NIH Publication No: 06-4513

BACKGROUND REVIEW DOCUMENT

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

Isolated Chicken 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 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).

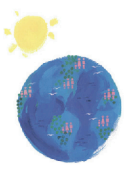
ICCVAM is comprised of representatives from 15 U.S. Federal regulatory and research agencies that use, generate, or disseminate toxicological information. ICCVAM conducts technical evaluations of new, revised, and alternative methods with regulatory applicability. The Committee promotes the scientific 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 Chicken 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. 06-4513

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

Printed: 08/08/2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

ALTBIB	Bibliography on Alternatives to Animal Testing
ASTM	American Society for Testing and Materials
BCOP	Bovine Corneal Opacity and Permeability
BP	British Pharmacopoeia
BRD	Background Review Document
CAS	Chemical Abstract Service
CASRN	Chemical Abstracts Service Registry Number
CEET	Chicken Enucleated Eye Test
CGRP	Calcitonin Gene Related Peptide
CO	Corneal opacity
CPSC	U.S. Consumer Product Safety Commission
CS	Corneal swelling
CTFA	Cosmetic, Toiletries and Fragrance Association
CV	Coefficient of variation
^{o}C	Degrees centigrade
DOD	U.S. Department of Defense
DOT	U.S. Department of Transportation
DTC	Days to clearing
EC	European Commission
EC/HO	European Commission/British Home Office
ECETOC	European Center for Ecotoxicology and Toxicology Of Chemicals
ECVAM	European Center for the Validation of Alternative Methods
EDTA	Ethylenediaminetetraacetic acid
EEC	European Economic Council
EPA	U.S. Environmental Protection Agency
EU	European Union
FDA	U.S. Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FHSA	Federal Hazardous Substance Act

FR	Fluorescein retention
FR	*Federal Register*
FRAME	Fund for the Replacement of Animals in Medical Experiments
g	Gram
GHS	Globally harmonized system
GLP	Good Laboratory Practice
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
I.N.	If necessary
INVITOXX	*In Vitro* Techniques in Toxicology (ERGATT FRAME ECVAM Data bank)
IRAG	Interagency Regulatory Alternatives Group
IR	Iritis
IRE	Isolated Rabbit Eye
kg	Kilogram
LVET	Low Volume Eye Test
MeSH	U.S. National Library of Medicine's Medical Subject Heading
mg	Milligram
min	Minutes
mL	Milliliter
MAS	Maximum average score
MMAS	Modified maximum average score
MMPs	Matrix metalloproteinases
N/A	Not available or not applicable
NI	Nonirritant
n.p.	Not provided
NS	Not specified
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods

NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NTP	U.S. National Toxicology Program
OECD	Organisation for Economic Cooperation and Development
OPPTS	EPA Office of Prevention, Pesticides and Toxic Substances
OSHA	U.S. Occupational Safety & Hazards Administration
OTWG	Ocular Toxicity Working Group
P&G	The Proctor and Gamble Company
P.L.	Public Law
PLA_2	Phospholipase A_2
PMS	Polymethylsiloxane
QA	Quality Assurance
r	Correlation coefficient
RN	Registry number
SCNM	Study criteria not met
SD	Standard deviation
STN	Scientific and Technical Information Network
%sw	Percent corneal swelling
t	Time
TG	Test Guideline
TNF	Tumor necrosis factor
TNO	TNO Nutrition and Food Research
TSA	Test substance applicator
TSCA	Toxic Substances Control Act
µL	Microliter
UN	United Nations
w/v	Weight to volume ratio
ZEBET	German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments

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 Wind, 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 ICE Test Method Background Review Document

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

ECVAM
Chantra Eskes, Ph.D.

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

TNO Nutrition & Food Research Institute
Menk Prinsen

PREFACE

During the past 60 years, government regulatory agencies have implemented safety testing requirements to identify potential hazards of various chemicals and products in order 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 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 (ICE) assay, and the Isolated Rabbit Eye (IRE) assay.

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

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 BRDs 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 resent 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 BCOP 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 reclassified 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 the BRD. The efforts of the many individuals who contributed to the preparation of this document 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

EXECUTIVE SUMMARY

This Background Review Document (BRD) reviews available data and information regarding the validation status of the Isolated Chicken Eye (ICE)[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 Labeling of Chemicals (UN 2003). The objective of this BRD is to describe the current validation status of the ICE test method, including what is known about its accuracy and reliability, the scope of the substances tested, and the availability of a standardized 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* notices requesting high quality *in vivo* rabbit eye test and *in vitro* ocular irritation data for ICE, the Isolated Rabbit Eye (IRE), the Hen's Egg Test – Chorioallantoic Membrane (HET-CAM), and the Bovine Corneal Opacity and Permeability (BCOP) test methods. An online literature search identified three publications that contained relevant ICE test results for an evaluation of test method accuracy[2] and reliability[3]. Submitted unpublished ICE data and detailed *in vivo* data for two additional studies allowed for an evaluation of test method accuracy[2] and reliability[3] for a total of five studies.

Other published and unpublished ICE test method studies are reviewed in **Section 9.0** (Other Scientific Reports and Reviews). This section discusses 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 ICE test method is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the chicken eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, opacity, and fluorescein retention. Each measurement is either converted into a quantitative score used to calculate an overall Irritation Index, or assigned a qualitative categorization that is used to assign an *in vitro* irritancy classification. Either of these outcomes can then be used to predict the *in vivo* ocular irritation potential of a test substance. A histopathological assessment can also be included on a case-by-case basis to discriminate

---

[1] In order to maintain consistency among the isolated eye methods, ICE is used throughout the BRD as opposed to CEET (Chicken Enucleated Eye Test), which is used by the test method developer.

[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.

borderline cases (i.e., substances that produce results that preclude assignment to a single category).

The ICE 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 ICE test method useful for the identification of ocular corrosives and severe irritants in a tiered testing strategy on a case-by-case basis. In this strategy, positive *in vitro* test results are considered in a weight-of-evidence decision as to whether to classify the substance as an ocular corrosive or severe irritant. Negative results and suspected false positive *in vitro* results proceed to standard *in vivo* testing or to *in vitro* test methods that are capable of detecting false negative corrosives and severe irritants.

The ICE test method protocols used in the various studies considered in this BRD are similar, but not identical. The essential principles of the test method include the enucleation of eyes from chickens obtained from a slaughterhouse, mounting in a specially-designed apparatus and testing for damage that may have occurred during the isolation process, treating the eyes with a test substance, collecting corneal thickness, opacity and permeability data, and evaluating the data in relation to a prediction model. The primary difference among these protocols was the number of treated eyes per test substance. Acceptable ranges for negative control responses, historical data used to establish these ranges, and procedures to determine the optimum quantity of test substance to be applied have not been published.

A total of 175 substances in five studies can be used to evaluate ICE test method accuracy