of less than 35% have been considered satisfactory for biologically-based test methods (Fentem et al. 1998; ICCVAM 2003). For IRE, there is moderate interlaboratory variability for each of the four endpoints, with corneal opacity at 1 hour representing the largest %CV (84%) with a range spanning 0 to 200% and corneal swelling at 4 hours representing the lowest %CV (53%) with a range of 10 to 118%. When only severe irritants (GHS Category 1[6], based on *in vivo* data [UN 2003]) are considered, the interlaboratory variability is lower for all endpoints, although corneal opacity at 1 hour retains the highest variability (47%CV) with a range of 0 to 200% and 4-hour corneal swelling the lowest (37%CV) spanning a range of 11 to 118%. The overall median of the 4-hour corneal opacity values was 68% (43.4/63.79) of the mean, whereas all other parameters ranged from 89% (74.6/84.1) to 93% (49.7/53.47) of their respective means. The overall median of the 4-hour corneal opacity for severe irritants was 83% (33.6/40.5) of the mean with the other parameters ranging from 87% (40.6/46.6) to 96% (35.5/36.9) of their respective means. These values suggest that efforts to increase the interlaboratory reproducibility of the test method might be warranted.

There do not appear to be physicochemical characteristics that are common to most of the substances with the most variable responses (defined arbitrarily as > 100%CV) in any of the endpoints). Of the 36 substances with significant variability in at least one endpoint, 17 are solids (of 19 tested) and 19 are liquids (of 40 tested). However, there are some chemical classes that predominate among the variable results with seven acetates/esters (of 7 tested), six surfactants (of 12 tested), six acids (of 6 tested), three heterocyclic compounds (of 6 tested), three alcohols (of 7 tested), and three pesticides (of 4 tested) represented among the 36 substances. However, in the absence of a larger dataset, the significance of these findings is not clear.

[6] One of these substances (sodium lauryl sulfate, 15%) is classified as R36 according to EU (EU 2001). Two other substances (cetylpyridinium bromide, 6% and dibenzoyl-L-tartaric acid) were not classified according to EPA system due to inadequate *in vivo* data with which to follow the EPA-specific classification rules (EPA 1996). Therefore, substances classified as severe irritants according to the GHS system (UN 2003) were used for this subanalysis in order to include the largest dataset.

Table 7-6 Quantitative Evaluation of the Interlaboratory Variability of the IRE Test Method (Balls et al. 1995)

Substance	CO 1 Hour Mean	CO 1 Hour (%CV)	CO 4 Hour Mean	CO 4 Hour (%CV)	CS 1 Hour Mean	CS 1 Hour (%CV)	CS 4 Hour Mean	CS 4 Hour (%CV)
1-Naphthalene acetic acid[1]	0.25	200	0.90	114	11.98	73	13.7	64
1-Naphthalene acetic acid, Na salt	1.00	115	2.68	18	57.03	51	107.6	45
2,2-Dimethylbutanoic acid	2.75	18	2.74	12	33.58	17	68.0	22
2,5-Dimethylhexanediol	0.33	142	0.42	120	15.30	87	16.4	79
2,6-Dichlorobenzoyl chloride	0.75	67	1.90	32	8.53	103	21.1	56
2-Ethyl-1-hexanol	0.25	200	1.43	35	10.7	44	20.3	17
4-Carboxybenzaldehyde	0.25	200	0.43	119	6.20	56	13.0	70
Acetone	0.43	119	1.05	105	15.3	78	31.9	95
Ammonium nitrate	0.00	0	0.00	0	7.30	43	10.2	111
Benzalkonium chloride (1 %)	0.93	90	2.43	28.0	23.9	23.0	52.8	48
Benzalkonium chloride (10%)	1.67	28	2.50	23.0	36.4	50	73.1	43
Benzalkonium chloride (5%)	1.33	71	3.00	0.00	32.3	40	99.2	23
Dibenzoyl-L-tartaric acid	1.00	141	1.90	60.0	18.2	118	24.5	70
Captan 90 concentrate	0.75	128	1.01	98.0	6.50	80	18.7	51
Cetylpyridinium bromide (0.1%)	0.00	0.00	0.00	0.00	14.7	47.0	19.8	50
Cetylpyridinium bromide (10%)	0.83	106	1.92	43	17.9	36	43.5	68

Substance	CO 1 Hour Mean	CO 1 Hour (%CV)	CO 4 Hour Mean	CO 4 Hour (%CV)	CS 1 Hour Mean	CS 1 Hour (%CV)	CS 4 Hour Mean	CS 4 Hour (%CV)
Cetylpyridinium bromide (6%)	0.58	88.0	1.75	43	21.4	41	32.0	31
Chlorhexidine	1.25	101	2.68	35	26.8	56	69.2	59
Cyclohexanol	1.08	77	2.50	23	24.3	41	82.1	26
Dibenzyl phosphate	0.50	115	1.08	64	9.5	44	16.4	55
Ethanol	1.72	45.0	2.58	20	26.8	60	52.6	18
Ethyl acetate	0.00	0.00	1.43	47	14.6	41	30.6	46
Ethyl trimethyl acetate	0.00	0.00	0.83	108	6.6	79	12.0	49
Ethyl-2-methylacetoacetate	0.42	120	1.68	50	16.3	68	21.2	67
Fomesafen	0.83	175	1.18	124	9.2	115	16.3	84
Gammabutyrolactone	0.25	200	1.67	63	21.4	19	38.3	26
Glycerol	0.00	0.00	0.33	145	7.7	40	7.6	47
Imidazole	2.50	23.0	2.75	18	44.8	11	74.7	11
Isobutanol	1.33	71.0	2.50	23	25.1	44	75.5	26
Isopropanol	1.34	68.0	1.92	51	16.0	70	35.8	57
L-aspartic acid	0.25	200	0.25	200	5.1	76	6.08	107
Maneb	1.00	141	1.00	115	24.0	82	26.6	87
Methyl acetate	0.50	115	1.59	77	15.1	28	30.6	43
Methyl cyanoacetate	0.08	200	0.66	138	5.0	29	6.9	21
Methyl ethyl ketone	0.92	91.0	2.41	18	21.2	30	61.2	34
Methyl isobutyl ketone	0.25	200	1.58	80	18.2	90	34.2	70
Methylcyclopentane	0.00	0.00	0.00	0	8.2	80	9.5	82
n-Butyl acetate	0.00	0.00	0.34	116	6.6	74	14.7	74
n-Hexanol	0.66	115	2.68	18	18.6	18	48.3	21
n-Octanol	0.00	0.00	1.45	36	11.8	48	21.7	34
Parafluoraniline	1.24	71	2.29	21	27.8	15	64.3	11
Polyethylene glycol 400	0.25	200	0.50	115	15.0	81	17.6	84
Potassium cyanate	0.00	0	0.00	0	5.2	59	5.3	113

Substance	CO 1 Hour Mean	CO 1 Hour (%CV)	CO 4 Hour Mean	CO 4 Hour (%CV)	CS 1 Hour Mean	CS 1 Hour (%CV)	CS 4 Hour Mean	CS 4 Hour (%CV)
Promethazine HCl	1.50	38	2.33	20	44.1	67	89.7	36
Pyridine	1.83	31	2.83	12	25.9	54	54.9	26
Quinacrine	0.00	0	0.18	200	7.1	82	81.0	89
Sodium hydroxide (1%)	0.99	72	2.75	18	50.2	22	93.5	26
Sodium hydroxide (10%)	2.93	24	4.00	0	101.6	13	138.3	18
Sodium lauryl sulfate (3 %)	0.00	0	0.50	115	9.8	37	15.4	35
Sodium lauryl sulfate (15 %)	0.08	200	1.33	63	16.3	21	23.4	10
Sodium oxalate	0.00	0	0.00	0	7.3	97	9.7	85
Sodium perborate	0.00	0	0.00	0	3.2	57	5.5	118
Tetraaminopyrimidine sulfate	0.75	128	0.75	128	4.3	129	10.3	98
Toluene	0.43	119	0.50	115	14.4	65	22.8	61
Trichloroacetic acid (3%)	0.68	70	0.75	128	8.1	34	18.4	72
Trichloroacetic acid (30%)	3.43	15	3.68	13	24.0	118	77.4	43
Triton X-100 (10 %)	0.67	141	2.33	20	27.1	51	56.8	64
Triton X-100 (5 %)	0.58	164	1.95	39	19.7	35	33.0	26
Tween 20	0.00	0	0.25	200	13.5	75	15.8	66

Substance	CO 1 Hour Mean	CO 1 Hour (%CV)	CO 4 Hour Mean	CO 4 Hour (%CV)	CS 1 Hour Mean	CS 1 Hour (%CV)	CS 4 Hour Mean	CS 4 Hour (%CV)
Mean for All Substances	**0.72**	**84.1**	**1.47**	**63.79**	**19.19**	**56.18**	**37.08**	**53.47**
Median for All Substances		**74.6**		**43.4**		**50.8**		**49.7**
Range for All Substances	**0-3.4**	**0-200**	**0-3.7**	**0-200**	**5-102**	**11-129_**	**6-108**	**10-118**
Mean for Severe Irritants (GHS)	**32.4**	**46.6**	**1.94**	**40.5**	**33.2**	**37.6**	**33.3**	**36.9**
Median for Severe Irritants		**40.6**		**33.6**		**36.0**		**35.5**
Range for Severe Irritants	**0-3.4**	**0-200**	**0-2.4**	**0-200**	**5-102**	**11-118**	**6-108**	**11-118**

CO = Corneal opacity; CS = Corneal swelling; SD = Standard deviation; %CV = Percent coefficient of variation
[1]Substances listed in bolded italics are classified *in vivo* as severe irritants (Category 1) according to GHS (UN 2003).

7.2.5 Additional Analysis of Interlaboratory Reproducibility

In the EC/HO validation study, Balls et al. (1995) determined the interlaboratory correlation between four specific IRE endpoints (corneal opacity at 1 and 4 hours; corneal swelling at 1 and 4 hours) as well as the summary endpoint generated by four independent laboratories. Correlation analyses were conducted for the total data set, along with specific subsets of substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients provided in **Table 7-7** (see **Appendix E** for all correlation coefficients derived from comparing each laboratory with every other laboratory).

Interlaboratory correlation coefficients varied considerably depending on the endpoint assessed and the subset of substances tested. In general, when the different endpoints were considered, the highest correlation and the most consistent data was produced with the 4-hour opacity and swelling measurements. Also, in general, compared to the individual 4-hour opacity and swelling measurements, the IRE summary score exhibited greater variability and a lower maximum correlation. The highest correlation was obtained for surfactants (0.696-0.853; 4-hour opacity, and 0.532-0.677; 4-hour swelling) and for liquids (0.402-0.759; 4-hour opacity, and 0.527-0.763; 4-hour swelling). For solids, the highest correlation was only 0.566 and the range of correlation values was increased considerably. Much of the discordance can be attributed to a single laboratory (laboratory b) for the entire range of substances. In general, there was good correlation between three of the four laboratories, including the lead laboratory. The other laboratories (laboratories c and d) contributed more to the discordance when the substances were solids or those insoluble in water.

7.3 Historical Positive and Negative Control Data

As noted in **Section 2.0**, positive controls have not been employed in the IRE test method publications or submitted data, and therefore, historical positive control data is not available. In addition, although negative/vehicle controls (isotonic saline) are traditionally run on at least one test eye with each experiment, these data have not been published and/or provided with data submitted for this BRD. Therefore, an analysis of historical negative control data also is not possible.

Table 7-7 Interlaboratory Correlation Ranges Determined for Various Subsets of Tested Substances in Balls et al. (1995)

Index Score	Interlaboratory Pearson's Correlation (r) of the *In Vitro* Data
Full set of substances (60)	
IREA-Mean Opacity Score, 1 Hour	0.407-0.502
IREB-Mean Opacity Score, 4 Hour	0.485-0.606
IREC-Corneal Swelling, 1 Hour	0.247-0.528
IRED-Corneal Swelling, 4 Hour	0.447-0.611
IRESUM-Summary Score	0.399-0.483
Chemicals soluble in water (30)	
IREA-Mean Opacity Score, 1 Hour	0.422-0.514
IREB-Mean Opacity Score, 4 Hour	0.341-0.516
IREC-Corneal Swelling, 1 Hour	0.246-0.492
IRED-Corneal Swelling, 4 Hour	0.329-0.552
IRESUM-Summary Score	0.471-0.560
Chemicals insoluble in water (18)	
IREA-Mean Opacity Score, 1 Hour	0.104-0.706
IREB-Mean Opacity Score, 4 Hour	0.422-0.730
IREC-Corneal Swelling, 1 Hour	0.177-0.762
IRED-Corneal Swelling, 4 Hour	0.342-0.763
IRESUM-Summary Score	0.156-0.502
Surfactants (12)	
IREA-Mean Opacity Score, 1 Hour	0.466-0.833
IREB-Mean Opacity Score, 4 Hour	0.696-0.853
IREC-Corneal Swelling, 1 Hour	0.204-0.690
IRED-Corneal Swelling, 4 Hour	0.532-0.677
IRESUM-Summary Score	0.513-0.666
Solids (20)	
IREA-Mean Opacity Score, 1 Hour	0.001-0.403
IREB-Mean Opacity Score, 4 Hour	0.231-0.564
IREC-Corneal Swelling, 1 Hour	-0.056-0.487
IRED-Corneal Swelling, 4 Hour	0.112-0.566
IRESUM-Summary Score	0.033-0.293
Solutions (14)	
IREA-Mean Opacity Score, 1 Hour	0.502-0.718
IREB-Mean Opacity Score, 4 Hour	0.657-0.763
IREC-Corneal Swelling, 1 Hour	0.157-0.564
IRED-Corneal Swelling, 4 Hour	0.240-0.686
IRESUM-Summary Score	0.631-0.770
Liquids (26)	
IREA-Mean Opacity Score, 1 Hour	0.197-0.595
IREB-Mean Opacity Score, 4 Hour	0.402-0.759
IREC-Corneal Swelling, 1 Hour	0.115-0.709
IRED-Corneal Swelling, 4 Hour	0.527-0.763
IRESUM-Summary Score	0.203-0.514

7.4 Conclusions

Evaluation of the intralaboratory repeatability and reproducibility of the IRE test method could not be conducted. Interlaboratory reproducibility was assessed based on a qualitative analysis (correct classification as a severe irritant or as a nonsevere irritant) of the individual laboratory test results obtained for the EC/HO validation study (Balls et al. 1995). However, it must be noted that the protocols for these studies were not always identical. This data suggested that the IRE test method may be generally reproducible with respect to identification of severe irritants (and ocular corrosives). For example, in the Balls et al. (1995) validation study, when *in vivo* data from four laboratories was assigned a regulatory classification and compared to irritancy defined using the IRE test method with decision criteria targeted for identification of severe irritants (i.e., Guerriero et al. 2004), 100% of the laboratories correctly identified the 14, 18, and 12 substances, respectively, tested as Category 1 GHS (UN 2003), Category I EPA (EPA 1996), or R41 EU (EU 2001) severe irritants. Discordance was greatest for false positives where only 45-83% of the substances were concordant among three of the four testing laboratories, and 45-50% were concordant among two of the four testing laboratories. By chemical class, the substances with the greatest levels of interlaboratory variability in all studies included alcohols, carboxylic acids, esters, and ketones. Solvent was the most common product class exhibiting a greater level of interlaboratory variability.

An evaluation of IRE interlaboratory variability using a CV analysis of corneal swelling, corneal opacity, and fluorescein retention also indicated generally reproducible results across laboratories when testing severe irritants (%CVs for severe irritants were approximately 40% for studies where the recommended protocol was not used). When all substances tested were considered, the %CV increased to 84%.

Based on the results from this limited dataset, the IRE test method appears to be generally reproducible among different laboratories with respect to the identification of severe irritants and false positives. However, there is not enough reliability data to draw definitive conclusions based on the limited available data. Reliability needs to be assessed using the standardized test method protocol (with all four ocular parameters) against an appropriate set of substances of varying levels of irritancy, physicochemical properties, chemical classes and product classes.

[This Page Intentionally Left Blank]

8.0 TEST METHOD DATA QUALITY

8.1 Adherence to National and International GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines, which are nationally and internationally recognized rules designed to produce high-quality laboratory records. GLPs provide a standardized approach to report and archive laboratory data and records, and information about the test protocol, to ensure the integrity, reliability, and accountability of a study (OECD 1998; U.S. EPA 2003a, 2003b; FDA 2003).

Based on the information available in the publications and from additional information provided (F. Guerriero, personal communication), it appears that Balls et al. (1995), Gettings et al. (1996), and Guerriero et al. (2004) conducted IRE studies in compliance with GLP guidelines. It could not be determined whether the IRE studies in CEC (1991) were conducted in accordance with GLP guidelines.

The *in vivo* reference studies used for Balls et al. (1995) appear to have adhered to GLP guidelines. Balls et al. 1995 used *in vivo* reference data from the ECETOC Eye Irritation Reference Data Bank (ECETOC 1992). These *in vivo* data were generated in studies carried out according to OECD Test Guideline 405 (OECD 1987) and following the principles of GLPs. The *in vivo* reference data from Guerriero et al. (2004) were also performed in accordance with GLP guidelines (F Guerriero, personal communication). Based on the available information, it could not be determined whether the *in vivo* reference data for the remaining test substances reported in Gettings et al. (1996) or CEC (1991) were obtained under GLP guidelines.

8.2 Data Quality Audits

Formal assessments of data quality, such as a quality assurance (QA) audit, generally involve a systematic and critical comparison of the data provided in a study report to the laboratory records generated for a study. No attempt was made to formally assess the quality of the *in vitro* IRE data included in this BRD or to obtain information about data quality audits from the authors of the IRE study reports. The published data on the IRE assay were limited to Draize (Balls et al. 1995; Gettings et al. 1996) or McDonald Shadduck (Guerriero et al. 2004) scoring of corneal opacity and/or area of involvement. Other measured parameters included scores for fluorescein penetration and description of endothelial integrity. Auditing these reported values would require obtaining the original data for each IRE experiment, which is not readily available.

An informal assessment of the IRE publications revealed limitations that complicate interpretation of the IRE data:

- *Incomplete substance information*: Some IRE study reports provided limited information about the substances tested. The CASRN, purity, and supplier of the test substances were not consistently reported. Thus, comparisons of data from different studies that evaluated test substances of the same chemical name must be interpreted with caution because of possible differences in purity and supplier of the test substances.
- *Data reporting*: Various scoring methods were utilized in the various reports, which makes it difficult to make comparisons between the studies or for compounds tested in different studies.
- *Methodology*: The methods were presented in varying levels of detail and completeness in the study reports.

Since the published data were not verified for their accuracy against the original experimental data, and the methods and data were presented in varying levels of detail and completeness, caution must be exercised when interpreting the analyses performed in **Sections 6.0** and **7.0**.

8.3 Impact of Deviations from GLP Guidelines

The impact of deviations from GLP guidelines was not evaluated for the reviewed IRE studies.

8.4 Availability of Laboratory Notebooks or Other Records

As noted in **Section 5.2**, the availability of notebooks or other records containing data from the reviewed IRE studies is unknown. Given the lack of availability of the original records, including the raw data for the studies used to evaluate the accuracy and reliability of the IRE test method in this document, the testing laboratory's summary judgment regarding the outcome of each study cannot be evaluated.

8.5 Need for Data Quality

Data quality is a critical component of the test method validation process. To ensure data quality, ICCVAM recommends that all of the data supporting validation of a test method be available with the detailed protocol under which the data were produced. Original data should be available for examination, as should supporting documentation, such as laboratory notebooks. Ideally, the data should adhere to national or international GLP guidelines (ICCVAM, 1997).

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Summaries of IRE Data from Published and Unpublished Studies

This section contains summaries of the available data from published or unpublished studies conducted using the IRE test method. In many of these reports, inadequate information on the substances tested (e.g., identity not specific) and/or on the results obtained from the *in vitro* or *in vivo* studies (e.g., qualitative but not quantitative IRE data, group mean but not individual *in vivo* animal scores) precluded an assessment of the performance of IRE. However, based on data received from contacting the authors or alternative sources (e.g., ECVAM), some substances included in these reports were used to assess the accuracy and reliability of IRE; these analyses are included in **Section 6.0**. This section provides a summary of reports (presented in alphabetic order by lead author) where such information was not available and the conclusions presented by the investigators. An explanation as to why the data presented in a report could not be used to independently assess the performance of IRE is provided. In addition, where applicable, an explanation as why some data could be used as part of the performance evaluation is provided.

9.1.1 Balls et al. (1995)

Under the auspices of the British Home Office and Directorate General XI of the European Commission, a validation study on proposed alternatives to the *in vivo* rabbit ocular toxicity test method was conducted. The goal of the evaluation was to identify at least one non-whole animal test method that could be proposed to regulatory authorities as a replacement for the currently accepted *in vivo* ocular toxicity test method. For the IRE test method, a total of 52 substances were evaluated in 60 tests in four laboratories. Four of the test substances were evaluated at two different concentrations and two substances were evaluated at three different concentrations. The ocular irritancy potential of the test substances were ranked in terms of MMAS (which ranged from 0 to 108). The test substances evaluated in the validation study were classified as acids (4), acyl halide (1), alcohols (9), aldehyde (1), alkalis (1), esters (6), heterocyclics (3), hydrocarbons (2), inorganic chemicals (4), ketones (3), organophate (1), pesticides (5), surfactants (6), and miscellaneous (6). *In vivo* data for 46 of the test substances, which were generated in compliance with OECD TG 405 (OECD 1987), was obtained from historical sources. *In vivo* rabbit eye data for 14 of the test substances were obtained from concurrent studies conducted in compliance with OECD TG 405 (OECD 1987).

Since the *in vivo* test results were expressed as MMAS, the data provided in this report could not be used to evaluate the accuracy of IRE for detecting ocular corrosives and severe irritants according to the GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) classification systems. However, using data provided by ECVAM, an evaluation was conducted of the ability of the IRE test method to identify severe ocular irritants or corrosives, as defined by the three classification systems (**Section 6.0)**, as well as to evaluate its interlaboratory reproducibility (**Section 7.0**).

The individual scores for each IRE test method endpoint were not included in the published report in tabular form. Rather, the study reports the relationship between each IRE test

method endpoint to the MMAS in graphic form for the entire set of test substances. The MMAS was chosen as the *in vivo* reference endpoint by the EC/HO working group and therefore, was the single *in vivo* endpoint included in the Balls et al. (1995) evaluation. A list of the 59 substances representing a wide-range of chemical classes and irritancy ranges tested in this study can be found in **Appendix B1**.

Spearman's rank correlation test and linear regression analysis were used to compare *in vivo* MMAS with irritancy in the IRE expressed as mean corneal opacity and mean corneal swelling, both measured at one and four hours. Spearman's rank correlation coefficients and Pearson's correlation coefficients were calculated for each participating laboratory for the entire test substance set, as well as for five subsets of test substances (water-soluble substances, surfactants, solids, solutions, and liquids). The ranges of the correlation coefficients for correlations between overall classification scores and MMAS that were obtained by each of the testing laboratories are presented in **Table 9-1**.

The resulting analysis showed that overall, the IRE test method (based on the Summary Score) was not highly predictive of the MMAS (Pearson's Correlation Coefficient: 0.40 to 0.48 for the full set of test substances). Correlations with individual *in vitro* endpoints (corneal opacity and swelling) versus the MMAS also were relatively low (r = 0.25 to 0.61). Subset analyses revealed some differences among specific groups of test substances with Pearson's Correlation Coefficients ranging from 0.31 to 0.56 for water-soluble test substances, 0.10 to 0.76 for water insoluble test substances, 0.20 to 0.85 for surfactants, 0 to 0.57 for solids, 0.16 to 0.73 for solutions, and 0.11 to 0.76 for liquids.

9.1.2 Chamberlain et al. (1997)

As part of the Organotypic Models Working Group, Chamberlain et al. (1997) reviewed IRE test method data submitted to the Interagency Regulatory Alternatives Group (IRAG) on the use of isolated eyes and ocular components used to predict eye irritation potential. The protocol for the IRE test method was a modification of that described by Burton et al. (1981). A total of 107 substances were evaluated using the IRE test method. The substances represented a wide range of chemical types. The majority of substances (89) had MAS values of 30 or less (and therefore considered mild to moderate irritants) and 13 substances had MAS values ranging from 31 to 55 (and therefore considered moderate to severe irritants). The five severe irritants had MAS values equal to or greater than 55 and produced > 15% corneal swelling (**Table 9-2**). Greater than 50% of substances with MAS values between 31 and 55 (n = 13) produced corneal swelling greater than 15% in the IRE test method. When all of the substances were considered, only 38% produced > 15% corneal swelling. A Pearson's correlation coefficient of 0.50 was obtained when the IRE test results were correlated against the *in vivo* rabbit eye test results, presented as MAS scores. Consistent with some of the previous reports considered in this section, corneal opacity was not a good predictor of *in vivo* irritancy. The authors concluded that the IRE test method is suitable for screening severely irritating substances before *in vivo* animal tests are conducted, but cautioned that relying solely on organotypic methods for evidence of lack of an eye irritation hazard was not warranted at the present time.

Table 9-1 *In Vitro*/*In Vivo* Correlation Coefficients from Balls et al. (1995)

Index Score	Pearson's Correlation Coefficient (r)	Spearman's Correlation Coefficient (r)
Full set of test substances (n = 59)		
IRE-Opacity, 1 Hour	0.407-0.502	0.316-0.510
IRE-Opacity, 4 Hours	0.485-0.606	0.451-0.606
IRE-Swelling, 1 Hour	0.247-0.528	0.166-0.515
IRE-Swelling, 4 Hours	0.447-0.611	0.364-0.624
IRE- Summary Score	0.399-0.483	0.473-0.603
Chemicals soluble in water (n = 30)		
IRE-Opacity, 1 Hour	0.422-0.514	0.238-0.377
IRE-Opacity, 4 Hours	0.341-0.516	0.226-0.440
IRE-Swelling, 1 Hour	0.305-0.492	0.329-0.552
IRE-Swelling, 4 Hours	0.329-0.552	0.293-0.511
IRE- Summary Score	0.471-0.560	0.311-0.426
Chemicals insoluble in water (n = 18)		
IRE-Opacity, 1 Hour	0.104-0.706	0.117-0.770
IRE-Opacity, 4 Hours	0.422-0.730	0.346-0.795
IRE-Swelling, 1 Hour	0.177-0.762	0.159-0.692
IRE-Swelling, 4 Hours	0.342-0.763	0.381-0.656
IRE- Summary Score	0.156-0.502	0.458-0.626
Surfactants (n = 12)		
IRE-Opacity, 1 Hour	0.466-0.833	0.486-0.855
IRE-Opacity, 4 Hours	0.696-0.853	0.623-0.828
IRE-Swelling, 1 Hour	0.204-0.690	0.007-0.720
IRE-Swelling, 4 Hours	0.532-0.677	0.504-0.746
IRE- Summary Score	0.513-0.666	0.613-0.839
Solids (n = 20)		
IRE-Opacity, 1 Hour	0.001-0.403	-0.056-0.373
IRE-Opacity, 4 Hours	0.231-0.564	0.130-0.534
IRE-Swelling, 1 Hour	-0.056-0.487	-0.182-0.504
IRE-Swelling, 4 Hours	0.112-0.566	-0.085-0.612
IRE- Summary Score	0.033-0.293	0.045-0.545
Solutions (n = 14)		
IRE-Opacity, 1 Hour	0.502-0.718	0.425-0.702
IRE-Opacity, 4 Hours	0.657-0.733	0.598-0.761
IRE-Swelling, 1 Hour	0.157-0.564	0.308-0.726
IRE-Swelling, 4 Hours	0.240-0.686	0.495-0.664
IRE- Summary Score	0.539-0.743	0.631-0.770
Liquids (n = 26)		
IRE-Opacity, 1 Hour	0.197-0.595	0.261-0.617
IRE-Opacity, 4 Hours	0.402-0.759	0.384-0.764
IRE-Swelling, 1 Hour	0.115-0.709	0.139-0.774
IRE-Swelling, 4 Hours	0.527-0.736	0.524-0.782
IRE- Summary Score	0.203-0.514	0.524-0.743

There was insufficient information in the IRAG report to assign GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) regulatory classifications to perform an accuracy analysis in this BRD. Furthermore, as the identity of the substances considered in the IRAG analysis were kept confidential and some of the data were likely to have been generated by studies considered elsewhere in this BRD, these data were not considered further.

Table 9-2 Relationship Between MAS *In Vivo* and the Ability to Cause More Than 15% Corneal Swelling *In Vitro* (Chamberlain et al. 1997)

MAS Range	N	Materials Causing >15% Corneal Swelling	
		Number of Substances	%
0-76 (all substances	107	41	38.3
≥ 55	5	5	100
31 to 55	13	7	53.8
≤30	89	29	32.6

9.1.3 Cooper et al. (2001)

Cooper and colleagues compared the IRE test method results on seven shampoo formulations to MAS values obtained from corresponding *in vivo* rabbit eye studies. The IRE protocol was modified from Burton et al. (1981) by inclusion of the evaluation of fluorescein penetration and histopathology.

The data generated in the study suggests that the IRE test method is useful for predicting the irritant potential of shampoo formulations that, in general, tend to produce mild to moderate rather than severe irritation (**Table 9-3**). In general, there appeared to be a concentration-dependent increase in irritancy for the shampoo formulations. Based on the IRE test results, one of five full strength shampoo formulations was overpredicted and one was underpredicted, when compared to *in vivo* rabbit eye test results. These authors also suggest, as demonstrated by Jones et al. (2001), that corneal swelling often occurs in the absence of corneal opacity.

Table 9-3 Comparison of IRE Test Method Results With *In Vivo* Data (Cooper et al. 2001)

Treatment[a]	IRE Irritancy Rating	*In Vivo* Irritancy Rating (MAS)
10% A	Moderate	No Test Data
10% B	Slight/Moderate	No Test Data
10% C	Moderate	No Test Data
10% D	Moderate	No Test Data
10% E	Slight/Moderate	No Test Data
10% F	Slight/Moderate	No Test Data
10% G	Very Slight/Slight	Mild (Predicted)
100% A	Moderate	Mild (14.3)
100% B	Moderate	Moderate (30.0)
100% C	No Test Data	Extreme (59.0)
100% D	Severe	Extreme (77.0)
100% E	Mild	Moderate (Predicted)
100% F	Moderate	Moderate (Predicted)
100% G	No Test Data	Mild (Predicted)

[a] Shampoo formulations (A is base formula, B is base with 1.5% ingredient X, C is base with 3.0% ingredient X, D is base with 6.0% ingredient X, E and F are reference controls, and G is a baby shampoo).

There was insufficient information in this report to assign a GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) regulatory classification for the accuracy analysis in **Section 6.0**.

9.1.4 Gettings et al. (1996)

As part of the Phase III CTFA validation study, Gettings et al. (1996) evaluated 25 surfactant-based personal care formulations using the IRE test method. *In vitro* responses were measured using either corneal swelling in the IRE (referred to in the report as the Rabbit Enucleated Eye Test or REET I) or scored according to severity (score ranging from 0 to 3) of the REET I corneal swelling results (referred to as the Rabbit Enucleated Eye Test II). Substances with *in vitro* scores greater than 18.6 for the REET I analysis or a score greater than 1.0 for the REET II were classified as irritants. Substances that did not meet these criteria were designated nonirritants. There was no attempt to distinguish severe irritants from moderate or mild irritants. The *in vitro* data obtained in the IRE were compared to *in vivo* rabbit eye test data obtained using the Draize scoring method (Draize et al. 1944) expressed as MAS or were classified as irritant or nonirritant based on the FHSA regulatory classification (FHSA 1988). The results of these analyses are shown in **Table 9-4.**

Table 9-4 The Results of the CTFA Evaluation of *In Vitro* Alternatives to the Draize Primary Eye Irritation Test (Phase III): Surfactant-Based Formulations (Getting et al. 1996)

Phase III Substances	REET I[a] (Percent of Control)	REET II[b] (Irritancy Score, 0-3)
Classified as Irritants by FHSA[b] Criteria		
HZQ	7.5	0.3
HZG	29.5	2.0
HZN	37.7	2.7
HZD	20.3	1.0
HZB	24.8	1.7
HZV	25.6	1.3
HZW	23.9	1.7
HZU	36.7	3.0
HZC	21.2	1.0
HZF	14.3	1.0
HZA	32.1	2.3
HZL	36.2	2.7
HZR	13.4	1.0
HZK	36.4	2.7
HZX	20.9	1.3
HZI	28.6	1.7
HZS	33.3	2.3
HZY	18.6	1.0
Classified as Non-Irritants by FHSACriteria		
HZH	7.7	0.0
HZZ	2.8	0.0
HZT	2.5	0.0
HZI	16.3	1.0
HZP	25.0	1.7
HZM	26.0	1.7
HZE	6.4	0.0

[a] Modified from Burton et al. (1981) using 20 µL test material at 10 second intervals for 1 minute. Represents percentage increase in mean corneal thickness compared to control. Score ≥ 18.6 considered irritant.
[b] Modified from Burton et al. (1981) using 20 µL test material at 10 second intervals for 1 minute. Represents a classification into one of four groups (0 to 3) based on the degree of corneal swelling. Score ≥ 1.0 considered irritant.

For the FHSA classification system (FHSA 1988) for identification of irritants, an accuracy of 80% (20/25), a sensitivity of 83% (15/18), a specificity of 71% (5/7), a false positive rate of 29% (2/7), and a false negative rate of 17% (3/18) were obtained for REET I. For the REET II test, an accuracy of 84% (21/25), a sensitivity of 94% (17/18), a specificity of 57% (4/7), a false positive rate of 43% (3/7), and a false negative rate of 6% (1/18) were obtained. The authors also calculated a separation index for each substance tested for REET I. The separation index represents the rate at which the *in vitro* endpoint (corneal swelling) and MAS do not agree. The mean of separation indices was 0.463 ± 0.026 (a standard error based on a Monte Carlo estimate of variability). A value of 1.0 indicates complete concordance with the *in vivo* outcome.

In vivo data from the Gettings et al. (1996) report were not used as provided, because FHSA classification does not include a severe irritant category. However, *in vivo* data were received from the CTFA in response to an *FR* notice that allowed for an accuracy analysis. This analysis is provided in **Section 6.0**.

9.1.5 Guerriero et al. (2004)

Guerriero and colleagues obtained data using the IRE test method protocol as described in **Section 5.1.3**. The study evaluated the response of 44 substances (30 pharmaceutical process materials, 14 ECETOC compounds) in the IRE test method. *In vitro* data were recorded as scores for corneal opacity and area, corneal swelling, scores for fluorescein intensity and area, and observations of epithelial integrity (pitting, mottling, sloughing). Test substances that produced an *in vitro* corneal opacity x area score ≥ 3, a fluorescein uptake intensity x area score ≥ 4, swelling ≥ 25, or produced corneal epithelial damage were designated as severe irritants. Test substances that did not exceed this score were classified as nonsevere irritants. Data obtained from concomitant *in vivo* rabbit eye irritation tests on these substances were classified for ocular irritancy according to the EU classification system (EU 2001). Using these multiple decision criteria, the authors correctly identified 100% (n = 15) of R41 substances. The authors concluded that use of the IRE assay supports the concept of the 3Rs (replacement, reduction, and refinement) and that the IRE assay is a valuable and practical screening tool to identify substances that are severe eye irritants.

In their 2004 report, Guerriero et al. provided a EU regulatory classification (EU 2001) for the *in vivo* data. Upon request, the authors kindly provided the individual animal *in vivo* response data, which permitted classification according to the GHS (UN 2003) and EPA (EPA 1996) classification systems. These results were used in the accuracy analysis described in **Section 6.0**.

9.1.6 Jacobs and Martens (1990)

Using an ultrasonic pachymeter, corneal swelling (expressed as a percentage) derived from the mean increase in corneal thickness produced in response to application of 34 test substances of varying irritancy levels at 4, 24, 48, and 72 hours *in vivo* was compared to that obtained in the Isolated Eye Test (IET) at two and four hours. Linear correlation between corneal swelling *in vitro* and *in vivo* tests at four hours was slight with r = 0.77. However, when test substances that produced epithelial opacity (notably acids) were omitted from the evaluation, the correlation between *in vitro* corneal swelling at two and four hours improved

to r = 0.91, when compared against the mean *in vivo* corneal swelling measured at 24, 48, and 72 hours (EU 2001). Linear correlation between mean percentage corneal opacity scores and mean corneal swelling was satisfactory with r = 0.89. In this study, a percentage increase in corneal swelling of 55% obtained in isolated rabbit eyes over two and four hours, corresponds to the limit of an irritant classification using the EEC (1984) regulatory classification system. When this criterion was applied to all of the substances excluding those that produced epithelial swelling, one false positive and no false negatives were observed.

9.1.7 Jacobs and Martens (1989)

The ultrasonic pachymeter has been shown to be more accurate than the optical pachymeter (Salz et. al. 1983; Thornton 1985) and has the advantage that it is easy to handle and transport, has rapid measuring speed, requires less operating skill, is not restricted to measurement of central corneal thickness and can be used in the presence of severe opacity (Jacobs and Martens 1988). Thirty-four chemically diverse test substances with a wide range of irritant responses were tested in the *in vivo* rabbit eye test for corneal swelling using an ultrasonic pachymeter and this data was compared to mean Draize corneal opacity, erythema, chemosis, and iritis scores. Mean corneal swelling at 24, 48, and 72 hours was determined. The eye irritation protocol described in EEC (1979) was used for the assay. Linear correlation between mean percent corneal swelling measurements and corneal opacity scores was r = 0.94. Linear correlation between mean percent corneal swelling measurements and chemosis scores were r = 0.87. Erythema scores were not linear with percent corneal swelling measurements, due to a limited erythema scale and the need for a minimum degree of erythema to be produced before corneal swelling can be measured. Mean percent corneal swelling at 24 and 72 hours using ultrasonic pachymetry were comparable to 24-hour optical pachymetry measures, while ultrasonic measures were lower than optical pachymetry measures at 72 hours. The authors suggest that addition of a quantitative and sensitive measure such as ultrasonic pachymetry to *in vivo* rabbit eye testing for ocular toxicity would reduce intra- and interlaboratory variability.

9.1.8 Jacobs and Martens (1988)

The ultrasonic pachymeter was used to measure corneal swelling (expressed as a mean percentage and standard deviation) in response to 11 substances tested in the enucleated rabbit eye test and compared to mean percentage corneal swelling results obtained on these substances in the enucleated rabbit eye test methods previously reported by Burton et al. (1981) using an optical pachymeter and by Köeter and Prinsen (1985) using an ultrasonic pachymeter. Mean percentage corneal swelling was determined 240 min after test substance application to four enucleated rabbit eyes after a 10 sec exposure to the test substance followed by saline rinse. Although the measured results were not identical, good correlation with an r-value of 0.98 was obtained by plotting a linear regression of 240-minute ultrasonic data and the optical pachymeter data from Burton et al. (1981). Standard deviations for both test methods were of the same order of magnitude, with the exception of acetone and ethanol which were higher for the ultrasonic pachymeter. Corneal opacity scores at 240 minutes compared to ultrasonic pachymetry with a Spearman's rank correlation coefficient of 0.91 ($p < 0.0005$). Using a mean epithelial damage score produced a less satisfactory correlation

(0.78; $p < 0.005$). Careful assessment of epithelial integrity in response to the applied test substances and to the ultrasonic pachymeter itself, revealed that the pachymeter did not significantly contribute to epithelial damage observed in response to the test substances. Higher values for the ultrasonic pachymeter against strongly irritating materials such as allyl alcohol, 1N sodium hydroxide, and butanol might be related to the fact that optical pachymetry units are not linear with swelling or that increased corneal opacity resulted in a concomitant decrease in refractivity of the cornea.

9.1.9 Jones et al. (2001)

Jones and colleagues published a study comparing ten shampoo formulations and seven conditioner formulations using five alternative test methods, including the IRE. The shampoos were tested at both 100% and 10% concentrations. The investigators modified the original Burton et al. (1981) IRE test method to include evaluation of fluorescein retention and evaluation of the epithelium. The investigators found generally good agreement between the irritancy ratings of the shampoo and conditioner formulations based on IRE data and their *in vivo* irritancy rating based on historical data. Eight of the 17 formulations classified as moderate irritants based on *in vivo* rabbit eye test results were either classified correctly or overpredicted, but never underpredicted (i.e., no false negatives were identified). A single severe ocular irritant formulation was correctly predicted by the IRE. However, for most test substances, corneal opacity alone was not as predictive as corneal opacity combined with corneal swelling and histology. Histology scoring appears to be responsible for some of the overpredicted classification, since a maximum number of layers lost rather than an average was used. For example, in cases where there was a wide range of responses of cell layers lost (e.g., two to seven), use of an average value instead of the maximum would have reduced the overall score. Furthermore, the conditioners tended to be overpredicted more frequently than the shampoos, perhaps because they contained predominately cationic surfactants versus the anionic and amphoteric surfactants contained in the shampoo formulations. The authors concluded that the data supports continued use of the IRE test method as an alternative to the *in vivo* rabbit eye irritation test with recognition that it can overpredict the irritancies of some formulations.

There was insufficient data provided in this report to assign GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) classifications for the tested formulations to perform an accuracy analysis in **Section 6.0**.

9.1.10 Koëter and Prinsen (1985)

A total of 34 substances were evaluated using the IRE test method and the data were compared to *in vivo* rabbit eye data obtained in the Draize test (**Table 9-5**). In this report, the test substances are indicated by code and therefore the substance names are unknown. However, physicochemical properties, including pH values, for some substances were provided. A mixture of hydrophilic (14) and hydrophobic (11) liquid substances and nine solid substances with pH values ranging from 1.8 to 13.5 were tested. In this assay, the Burton et al. (1981) protocol was modified to include fluorescein penetration and histology.

Table 9-5 Comparison of IRE *In Vitro* Irritancy Grades to *In Vivo* Rabbit Eye Test Irritancy Classifications (Koëter and Prinsen 1985)

Test Substance	Irritancy Grade	
	***In Vitro*[a]**	***In Vivo*[b]**
1	Not Irritant	Not Irritant
2	Slight	Slight
3	Slight	Slight
4	Moderate/Severe	Severe
5	Slight	Slight
6	Severe	Severe
7	Slight	Severe
8	Severe	Severe
9	Slight	Not Irritant
10	Negligible	Not Irritant
11	Not Irritant	Not Irritant
12	Not Irritant	Not Irritant
13	Moderate	Moderate
14	Slight	Not Irritant
15	Moderate	Not Irritant
16	Not Irritant	Not Irritant
17	Severe	Severe
18	Slight	Slight
19	Negligible	Not Irritant
20	Not Irritant	Not Irritant
21	Slight	Slight
22	Negligible	Not Irritant
23	Negligible	Not Irritant
24	Negligible	Not Irritant
25	Severe	Severe
26	Not Irritant	Not Irritant
27	Negligible	Not Irritant
28	Not Irritant	Not Irritant
29	Not Irritant	Not Irritant
30	Slight	Severe
31	Slight	Slight
32	Negligible	Not Irritant
33	Moderate	Not Irritant
34	Severe	Severe

[a] Based on overall Irritancy Rating
[b] Based on Draize score according to FDA guidelines (FDA 1980)
For identification of severe irritants, the accuracy was 91% (31/34), sensitivity was 63% (5/8), specificity was 93% (26/28), the false positive rate was 7% (2/28), and the false negative rate was 38% (3/8).

Corneal opacity was scored and corneal swelling was calculated based on the percentage increase in corneal thickness at each time point relative to a preapplication measurement, but modified with respect to the inclusion of the additional parameters -- histological assessment of the cornea and fluorescein penetration. Based upon averaging the final scores of all four *in vitro* endpoints, an overall Irritancy Rating was assigned. A comparative analysis of the IRE test results and the Draize rabbit eye test scores indicates that 28 of the 34 substances (82%) had similar irritancy ratings *in vitro* and *in vivo*. In general, the irritancy ratings were

predictive throughout the range of irritancy with a few exceptions. Two substances (6%) were underpredicted and four substances (12%) were overpredicted. Importantly, the two underpredicted substances were classified as severe ocular irritants *in vivo* on the basis of persistence of adverse effects and not the severity of the effect. The authors conclude that the IRE test method is a useful and sensitive test system for the evaluation of ocular irritation. A performance analysis on the reported data for identification of severe irritants indicated that the accuracy was 91% (31/34), sensitivity was 63% (5/8), specificity was 93% (26/28), the false positive rate was 7% (2/28) and the false negative rate was 38% (3/8).

There was insufficient information in this report to assign GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) regulatory classifications to perform an accuracy analysis in **Section 6.0**.

9.1.11 Lewis et al. (1994)

Lewis and colleagues published a report on the use of an *in vitro* test battery as a prescreen in the assessment of ocular irritancy. The authors describe a trypan blue exclusion assay using a human myeloid cell line as an initial screening test for severe irritants based on cytotoxicity. Test substances that produced < 15% cytotoxicity were tested *in vivo* using the rabbit eye test method while substances that produced > 15% cytotoxicity were tested using the IRE test method. In the IRE test method, if a substance produces less than 15% corneal swelling, one animal is tested *in vivo* since there is little likelihood of a severe irritant response. Those test substances producing greater than 15% corneal swelling are likely to be severe irritants; therefore, only one animal is tested initially using the low volume eye test in which the quantity dosed is 0.01 mL or 0.01 g. A total of 93 substances were evaluated using this tiered *in vitro* approach.

Among these 93 substances, a complex fiber formulation and a research agrochemical were classified as false negatives. Eight false positives were identified. Using nonparametric analysis, it was concluded that the majority of severe eye irritants were correctly predicted *in vitro*, with a sensitivity (ability to predict severe irritants) of 83% and a specificity (ability to identify less than severe irritants) of 90%. The authors concluded that although 10 of 11 severe eye irritants were predicted correctly using the IRE test method and 11 of 12 severe eye irritants were predicted by the trypan blue exclusion assay, the incidence of false positive responses in each of the assays still precludes their routine use as complete replacements for the *in vivo* rabbit eye test. However, the authors added that the *in vitro* battery assay approach does reduce the number of animals used and is clearly superior to reliance on skin testing data as an indicator of potential ocular effect. Using this approach, the authors report a reduction of 85% in the number of laboratory animals treated in the traditional *in vivo* rabbit eye test.

There was insufficient information in this publication to assign GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) regulatory classifications for the accuracy analysis in **Section 6.0**.

9.1.12 Price and Andrews (1985)

Price and Andrews evaluated the *in vivo* predictive accuracy of 60 substances using the IRE test method. The 60 substances included 25 industrial chemicals and 32 formulations (three

unformulated agrochemicals, 14 formulated lubricating oils and 18 formulated agrochemicals). The results were presented as a ratio of the *in vitro* prediction of irritancy with an *in vivo* classification expressed as a percentage. In this study, the Burton et al. (1981) protocol was modified to include evaluation of fluorescein penetration. Corneal thickness measurements along with evaluations of corneal appearance were recorded at regular intervals for up to five hours. Fluorescein penetration was recorded at four hours, if damage was present. Irritancy criteria for the *in vivo* eye test were based on OECD guidelines (OECD 1983). The scoring system for determination of severe irritancy *in vitro* was based on the time for corneal swelling to equal or greater than 20% (Grade IV, maximum). Lesser grades were assigned if it took longer to achieve this level of swelling (two hours, Grade III; five hours, Grade II, or less than 20% swelling in five hours, Grade I, minimal). Using these decision criteria, the results demonstrated that 10 (83%) of the 12 *in vivo* Class IV (severe) irritants and 33 (97%) of the 34 Class I (nonirritants or very mild) irritants were correctly identified by the IRE test method. For the detection of severe irritants only, a retrospective performance analysis indicated that the accuracy was 97% (58/60), sensitivity was 83% (10/12), specificity was 100% (48/48), the false positive rate was 0% (0/48) and the false negative rate was 17% (2/12).

There was insufficient information in this publication to assign GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) regulatory classifications for the accuracy analysis in **Section 6.0**.

9.1.13 Whittle et al. (1992)

In an interlaboratory trial of the IRE test method, Whittle and colleagues studied the ocular effect of 27 substances (17 liquids and 10 solids) representing a variety of chemicals and surfactants using the IRE test method. A modification of the IRE protocol described by Burton et al. (1981) was used that included an assessment of fluorescein retention and an evaluation of epithelial cell erosion. For two laboratories, the exposure duration (ten seconds) was the same as that proposed by Burton; in the third laboratory, the exposure duration was increased to one minute. The two laboratories that used the ten-second exposure protocol were able to separate severe/moderate from the mild eye irritants. *In vivo* irritancy was rated as severe, moderate/severe, moderate, slight/moderate, or slight, using in-house historical data on the *in vivo* rabbit eye test.

For the majority of test substances, evaluation of corneal swelling with a ten second exposure was a better indicator of irritancy than corneal opacity. For example, for the 17 liquid substances tested, all seven moderate to severe irritants induced corneal swelling of greater than 11% in both laboratories. However, corneal opacity was induced by only two of the seven-moderate/severe substances in both laboratories and by another substance in only one of the two laboratories. For the ten solid substances tested, corneal swelling was >12.5% for the three moderate to severe irritants in both laboratories, while corneal opacity was induced by two of three moderate to severe irritants and only in one of two laboratories. Evaluation of results from the 60-second exposure did not appear to provide additional benefit in identifying severe irritants. The investigators concluded that the IRE test method was useful for separating moderate to severe eye irritants from the milder eye irritants. However, it was also clear from the study that corneal opacity alone was not predictive of mild/moderate or moderate irritants using a ten-second exposure. The consistency of rating of irritancy

between laboratories was considered excellent for liquids, but was less impressive for solid materials.

There was insufficient information in this study to conduct an accuracy analysis as described in **Section 6.0**.

9.1.14 York et al. (1982)

York and colleagues published a report describing preliminary findings of an *in vitro* test for the assessment of eye irritancy in consumer products. A modification of the Burton et al. (1981) protocol was used in which evaluation of fluorescein penetration and histopathology were included. Eleven test substances with a span of irritancy ranging from no effect to very severe ocular damage (as reported in literature) were evaluated. The authors compared their *in vitro* irritancy ratings (mild to severe) to an *in vivo* Irritancy Grade (1-10; 10 being the most severe) described by Carpenter and Smyth (1946). Of 10 substances graded using the Carpenter and Smyth scale, three substances rated severe *in vitro* had *in vivo* grades of 10, 9 and 8, respectively, and were correctly predicted. Allyl alcohol was rated moderate/severe *in vitro*, assigned a five (moderate) on the *in vivo* scale, and therefore overpredicted. Toluene was underpredicted *in vitro* as negligible/slight whereas it had a scale of 7 (moderate/severe) *in vivo*. Overall, the authors conclude that the IRE test method is a valid model to use as a screening procedure for strong irritants.

There was insufficient information in this report to assign a GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) classification for the accuracy analysis in **Section 6.0**.

9.2 Data Received in Response to the ICCVAM *Federal Register* Notice or from Study Authors

An *FR* notice (Vol. 69, No. 57, pp. 13859-13861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original IRE test method data and *in vivo* reference data, was published on March 24, 2004. In addition, authors of published IRE studies were contacted to request original IRE data and *in vivo* reference data. In response to the *FR* notice, Guido Jacobs of the Institute for Hygeine and Epidemiology (Brussels, Belgium) and Dan Marsman of Proctor and Gamble (P&G; Cincinnati, Ohio) submitted reports of IRE test method data and *in vivo* rabbit eye test data.

9.2.1 Jacobs and Martens (January 1987)

Twenty-one substances were tested in the *in vivo* rabbit eye test (EEC 1979) and results were obtained for erythema, edema, corneal opacity, iritis, pain response, damage of the corneal epithelium, healing, and corneal swelling. This *in vivo* data was compared to the enucleated eye test of Burton et al. (1981) using the same set of substances. Mean percentage corneal swelling was determined in three rabbits over 24, 48, and 72 hours. Mean percentage corneal swelling in the enucleated eye test was obtained over 0.5, 1, 2, 4, and 5 hours. *In vitro* corneal swelling with various *in vivo* endpoint results correlated with corneal opacity ($r = 0.92$), erythema ($r = 0.91$), and percent fluorescein retention ($r = 0.94$). Correlation between mean percentage corneal swelling at four hours and the mean calculated over all observation times (24, 48, and 72 hours) was not as good ($r = 0.82$). Erythema appeared to be the most

sensitive indicator of ocular damage, and some degree of erythema was required before corneal opacity or chemosis were triggered. No correlation between pain response and production of ocular lesions was found. Test substances could be divided into two groups, one in which corneal swelling was increasing at five hours and one in which it had reached a maximum level by five hours. When *in vivo* clinical observations are considered (i.e., corneal opacity, erythema, chemosis, and iritis scores), the first group represents moderate to severe ocular irritants, whereas the latter group represents mild to moderate ocular irritants. The authors concluded that the enucleated eye test is a valid screening method for ocular irritation, although eye irritation classification cannot be based on the results of percentage corneal swelling alone or based on evaluation of a relatively small set of test substances.

9.2.2 Jacobs and Martens (May 1987)

An ultrasonic pachymeter was used to measure the percentage corneal swelling using the enucleated eye technique described by Köeter and Prinsen (1985) on the irritancy of 11 test substances reported by Burton et al. (1981) using optical pachymetry. Pachymetry data from one enucleated rabbit eye per test substance at 240 min was compared to the same substance tested in three enucleated rabbit eyes performed after 5, 30, 60, 120, 240, and 300 min. Using the 240 min readings on the four rabbit eyes evaluated with the ultrasonic pachymeter, a good correlation of $r = 0.98$ with the optical data was obtained. In addition, corneal swelling correlated well with corneal opacity scores at 240 min from Burton et al. (1981) with a Spearman rank correlation coefficient of $r = 0.91$ ($p = <0.0005$). Disadvantages of the optical pachymeter include changing refractive index by stromal swelling and a nonlinear correlation between actual and apparent (as viewed by the angle of the optical glass plate) corneal thickness. The study reported that the ultrasonic technique was a considerable improvement over the optical technique in: 1) simplicity of use, 2) short measuring time with ability to measure multiple eyes at each time point, 3) 10-fold increase in resolution, 4) wider range of corneal swelling is covered, since measurement is not hampered by corneal opacity, 5) measurement possible at all sites on corneal surface, 6) subjective aspects of optical pachymeter are not an issue with the probe tip of the ultrasonic instrument. One potential issue is damage to the epithelium by contact with the probe tip, although no adverse effects were observed in the study.

9.2.3 Proctor and Gamble (P&G) Submission from Drs. Daniel Marsman and Karen Acuff

9.2.3.1 *Summary of P&G Confocal Ocular Test Method*

The method of evaluation and scoring of the ocular toxicity of test substances used by P&G is substantially different from that used by many other investigators. The major difference is that confocal microscopy is used to determine the depth of corneal injury in addition to the area of involvement using a low volume eye test (LVET). This published methodology has been applied mainly to the testing of surfactant-based products (Jester et al. 1996; Maurer et al. 1996, 1997, 1998; Jester et al. 1998).

P&G has optimized this experimental methodology for use in the IRE (referred to as the *Ex Vivo* Rabbit Eye Test (ExRET) by P&G. P&G developed a Depth of Injury (DOI) method of evaluating the area and depth of corneal injury that is particularly important in evaluating an ocular response to surfactant-based substances. This measurement is obtained by staining the

eyes with Syto 10®, a fluorescent nucleic acid stain that penetrates cell membranes and labels all cells. Dead Red® is a cell-impermeant nucleic acid stain that labels only cells with compromised membranes. Measurement of the depth of corneal penetration is based on the depth at which no further staining of dead cells (as evidenced by dead cell staining) is observed and only live cells are present. The Normalized Depth of Injury (NDI) is the lone endpoint in the ExRET test method and is measured after 30-second exposure to the test substance using measurements in five regions of the cornea (center and four corresponding quadrants). The NDI is calculated as the mean of these five regions of the cornea divided by the overall corneal thickness (measured as the distance between the endothelial membrane and the basement membrane). The NDI is expressed as a percentage and is calculated by dividing the measured depth of injury by the overall corneal thickness and multiplying by 100. The experimental mean of NDI values for five eyes is expressed as a percentage. The final reported value is the average NDI obtained in three separate experiments. Liquid test substances are generally tested neat or may be diluted in water. One rabbit eye is treated for 30 seconds with 10 µL of D-MEM without phenol red containing 0.3% AlbuMax and 1% Dextran as a negative control. Five rabbit eyes are treated for 30 seconds with 10 µL of test substance. Two rabbit eyes are treated with the positive control for 30 seconds. Assays are conducted at room temperature and the eyes are rinsed with phosphate-buffered saline (PBS) to remove the test substance. Each test substance is tested three times for a total of 15 eyes per test substance, six eyes for the positive control and three for the negative control. A valid negative control response has an NDI = 0, and the NDI of the positive control should be within two standard deviations of the historical mean positive control.

9.2.3.2 *P&G Data*

P&G submitted data from the ExRET. Irritancy data obtained in the ExRET assay was compared to *in vivo* rabbit eye data obtained using confocal microscopy *in vivo*. In some studies, ExRET irritancy data was compared to data obtained using either conventional histopathology of LVET-treated tissues or a standard LVET *in vivo* rabbit model. Products tested included surfactants, general chemicals, surfactant-based dishwashing products and bleach-containing laundry additive products. Summarized NDI measurements and/or histopathology with predicted irritancy categories were presented in tabular and graphical form for each test substance. The data provided allowed for the development of an ExRET prediction model that contains: 1) a definition of the specific purposes for which the test was conducted; 2) definition of all possible results that may be obtained; 3) an algorithm to convert each test result into a prediction of the toxic effect of interest; and 4) the probability of the accuracy of the prediction for three irritancy categories (slight, mild/moderate, or severe).

The irritancy of anionic, nonionic and cationic surfactants as determined by confocal microscopy *in vivo* and *in vitro* is shown in **Table 9-6**. There is a general agreement in the assigned irritancy classification between *in vivo* data and the ExRET test method. The irritancy ratings assigned to three anionic, three nonionic and four cationic surfactants (including two severe irritants) by *in vivo* and ExRET test methods were in agreement. For the set of ten general substances (**Table 9-7**) tested *in vivo*, three (8% sodium hydroxide, 12% hydrogen peroxide and 15% hydrogen peroxide) were classified as severe irritants. Of these, all three were underpredicted as mild/moderate irritants by the ExRET test method.

Three substances (cyclohexanol, p-fluoroaniline, and formaldehyde) were overpredicted *in vitro*. In **Table 9-8**, the irritancy results from the LVET test method, conventional histopathology, and the ExRET test method *in vitro* are compared. The LVET irritancy ratings for three products, LDL659, LDL298, and LDL645 were based on MAS of 45.9, 50.3, and 53 and ratings of moderate, moderate and severe were assigned, respectively. Using histopathology, a level of mild/moderate was assigned to all three formulations, which was an underprediction when compared to LVET.

Table 9-6 Irritancy of Surfactant-Based Products Using P&G *In Vivo* and ExRET *In Vitro* Confocal Microscopy Test Methods

Test Substance	Conc (%)	Irritancy Rating (Confocal Microscopy Test Method)	
		In Vivo (n)[1]	ExRET *In Vitro* (n)
Anionic Surfactants			
Sodium lauryl sulfate	5	Slight (24)	Slight (50)
Sodium linear alkyl benzene sulfonate	35	Mild/Mod (43)	Mild/Mod (75)
Sodium alkyl ethoxylate sulfate	42.75	Mild/Mod (20)	Mild/Mod (90)
Nonionic Surfactants			
Polyoxyethylene glycol monoalkyl ether	100	Slight[2]	Slight (75)
Polyoxyethylene sorbitan	100	Slight[2]	Slight (75)
Alkyl E7(avg)ethoxylate	99	Mild/Mod (23)	Mild/Mod (50)
Cationic Surfactants			
3-Isotridecyloxypropyl-bis(polyoxyethylene) ammonium chloride	100	Slight (24)	Slight (75)
3-Decyloxypropyl-bis(polyoxyethylene) amine	100	Mild/Mod (6)	Mild/Mod (75)
Alkylbenzyldimethylammonium chloride	100	Severe (5)	Severe (40)
Cetyltrimethylammonium chloride	100	Severe (15)	Severe (45)
Cetyltrimethylammonium chloride	75	NT	Severe (25)
Cetyltrimethylammonium chloride	50	NT	Mild/Mod (25)
Cetyltrimethylammonium chloride	25	NT	Severe (25)
Cetyltrimethylammonium chloride	10	NT	Severe (50)

[1]Represents the total number of eyes used.
[2]n value was not available at time of submission.
NT = Not tested; Conc = Concentration; Mod = Moderate

Table 9-7 Irritancy of General Chemicals Using P&G *In Vivo* and ExRET *In Vitro* Confocal Microscopy Test Methods

Test Substance	Conc (%)	Irritancy Rating (Confocal Microscopy Test Method)	
		***In Vivo* (n)[1]**	**ExRET *In Vitro* (n)**
Acid			
Acetic acid	3	Slight (26)	Mild/Mod (75)
Acetic acid	10	Mild/Mod (32)	Mild/Mod (75)
Alkali			
Sodium hydroxide	2	Slight (26)	Mild/Mod (75)
Sodium hydroxide	8	Severe (20)	Severe (75)
Bleach			
Sodium perborate monohydrate		Slight (26)	Mild/Mod (75)
Sodium hypochlorite		Slight (26)	Mild/Mod (75)
Hydrogen peroxide	6	NA	Slight (25)
Hydrogen peroxide	10	NA	Slight (75)
Hydrogen peroxide	12	NA	Severe (25)
Hydrogen peroxide	15	NA	Severe (75)
Alcohol			
Cyclohexanol		Severe (31)	Mild/Mod (75)
Aromatic amine			
p-Fluoroaniline		Severe (33)	Mild/Mod (75)
Ketone			
Acetone		Slight (55)	Slight (55)
Aldehyde			
Formaldehyde Old	37	Severe (24)	Slight (75)
Formaldehyde New[2]	37	NT	Mild/Mod (25)

[1]Represents the total number of eyes tested.
[2]Includes zone of dead cells in calculation of NDI
NA - Data was not available at time of submission.
NT = Not tested; Conc = Concentration; Mod = Moderate

Table 9-8 Irritancy of Surfactant-Based Liquid Dishwashing Formulations Using LVET and Histopathology *In Vivo* and P&G ExRET Confocal Microscopy *In Vitro* Test Method

Product Name	*In Vivo*		*In Vitro*
	LVET MAS/DTC[1]	Histopathology[2]	Ex RET (n^3)
LDL659	Moderate	Mild/Mod	Slight (75)
LDL298	Severe	Mild/Mod	Mild/Mod (75)
LDL645	Severe	Mild/Mod	Mild/Mod (75)

[1]Maximum Average Score (MAS) and Days to Clear (DTC). LDL659 had a MAS of 45.9 clearing in 7 days. LDL298 and LDL645 had MAS values of 50.3 and 53, respectively, and cleared in 21 days.
[2]Conventional histopathology
[3]Represents the total number of eyes tested (usually multiples of 15 eyes/test article from three experiments).
LVET = Low volume eye test; Mod = Moderate

However, the histopathology ratings *in vivo* were in agreement for two of the three formulations and one of the three (LDL659) was underpredicted by the ExRET test method. For bleach-containing laundry additives using the same battery of test methods (**Table 9-9**), two of the four test substances (Peroxi694 and Peroxi695) were underpredicted *in vitro*. Another substance (Hypo686) was overpredicted. In general, the ExRET test method appears to be optimized for evaluation of surfactant-based chemicals, but was not optimized for evaluation of test substances from general chemical classes or from other formulation-based product classes.

Table 9-9 Irritancy of Bleach-Containing Laundry Additive Products Using LVET *In Vivo* and P&G ExRET *In Vitro* Confocal Microscopy Test Methods

Product Name	*In Vivo*		*In Vitro*
	LVET MAS/DTC[1]	Histopathology[2]	Ex RET (n^3)
Peroxi694	Moderate	Mild/Moderate	Slight (75)
Peroxi695	Moderate	Mild/Moderate	Mild/Moderate (75)
Hypo686	Severe	Mild/Moderate	Mild/Moderate (95)
Hypo580	Severe	Mild/Moderate	Mild/Moderate (90)

[1]Maximum Average Score (MAS) and Days to Clear (DTC).
[2]Conventional histopathology
[3]Represents the total number of eyes tested (usually multiples of 15 eyes/test article from three experiments).
LVET = Low volume eye test

[This Page Intentionally Left Blank]

10.0 ANIMAL WELFARE CONSIDERATIONS

10.1 How the IRE Test Method Will Refine, Reduce, or Replace Animal Use

ICCVAM promotes the scientific validation and regulatory acceptance of new methods that refine, reduce, or replace animal use where scientifically feasible. Refinement, Reduction, and Replacement are known as the "Three Rs" of animal protection. These principles of humane treatment of laboratory animals are described as:

- refining experimental procedures such that animal suffering is minimized
- reducing animal use through improved science and experimental design
- replacing animal models with nonanimal procedures (e.g., *in vitro* technologies), where possible (Russell and Burch 1992)

The IRE was initially developed as an organotypic, *in vitro* assay for the detection of severe eye irritants to avoid the testing of such substances in live animals (Burton et al. 1981). The IRE test method reduces animal use when eyes are obtained from rabbits raised for food or by obtaining them from rabbits sacrificed after use in other laboratory procedures that do not adversely affect the eye. The IRE test method is a refinement of the *in vivo* rabbit eye test in that the animals are sacrificed prior to application of the test substance and, therefore, the animals do not experience pain and suffering when an ocular irritant is directly applied to the eye. Furthermore, since the IRE test method was adapted from the Draize *in vivo* eye irritation test method specifically to reduce the need for live animals for ocular irritation testing, pain and suffering of the animals is eliminated and the overall number of animals needed for ocular toxicity screening is reduced.

10.2 Requirement for the Use of Animals

Although rabbits are required as a source of corneas for this organotypic assay, only rabbits sacrificed for food or used for other laboratory purposes are typically used as eye donors (i.e., no live animals are used in this assay).

[This Page Intentionally Left Blank]

11.0 PRACTICAL CONSIDERATIONS

Several issues are taken into account when assessing the practicality of using an *in vitro* test method in place of an *in vivo* test method. In addition to reliability and accuracy evaluations, assessments of the equipment and supplies needed for the *in vitro* test method, level of personnel training, costs of the *in vitro* test method, and time to complete the method are necessary. This information provides additional information as whether the time, personnel cost, and effort required to conduct the test method are considered reasonable

11.1 Transferability of the IRE Test Method

Test method transferability addresses the ability of a method to be accurately and reliably performed by different, competent laboratories (ICCVAM 2003). Issues of transferability include laboratories experienced in the particular type of procedure, and otherwise competent laboratories with less or no experience in the particular procedure. The degree of transferability of a test method affects its interlaboratory reproducibility.

11.1.1 Facilities and Major Fixed Equipment

If standard laboratory rabbits are to be used to provide the eyes for the IRE test method, then a standard animal housing facility approved by IUCAC and approved by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) and in accordance with the Animal Welfare Act (7 U.S.C. 2131 et. seq.) or through equivalent accreditation/ legislation in other countries is needed. This same facility would be needed to conduct in-house *in vivo* rabbit eye tests. Such facilities require strict adherence to animal welfare considerations with controlled temperature and humidity, cage size and construction, feed and watering requirements, and are likely to be available in any toxicology-testing laboratory that involves standard animal testing. The IRE test method does not require an animal facility unless the animals are purchased specifically for use in the assay. However, holding cages or boxes may be required for temporary storage of live animals. The capital or fixed item equipment required for the IRE test method include a slit-lamp (e.g., Haag-Streit) with a depth-measuring device (Haag-Streit #1 attachment) or ultrasonic pachymeter, a vibration-free table for the slit-lamp observation, and a water-jacketed, Perspex superfusion chamber with black-walled cells to hold a sufficient number of eyes (Burton et al. 1981).

11.1.2 General Availability of Other Necessary Equipment and Supplies

Noncapital equipment includes a water bath and peristaltic pump to recirculate the water in the superfusion jacket at a rate of approximately 4 liters/minute for maintenance of a temperature in the chamber cells of 32 ± 1.5°C, an additional peristaltic pump for saline infusion at a rate of 0.1 to 0.4 mL/minute, Perspex or stainless steel eye holders that fit into the superfusion chamber, stainless steel saline drip tubes, and surgical equipment for enucleation and fine dissection (scissors, forceps). Other items such as syringes, weighing boats, physiological salt solutions, fluorescein solution, sodium pentobarbital and other items are readily available for purchase commercially.

Similarly, the remaining equipment and supplies necessary for conducting the *in vivo* rabbit eye test are readily available in most toxicity testing laboratories or could be readily obtained from any of a number of scientific laboratory equipment vendors.

11.2 Training Considerations

Training considerations are defined as the level of instruction needed for personnel to conduct the test method accurately and reliably (ICCVAM 2003). Evaluation of the level of training and expertise needed to conduct the test method reliably and accurately, as well as the training requirements needed to ensure that personnel are competent in the test method, are discussed below.

11.2.1 Required Level of Training and Expertise Needed to Conduct the IRE Test Method

The most important difference between the *in vivo* and *in vitro* assays is the training required for administration of anesthetic for euthanasia of the rabbits and for enucleation and dissection of the eyes in the IRE test method. Although procurement of animals and administration of anesthetic to and dissection of animals at necropsy is standard practice in a toxicology-testing laboratory, proper training is required to understand shipment requirements, proper storage of the eyes if received from a vendor, or actual enucleation and dissection of the eye in a manner that prevents loss of intraocular pressure. Personnel familiar with the use of state-of-the-art procedures should train the laboratory personnel conducting the experiment. A training video or other visual media to provide guidance on the development of endpoints may be considered for use.

Once the Perspex superfusion apparatus and associated equipment is set up and running in a laboratory, minimal training is needed to place the enucleated eyes in the holders without damaging the cornea or affecting intraocular pressure, to control the temperature in the superfusion cells using the water bath, and to control the temperature and drip rate of saline flowing over the isolated cornea in the IRE test method. Some additional training in maintenance and changing of peristaltic tubing may be required.

To carry out the IRE test method, additional training principally involves the ability to measure and/or score the appropriate ocular parameters (i.e., corneal opacity and area of involvement, corneal thickness and swelling, fluorescein retention or penetration, epithelial cellular effects). Corneal opacity and area measurements and/or observations are performed with a slit-lamp and are similar to those performed *in vivo*. Personnel experienced in the state-of-the-art use of this equipment should be used to train new personnel in the use of the instrument, as well as in corneal observation and scoring methods. However, iridal and conjunctival observations needed for the *in vivo* test method are not required for the *in vitro* test method, since these tissues are removed or inoperative in the isolated eye due to lack of perfusion or muscular activity. Measurement of corneal thickness, calculation of corneal swelling and fluorescein retention are performed both *in vivo* and *in vitro* in some laboratories as additional endpoints to the Draize system. Training for the *in vitro* IRE test method is therefore no more complex than that for the *in vivo* assay. In fact, those trained to perform the *in vivo* test method could easily adapt to the *in vitro* assay

In general, personnel performing the IRE test method should be as proficient as possible. The trainers should insure that new laboratory personnel carry out their *in vitro* ocular testing appropriately, particularly when using the slit-lamp for observation and measurement. Personnel should demonstrate proficiency in the ability to procure laboratory animals and work with live animals if necessary, to administer anesthetic, to perform dissection procedures such as enucleation with reasonable speed while keeping the eyes free of corneal damage during the process. Personnel should be able to maintain the *in vitro* superfusion testing apparatus in an appropriate state by regulation of the temperature in the holding cells and the flow rate and temperature of the saline drip. Benchmark and standard ocular irritants with varying degrees of severity that represent various types of chemical substances should be scored by the trainee. The irritation scores obtained by the trainee should approximately match those obtained by someone trained in state-of-the-art techniques. For example, surfactants could produce a different type of corneal opacity than alcohols or acids and bases (e.g., diffuse rather than punctate lesions) and the testing personnel should be trained to understand the differences, particularly in how these various types of lesions are scored. Furthermore, the laboratory personnel should be proficient in applying fluorescein solutions to the eye and in scoring the degree of penetration using benchmark or standard irritants. Additionally, laboratory personnel involved in the IRE test method should demonstrate proficiency in standard laboratory procedures such as preparation and handling of solutions, weighing solids, sterile technique if required (e.g., media preparation), safe laboratory procedures, safe and appropriate storage practices, and other standard laboratory practices.

11.3 Cost Considerations

The current cost for a GLP compliant IRE assay (without the inclusion of a concurrent positive control) at SafePharm Laboratories, Ltd. (United Kingdom) is approximately $1070 per test substance (Guest R, personal communication). In comparison, a GLP-compliant EPA OPPTS Series 870 Acute Eye Irritation test (EPA 1996) in the rabbit ranges from $765 for a three day/three animal study up to $1665 for a 21 day/three animal study at MB Research Laboratories (MB Research laboratories, personal communication).

11.4 Time Considerations

Use of the IRE test method would significantly reduce the time needed to assess the ability of a test substance to induce ocular corrosivity or severe irritancy, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days. Depending upon the severity of ocular effects produced by a test substance, the method can be extended for up to 21 days. Comparatively, the standard IRE test method can be completed, from the onset of treatment, in about four hours.

[This Page Intentionally Left Blank]

12.0 REFERENCES

ASTM. 1999. Standard practice for conducting an interlaboratory study to determine the precision of a test method. ASTM E691-92. In: Annual Book of ASTM Standards. Philadelphia, PA:American Society for Testing and Materials.

Balls M, Botham PA, Bruner LH, Speilman H. 1995. The EC/HO international validation study on alternatives to the Draize Eye Irritation Test. Toxicol In Vitro 9(6):871-929.

Barkman R, Germanis M, Karpe G, Malmborg AS. 1969. Preservatives in drops. Acta Opthalmol. 47:461-475.

Berdasco N, Gilbert K, Lacher J, and Mattsson J. 1996. Low rate of severe injury from dermal and ocular irritation tests and the validity of using fewer animals. Journal of the American College of Toxicology. 15:177-193.

Berta A. 1992. Chapter 17. In: Enzymology of the tears. (Berta A. ed.). Boca Raton:CRC Press, 285-295.

BLS. 2004. US Dept. of Labor, Bureau of Labor Statistics, Case and Demographic Characteristics for Work-related Injuries and Illnesses Involving Days Away From Work. Available: http://www.bls.gov/iif/oshcdnew.htm [accessed 26 October 2004].

Bruner L. 1992. Chapter 7: Ocular irritation. In: In Vitro Toxicology Testing (Frazier J, ed). New York:Marcel Dekker, Inc., 149-190.

Bruner LH, Evans MG, McPherson JP, Southee JA, Williamson PS. 1998. Investigation of ingredient interactions in cosmetic formulations using isolated bovine corneas. Toxicology In Vitro 12:669-690.

Burton ABG. 1972. A method for the objective assessment of eye irritation. Fd Cosmet Toxicol 10:209-217.

Burton ABG, York M, Lawrence RS. 1981. The *in vitro* assessment of severe eye irritants. Food and Cosmetic Toxicol 19:471-480.

Butscher P. 1953. Beitrag zur therapie von augenschadigunen durch thioglykolsaur bei der herstellung der sogenannten kaltwelle. Klin Monatsbl Augenheilkd. 122:349-350.

Calabrese EJ. 1983. Dermatotoxicity: Predictive Models. In: Principles of Animal Extrapolation. New York:Wiley Interscience.

Carpenter CP, Smyth HF. 1946. Chemical burns of the rabbit cornea. Am J Ophthal. 29:1363-1372.

Cater K, Patrick E, Harbell J, Merrill J, Schilcher S. 2004. Comparison of in vitro eye irritation potential by BCOP assay to erythema scores in human eye sting test of surfactant-based formulations. [Abstract] Toxicol Sci 78:1307.

Chamberlain M., Gad SC, Gautheron P, Prinsen MK. 1997. Organotypic Models for the Assessment/Prediction of Ocular Irritation. Food and Chemical Toxicology 35:23-37.

Clean Air Act Amendments of 1990. 1990. Public Law 101-549.

Commission of the European Communities (CEC). 2001. Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test. Doc. XI/632/91/V/E/1/131/91 Part I and II.

Cooper KJ, Earl LK, Harbell J, Raabe H. 2001. Prediction of the ocular irritancy of prototype shampoo formulations by the isolated rabbit eye (IRE) test and bovine corneal opacity and permeability assay (BCOP) assay. Toxicology In Vitro 15:95-103.

CPSC. 1984. Animal Testing Policy. *Fed Reg* 49:22522-22523.

CPSC. 1995. Test for eye irritants. 16CFR1500.42.

CPSC. 2003. Commercial practices. Hazardous Substances and Articles; Administration and Enforcement Regulations. 16 CFR 1500.42.

CPSC. 2004. Definitions. Hazardous Substances and Articles; Administration and Enforcement Regulations. 16 CFR 1500.3.

Curren RD, Harbell JW. 1998. *In vitro* alternatives for ocular irritation. Envir Health Persp 106:485-492.

Dalbey W, Rodriguez S, Wilkins K, Cope C. 1993. Reducing the number of rabbits in eye and skin irritancy tests. Journal of the American College of Toxicology 12:347-357.

DeSousa D, Rouse A, Smolon W. 1984. Statistical consequences of reducing the number of rabbits utilized in eye irritation testing: Data on 67 petrochemicals. Toxicology and Applied Pharmacology 76:234-242.

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 Exp Ther 82:377-390.

Dua HS, Azuara-Blanco A. 2000. Limbal stem cells of the corneal epithelium. Survey of Ophth 44(5):415-425.

Duane TD. 1949. The steady state of corneal hydration. Am J Ophth 32:203-207.

EC. 2004. Manual of Decisions for Implementation of the Sixth and Seventh Amendments to Directive 67/548/EEC on Dangerous Substances (Directives 79/831/EEC and 92/32/EEC). European Chemicals Bureau, IHCP, European Commission DG-JRC, report n. EUR 20519 EN, updated version of July 2004. Available: http//ecb.jrc.it/new-chemicals [accessed 14 October 2004].

ECETOC. 1992. Eye irritation reference chemicals data bank. Technical Report No. 48. European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), Brussels.

ECETOC. 1998. Eye Irritation Reference Chemicals Data Bank (Second Edition). Technical Report No. 48 (2). (ECETOC). Brussels, June 1998.

EEC. 1979. Annex V, part B of the Council Directive 79/831/EEC (19 September 1984) concerning the methods for the determination of toxicity (84/449/EEC). Off J Eur Comm. N°. (L251), 94.

EEC. 1983. Annex VI, part IID of Council Directive 79/831/EEC (16 September 1983). Concerning the guide to the classification and labeling of dangerous substances and preparations; criteria for the choice of phrases indicating special risks (R-phrase) and safety advice (S-phrases) (83/467/EEC). Off. J Eur Commun. 26 (L257), 1.

EEC. 1984. Acute Toxicity – Eye Irritation. In Directive 67/548 (6th adaption); Annex V, Part B: Methods for the Determination of Toxicity. Official Journal of the European Community 27 L251, 109.

EPA. 1996. Label Review Manual: 2nd Ed. EPA 737-B-96-001. Washington, DC: US Environmental Protection Agency.

EPA. 1998. Health Effects Test Guideline, OPPTS 870.2400. Acute Eye Irritation. EPA 712-C-98-195. Washington, DC: US Environmental Protection Agency.

EPA. 2003a. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR 792.

EPA. 2003b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and Rodenticide Act. 40 CFR 160.

Eskes C, Bessou S, Bruner L, Curren R, Harbell J, Jones P, KreilingR, Liebsch M, McNamee, Pape W, Prinsen M, Seidle T, Vanparys P, Worth A, Zuang V. 2005. Eye Irritation. In: Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Perspectives (Eskes C, Zuang V, eds.). ATLA 33 (Supplement 1): 47-81.

Estable JL. 1948. The ocular effect of several irritant drugs applied directly to the conjunctiva. Am J Opthalmol. 31:837-844.

EU. 1992. Commission Directive 92/69/EEC of 31 July 1992 adapting to technical progress for the seventeenth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. Official Journal of the European Communities L255:1-333.

EU. 2001. Commission 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 labeling of dangerous substances. Official Journal of the European Communities L 255:1-333.

FDA. 1980. Code of Federal Regulations. FED method for testing primary irritant substances. 16 CFR 1000.1 paragraph 1500.41, p 294.

FDA. 2003. Good Laboratory Practices for Nonclinical Laboratory Studies. 21 CFR 58.

Federal Environmental Pesticide Control Act of 1972. 1972. Public Law 92-516.

FHSA. 1964. Federal Hazardous Substances Act. Public Law 86-613.

Federal Hazardous Substances Act (FHSA), CFR. 1988. Title 16, Section 150042: Federal Hazardous Substances Act Regulations, pp 350-351, US Government Printing Office, Washington, DC.

Federal Insecticide, Fungicide and Rodenticide Act of 1947. 1947. Public Law 80-102.

Fentem J, Archer GEB, Balls M, Botham PA, Curren RD, Earl LK, Esdaile DJ, Holzhütter, Liebsch M. 1998. The ECVAM International Validation Study on in vitro tests for skin corrosivity. 2. Results and evaluation by the Management Team. Toxic In Vitro 12:483-524.

Fox DA, Boyes WK. 2001. Toxic responses of the ocular and visual system. In: Casarett and Doull's Toxicology, the Basic Science of Poisons (Klaassen CD, ed.). 3rd ed. New York:McGraw-Hill, 3rd edition, 565-595.

Food, Drug and Cosmetic Act of 1938. 1938. Public Law 75-717.

Freeberg FE, Nixon GA, Reer PJ, Weaver JE, Bruce RD, Griffith JF, Sanders, LW 3d. 1986. Human and rabbit eye responses to chemical insult. Fundam Appl Toxicol 7:626-634.

Friedenwald JS, Hughes WF, Hermann H. 1946. Acid burns of the eye. Arch Ophth. 35:98-108.

Gartner S. 1944. Blood vessels of the conjunctiva. Arch Opthalmol. 36:464-471.

Gautheron P, Giroux J, Cottin M, Audegond L, Morilla A, Mayordomo-Blanco L, Tortajada A, Haynes G, Vericat JA, Pirovano R, Tos EG, Hagemann C, Vanparys P, Deknudt G,

Jacobs G, Prinsen M, Kalweit S, Spielmann H. 1994. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) assay. Toxicol In Vitro 8(3):381-392.

Gettings SD, Teal JJ, Bagley DM, Demetrulias JL, DiPasquale LC, Hintze KJ, Rozen MG, Weise SL, Chudowski M, Marenus KD, Pape WJW, Roddy M, Schnetzinger R, Silber PM, Glaza SM, Kurtz PJ. 1991. The CTFA Evaluation of alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test (Phase 1) hydro-alcoholic formulations; (Part 2) Data analysis and biological significance. *In Vitro* Toxicol 4:(4):247-288.

Gettings SD, Dipasquale LC, Bagley DM, Casterton PL, Chudkowski M, Curren RD, Demetrulias JL, Feder PI, Galli CL, Gay R, Glaza SM, Hintze KL, Janus J, Kurtz, PJ, Lordo A, Marenus KD, Moral J, Muscatiello M, Pape WJW, Renskers KJ, Roddy MT, Rozen MG. 1994. The CTFA evaluation of alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase II) Oil/water emulsions. Fd Chem Toxicol 32(10):943-976.

Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, Curren RD, Demetrulias JL, Dipasquale LC, Earl LK, Feder PI, Galli CL, Glaza SM, Gordon VC, Janus J, Kurtz PJ, Marenus KD, Moral J, Pape WJW, Renskers KJ, Rheins LA, Roddy MT, Rozen MG, Tedeschi JP, Zyracki J. 1996. The CFTA evaluation of alternatives program: An evaluation of *in vivo* alternatives to the Draize primary rabbit eye irritation test. (Phase III) Surfactant-based formulations. Fd Chem Toxicol 34:79-117.

Grant WM. 1974. Toxicology of the eye. In: Toxicology of the eye. 2nd ed. (Grant WM, ed.) Springfield, IL:Charles C Thomas.

Griffith JF, Nixon GA, Bruce RD, Reer PJ, Bannan EA. 1980. Dose-response studies with chemical irritants in the albino rabbit eye as a basis for selecting optimum testing conditions for predicting hazard to the human eye. Toxicol Appl Pharmacol 55:501-513.

Guerriero FJ, Seaman CW, Sutton TJ, Sprague Gl, Guest R, Whittingham A. 2002. The rabbit enucleated eye test (REET) as a screen to reduce the use of animals in worker safety studies. [Abstract No. 1607] Toxicol Sci (The Toxicologist Supplement) 66(1-S).

Guerriero FJ, Seaman CW, Olson MJ, Guest R, Whittingham A. 2004. Retrospective assessment of the rabbit enucleated eye test (REET) as a screen to refine worker safety studies. [Abstract No. 1282] Toxicol Sci (The Toxicologist Supplement) 78(1-S).

Hackett RB, McDonald TO. 1991. Eye irritation. In: Advances in Modern Toxicology: Dermatoxicology. 4th ed. (Marzulli FN, Maibach HI, eds.). Washington, DC:Hemisphere Publishing Corporation, 749-815.

Harbell JW, Curren RD. 2002. In vitro methods for the prediction of ocular and dermal toxicity. In Handbook of Toxicology (Eds. MJ Derelanko and MA Hollinger). 835-866. CRC Press: Boca Raton, Florida.

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjostrom M. 1996. Recommendations of the application of biostatistical methods during the development and validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. ATLA, 24:511-530.

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: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:National Toxicology Program.

ICCVAM Authorization Act of 2000. 2000. Public Law 106-545. [114 Stat. 2721]. Available: http://iccvam.niehs.nih.gov/about/overview.htm [accessed 21 October 2004].

INVITTOX 1994. Protocol IP-85, May 1994, European Commission for the Validation of Alternative Methods (ECVAM) Scientific Information Service (ECVAMSIS).

Jacobs G, Martens M. 1987a. Introduction of the ultrasonic pachometer in the *in vitro* corneal injury technique (Enucleated Eye Test) as a valid alternative for the optic pachometer. Report submitted by G Jacobs dated 05/06/1987.

Jacobs G, Martens M. 1987b. Evaluation of the enucleated eye test against the *in vivo* eye irritation test in the rabbit. Report submitted by G. Jacobs dated 01/12/1987.

Jacobs G, Martens M. 1988. The enucleated eye test: a comparison of the use of ultrasonic and optical pachymeters. Toxicology *In vitro* 2:253-256.

Jacobs G, Martens M. 1989. An objective method for the evaluation of eye irritation *in vivo*. Fd Chem Toxic 27(4):255-258.

Jacobs G, Martens M. 1990. Quantification of eye irritation based upon *in vitro* changes of corneal thickness. ATLA 17:255-262.

Jester JV, Maurer JK, Pertoll WM, Wilkie DA, Parker RD, Cavanaugh HD. 1996. Application of *in vivo* confocal microscopy to the understanding of surfactant-induced ocular irritation. Tox Path. 24:412-428.

Jester JV, Petroll WM, Bean J, Parker RD, Carr GJ, Cavanaugh HD, Maurer JK. 1998. Area and depth of surfactant-induced corneal injury predicts extent of subsequent ocular responses. Invest Ophth Vis Sci. 39(13):2610-2625.

Jester JV, Li L, Molai A, Maurer JK. 2001. Extent of corneal injury as a basis for alternative eye irritation tests. Toxicol *In vitro* 15:115-130.

Jones PA, Budynska E, Cooper KJ, Decker D, Griffiths HA, Fentem JH. 2001. Comparative evaluation of five *in vitro* tests for assessing the eye irritation potential of hair-care products. ATLA 29:669-692.

Kinoshita J. 1962. Some aspects of the carbohydrate metabolism of the cornea. Inves Ophth. 1:178-186.

Koëter BWM, Prinsen MK. 1985. Comparison of *in vivo* and *in vitro* eye irritancy test systems: A study with 34 substances. In: Alternative Methods in Toxicology (Goldberg AM, ed.). Vol 3, Chapter A9. New York:Mary Ann Liebert, 571-579.

Kuckelkorn R, Schrage N, Keller G, REdbrake C. 2002. Emergency treatment of chemical and thermal eye burns. Acta Ophth Scand 80:4-10.

Leopold IH. 1945. Local toxic effect of detergents on ocular structures. Arch Opthalmol. 34:99-102.

Lewin L, Guillery H. 1913. Die Wirkungen von Arzneimitteln und Giften auf das Auge. Hirschwald, Berlin. 2nd edition.

Lewis RW, McCall JC, Botham PA. 1994. Use of an in vitro test battery as a prescreen in the assessment of ocular irritancy. Toxic In Vitro 8(1):75-79.

Marsh RJ, Maurice DM. 1971. The influence of non-ionic detergents and other surfactants on human corneal permeability. Exp Eye Res. 11:43-48.

Marzulli F, Ruggles D. 1973. Rabbit eye irritation test: collaborative study. Journal of the Association of Analytical Chemists 56:905-914.

Maurer JK, Parker RD. 1996. Light microscopic comparison of surfactant-induced eye irritation in rabbits and rats at 3 hours and recovery/day 35. Toxicol Path 24:403-411.

Maurer JK, Li HF, Petroll WM, Parker RD, Cavanaugh HD, Jester JV. 1997. Confocal microscopic characterization of initial corneal changes of surfactant-induced eye irritation in the rabbit. Tox Appl Pharm 143:291-300.

Maurer JK, Parker RD Carr GJ. 1998. Ocular irritation: microscopic changes occurring over time in the rat with surfactants of known irritancy. Toxicol Path 26:217-225.

Maurer JK, Parker RD Jester JV. 2002. Extent of initial corneal injury as the mechanistic basis for ocular irritation: Key findings and recommendations for the development of alternative assays. Reg Tox Pharmacol 36:106-117.

Maurice DM. 1957. The structure and transparency of the cornea. J Physiol 136:263-286.

McCulley JP. 1987. Chemical injuries. In: Cornea; scientific foundations and clinical practice (Smolin G, Thoft RA, eds.). Boston:Little, Brown, 527-542.

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

McDonald TO, Seabaugh V, Shadduck JA, Edelhauser HF. 1987. Eye irritation. In: Dermatotoxicology. (Marzulli FN, Maibach HI, eds). 3rd edition. Washington:Hemisphere Publishing Corporation, 641-696.

McLaughlin RS. 1946. Chemical burns of the human cornea. Am J Opthalmol. 29:1355-1362.

Mishima S, Hedbys BO. 1968a. Physiology of the cornea. Int Opt Clin. 8(3):527-560.

Mishima S, Hedbys BO. 1968b. Measurement of corneal thickness with the Haag-Streit pachymeter. Arch Ophthal. 80:710-713.

Nakano M. 1958. Effect of various antifungal preparations on the conjunctiva and cornea of rabbits. Yakuzaigaku. 18:94-99.

NIOSH. 2004. Work-Related Injury Statistics Query System. Available: http://www2a.cdc.gov/risqs/ [accessed 26 October 2004].

Norn MS. 1971. Vital staining of cornea and conjunctiva. Eye Ear Nose Throat Mon 50:294-299.

Nourse WL, Tyson, CA, Bednarz RM. 1995. Mechanisms of Mild Ocular Irritation. Toxicol In Vitro. 9(6):967-976.

Occupational Safety and Health Act of 1970 (OSHA). 1970. Public Law 91-596

Oksala O, Stjernschantz J. 1988. Effects of calcitonin gene-related peptide in the eye. Inv Opth Vis Sci. 29(7):1006-1011.

OECD. 1984. Provisional Data Interpretation Guides for Initial Hazard Assessment of Chemicals. OECD. Paris.

OECD. 1987. Test guideline 405, Acute eye irritation/corrosion, adopted February 24, 1987. In OECD Guidelines for Testing of Chemicals. OECD, Paris.

OECD. 1996. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Methods. ENV/MC/CHEM/TG (96)9. OECD, Paris.

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD principles on Good Laboratory Practice (as revised in 1997). ENV/MC/CHEM(98)17. OECD, Paris.

OECD. 2002. Test guideline 405. Acute eye irritation/corrosion, adopted April 24, 2002. In: OECD Guidelines for Testing of Chemicals. OECD, Paris.

Parish W. 1985. Ability of in vitro (corneal injury-eye organ-and chorioallantoic membrane) tests to represent histopathological features of acute eye inflammation. Food and Chem Tox. 23:215-227.

Pasquale LC, Hayes AW. 2001. Chapter 18. Acute Toxicity and Eye Irritancy. In: Principles and Methods of Toxicology 4th Ed. (Hayes AW, ed.). Philadelphia:Taylor and Francis, 853-915.

Pfister R. 2005. Chemical Trauma. In: Smolin and Thoft's, The Cornea: Scientific Foundations and Clinical Practice (Foster CS, Azar DT, and Dohlman CH eds). New York:Lippincort Williams and Wilkins, 781-796.

Price JB, Andrews IJ. 1985. The *in vitro* assessment of eye irritancy using isolated eyes. Food and Chem Tox 25(2):313-315.

Russell WMS, Burch RL. 1992. The principles of humane experimental technique. 14th ed. South Mimms, Potters Bar, Herts, England:Universities Federation for Animal Welfare:.

Salz JJ, Anen SP, Berstein J, Caroline P, Villasensor RA, Schanzlin DJ. 1983. Evaluation and comparison of sources of variability in the measurement of corneal thickness with ultrasonic and optical pachymeters. Ophth. Surg. 14(9):750-754.

Sina JF, Gautheron PD. 1994. A multitest approach to evaluating ocular irritation in vitro. Toxicol Meth 4(1):41-49.

Solti J, Freeman JJ. 1988. Effect of reducing the number of animals in acute toxicity/irritation tests on U.S. and European labeling requirements. The Toxicologist 8:263.

Springer J, Chambers W, Green S, Gupta K, Hill R, Hurley P, Lambert L, Lee C, Lee J, Liu P, Lowther D, Roberts C, Seabaugh V, Wilcox N. 1993. Number of animals for sequential testing. Food and Chemical Toxicology 31:105-109.

Suker GF. 1913. Injury to cornea from oxalic acid. Opthalmol Rec. 23:40-47.

Swanston DW. 1985. Assessment of the validity of animal techniques in eye irritation testing. Food Chem Toxic 23:169-173.

Talsma D, Leach C, Hatoum N, Gibbons R, Roger J-C, Garvin P. 1988. Reducing the number of rabbits in the Draize eye irritancy test: A statistical analysis of 155 studies conducted over 6 years. Fundamental and Applied Toxicology 10:146-153.

Thoft RA. 1979. Chemical and thermal injury. Int Ophthalmol Clin Summer 19(2):243-256.

Thornton SP. 1985. A guide to pachymeters. Ophthal. Surg. 15:993.

Toxic Substances Control Act of 1976. 1976. Public Law 94-469.

UN. 2003. Globally Harmonised System of Classification and Labelling of Chemicals (GHS). New York & Geneva:United Nations Publications.

Unger WG. 1990. Review: Mediation of the Ocular Response to Injury. J Ocular Pharmacol 6(4):337-353.

Unger WG. 1992. Pharmacological and neural bases for eye irritation. Ann NY Acad Sci 641:176-186.

Vaughn D, Asbury T, Riordan-Eva P. 1999. General Ophthalmology. 15th ed. New York:McGraw-Hill.

Whittle E, Basketter D, York M, Kelly L, Hall T, McCall J, Botham P, Esdaile D, Gardner J. 1992. Findings of an interlaboratory trial of the enucleated eye method as an alternative to the eye irritation test. Tox Methods 2(1):30-41.

Wilhelmus K. 2001. The Draize eye test. Survey Ophthalmol. 45:493-515.

Wilkie DA,Wyman M. 1991. Comparative anatomy and physiology of the mammalian eye. In *Dermal and Ocular Toxicology, Fundamentals and Methods* (Ed. DW Hobson). 433-491, Boca Raton, Florida:CRC Press, Inc.

Williams SJ, Graepel GJ, Kennedy GL. 1982. Evaluation of ocular irritancy potential: Interlaboratory variability and effect of dosage volume. Toxicol Letters 2: 235-241.

Wilson SE, Mohan RR, Mohan RR, Ambrósio, Jr. R, Hong JW, Lee JS. 2001. The corneal wound healing response: Cytokine-mediated interaction of the epithelium, stroma, and inflammatory cells. Prog Retin and Eye Res 20(5): 625-637.

Wolff E. 1946. The muco-cutaneous junction of the lid margin and the distribution of tear fluid. Trans. Ophthal. Soc U.K. 66:291-308.

York M, Lawrence RS, Gibson GB. 1982. An *in vitro* test for the assessment of eye irritancy in consumer products – preliminary findings. Int J of Cosm Sci 4:223-234.

York M, Steiling W. 1998. A critical review of the assessment of eye irritation potential using the Draize rabbit eye test. J App Tox 18: 233-240.

[This Page Intentionally Left Blank]

13.0 GLOSSARY[1]

Accuracy[2]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance" (see also "two-by-two" table). Accuracy is highly dependent on the prevalence of positives in the population being examined.

Assay[2]: The experimental system used. Often used interchangeably with "test" and "test method."

Benchmark substance: A substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties:

- a consistent and reliable source(s)
- structural and functional similarity to the class of substances being tested
- known physical/chemical characteristics
- supporting data on known effects
- known potency in the range of the desired response

Benchmark control: A sample containing all components of a test system and treated with a known substance (i.e., the benchmark substance) to induce a known response. The sample is processed with test substance-treated and other control samples to compare the response produced by the test substance to the benchmark substance to allow for an assessment of the sensitivity of the test method to assess a specific chemical class or product class.

Blepharitis: Inflammation of the eyelids.

Bulbar conjunctiva: The portion of the conjunctiva that covers the outer surface of the eye.

Chemosis: A form of eye irritation in which the membranes that line the eyelids and surface of the eye ("conjunctiva") become swollen.

Classification system: An arrangement of quantified results or data into groups or categories according to previously established criteria.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

---

[1] The definitions in this Glossary are restricted to their uses with respect to the Draize rabbit eye test method and the IRE test method.

Coefficient of variation: A statistical representation of the precision of a test. It is expressed as a percentage and is calculated as follows:

$$\left(\frac{standard\ deviation}{mean}\right) \times 100\%$$

Concordance[2]: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of "relevance". The term is often used interchangeably with "accuracy" (see also "two-by-two" table). Concordance is highly dependent on the prevalence of positives in the population being examined.

Conjunctiva: The mucous membrane that lines the inner surfaces of the eyelids and folds back to cover the front surface of the eyeball, except for the central clear portion of the outer eye (the cornea). The conjunctiva is composed of three sections: palpebral conjunctiva, bulbar conjunctiva, and fornix.

Conjunctival sac: The space located between the eyelid and the conjunctiva-covered eyeball. Substances are instilled into the sac to conduct an *in vivo* eye test.
Cornea: The transparent part of the coat of the eyeball that covers the iris and pupil and admits light to the interior.

Corneal opacity: Measurement of the extent of opaqueness of the cornea following exposure to a test substance. Increased corneal opacity is indicative of damage to the cornea. Opacity can be evaluated subjectively as done in the Draize rabbit eye test, or objectively with an instrument such as an "opacitometer."

Corneal permeability: Quantitative measurement of damage to the corneal epithelium by a determination of the amount of sodium fluorescein dye that passes through all corneal cell layers.

Corneal swelling: An objective measurement in the IRE test of the extent of distention of the cornea following exposure to a test substance. It is expressed as a percentage and is calculated from corneal thickness measurements that are recorded at regular intervals during the IRE test. Increased corneal swelling is indicative of damage to the corneal epithelium.

Corneal thickness: The depth of the cornea measured using an ultrasonic pachymeter or a depth-measuring attachment on a slit-lamp.

Corrosion: Destruction of tissue at the site of contact with a substance.

Corrosive: A substance that causes irreversible tissue damage at the site of contact.

Endpoint[2]: The biological process, response, or effect assessed by a test method.

Enucleate: To remove without cutting into.

False negative[2]: A substance incorrectly identified as negative by a test method.

False negative rate[2]: The proportion of all positive substances falsely identified by a test method as negative (see "two-by-two" table). It is one indicator of test method accuracy.

False positive[2]: A substance incorrectly identified as positive by a test method.

False positive rate[2]: The proportion of all negative substances that are falsely identified by a test method as positive (see "two-by-two" table). It is one indicator of test method accuracy.

Fibrous tunic: The outer of the three membranes of the eye, comprising the cornea and the sclera; also called *tunica fibrosa oculi.*

Fluorescein penetration/retention: A subjective measurement in the IRE test of the extent of fluorescein sodium that is retained by epithelial cells in the cornea following exposure to a test substance. Increased fluorescein retention is indicative of damage to the corneal epithelium.

Globally Harmonized System (GHS): A classification system presented by the United Nations that provides (a) a harmonized criteria for classifying substances and mixtures according to their health, environmental and physical hazards, and (b) harmonized hazard communication elements, including requirements for labeling and safety data sheets.

Good Laboratory Practices (GLP)[2]: Regulations promulgated by the U.S. Food and Drug Administration and the U.S. Environmental Protection Agency, and principles and procedures adopted by the Organization for Economic Cooperation and Development and Japanese authorities that describe record keeping and quality assurance procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Hazard[2]: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Interlaboratory reproducibility[2]: A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test method can be transferred successfully among laboratories.

Intralaboratory repeatability[2]: 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.

Intralaboratory reproducibility[2]: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

In vitro: In glass. Refers to assays that are carried out in an artificial system (e.g., in a test tube or petri dish) and typically use single-cell organisms, cultured cells, cell-free extracts, or purified cellular components.

In vivo: In the living organism. Refers to assays performed in multicellular organisms.

Iris: The contractile diaphragm perforated by the pupil and forming the colored portion of the eye.

Negative control: An untreated sample containing all components of a test system, except the test substance solvent, which is replaced with a known nonreactive material, such as water. This sample is processed with test substance-treated samples and other control samples to determine whether the solvent interacts with the test system.

Negative predictivity[2]: The proportion of correct negative responses among substances testing negative by a test method (see "two-by-two" table). It is one indicator of test method accuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

Neuroectodermal tunic: The innermost of three membranes of the eye, comprising the retina.

Nictating (nictitating) membrane: The membrane that moves horizontally across the eye in some animal species (e.g., rabbit, cat) to provide additional protection in particular circumstances. It may be referred to as the "third eyelid."

Nonirritant: (a) A substance the produces no changes in the eye following application to the anterior surface of the eye. (b) Substances that are not classified as GHS Category 1, 2A, or 2B; or EU R41 or R36 ocular irritants.

Nonsevere irritant: (a) A substance that causes tissue damage in the eye following application to the anterior surface of the eye; the tissue damage is reversible within 21 days of application and the observed adverse effects in the eye are less severe than observed for a severe irritant. (b) Substances that are classified as GHS Category 2A or 2B; EPA Category II, III, or IV; or EU R36 ocular irritants.

Ocular: Of or relating to the eye.

Ocular corrosive: A substance that causes irreversible tissue damage in the eye following application to the anterior surface of the eye.

Ocular irritant: A substance that produces a reversible change in the eye following application to the anterior surface of the eye.

Opacitometer: An instrument used to measure "corneal opacity" by quantitatively evaluating light transmission through the cornea. The instrument has two compartments, each with its own light source and photocell. One compartment is used for the treated cornea, while the other is used to calibrate and zero the instrument. The difference between photocell signals in the two compartments is measured electronically as a change in voltage, and is displayed digitally, generating numerical opacity values with arbitrary units.

Palpebral conjunctiva: The part of the conjunctiva that covers the inner surface of the eyelids.

Pannus: A specific type of corneal inflammation that begins within the conjunctiva, and with time spreads to the cornea. Also referred to as "chronic superficial keratitis."

Performance[2]: The accuracy and reliability characteristics of a test method (see "accuracy", "reliability").

pH: A measure of the acidity or alkalinity of a solution. pH 7.0 is neutral; higher pHs are alkaline, lower pHs are acidic.

Positive control: A sample containing all components of a test system and treated with a substance known to induce a positive response, which is processed with the test substance-treated and other control samples to demonstrate the sensitivity of each experiment and to allow for an assessment of variability in the conduct of the assay over time.

Positive predictivity[2]: The proportion of correct positive responses among substances testing positive by a test method (see "two-by-two" table). It is one indicator of test method accuracy. Positive predictivity is a function of the sensitivity of the test method and the prevalence of positives among the substances tested.

Prevalence[2]: The proportion of positives in the population of substances tested (see "two-by-two" table).

Protocol[2]: The precise, step-by-step description of a test method, including a listing of all necessary reagents, criteria and procedures for evaluation of the test data.

Quality assurance[2]: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures is assessed independently by individuals other than those performing the testing.

Reduction alternative[2]: A new or modified test method that reduces the number of animals required.

Reference test method[2]: The accepted *in vivo* test method used for regulatory purposes to evaluate the potential of a test substance to be hazardous to the species of interest.

Refinement alternative[2]: A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

Relevance[2]: The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the "accuracy" or "concordance" of a test method.

Reliability[2]: A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

Replacement alternative[2]: A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility[2]: The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

Sclera: The tough, fibrous tissue that extends from the cornea to the optic nerve at the back of the eye.

Sensitivity[2]: The proportion of all positive substances that are classified correctly as positive in a test method. It is a measure of test method accuracy (see "two-by-two" table).

Secondary bacterial keratitis: Inflammation of the cornea that occurs secondary to another insult that compromised the integrity of the eye.

Slit-lamp microscope: An instrument used to directly examine the eye under the magnification of a binocular microscope by creating a stereoscopic, erect image. In the IRE test method, this instrument is used to view the anterior structures of the rabbit eye as well as to objectively measure corneal thickness with a depth-measuring device attachment.

Severe irritant: (a) A substance that causes tissue damage in the eye following application to the anterior surface of the eye that is not reversible within 21 days of application or causes serious physical decay of vision. (b) Substances that are classified as GHS Category 1, EPA Category I, or EU R41 ocular irritants.

Solvent control: An untreated sample containing all components of a test system, including the solvent that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the

same solvent. When tested with a concurrent negative control, this sample also demonstrates whether the solvent interacts with the test system.

Specificity[2]: The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy (see "two-by-two" table).

Superfusion apparatus: Water-jacketed, temperature-controlled, custom-built apparatus usually made from Perspex® plastic that houses isolated rabbit eyes in removable holders placed in isolated chambers to provide short-term maintenance of metabolic and physiological activity. The chambers have darkened walls to permit slit-lamp examination, and drip tubes positioned over the eyes to provide a continuous saline flow to maintain the eyes during an experiment.

Test[2]: The experimental system used; used interchangeably with "test method" and "assay."

Test method[2]: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay." See also "validated test method" and "reference test."

Test method component: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures.

Tiered testing: A testing strategy where all existing information on a test substance is reviewed, in a specified order, prior to *in vivo* testing. If the irritancy potential of a test substance can be assigned, based on the existing information, no additional testing is required. If the irritancy potential of a test substance cannot be assigned, based on the existing information, a step-wise animal testing procedure is performed until an unequivocal classification can be made.

Toxic keratoconjunctivitis: Inflammation of the cornea and conjunctiva due to contact with an exogenous agent. Used interchangeably with "contact keratoconjunctivitis, irritative keratoconjunctivitis, and chemical keratoconjunctivitis."

Transferability[2]: The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two table[2]: The two-by-two table can be used for calculating accuracy (concordance) ([a+d]/[a+b+c+d]), negative predictivity (d/[c+d]), positive predictivity (a/[a+b]), prevalence ([a+c]/[a+b+c+d]), sensitivity (a/[a+c]), specificity (d/[b+d]), false positive rate (b/[b+d]), and false negative rate (c/[a+c]).

		New Test Outcome		
		Positive	**Negative**	**Total**
Reference Test Outcome	**Positive**	a	c	a + c
	Negative	b	d	b + d
	Total	a + b	c + d	a + b + c + d

Uvea tract: The middle of three membranes of the eye, comprising the iris, ciliary body, and choroid. Also referred to as the "vascular tunic."

Validated test method[2]: An accepted test method for which validation studies have been completed to determine the relevance and reliability of this method for a specific proposed use.

Validation[2]: The process by which the reliability and relevance of a procedure are established for a specific purpose.

Vascular tunic: The middle of three membranes of the eye, comprising the iris, ciliary body, and choroid. Also referred to as the "uvea."

Weight of evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.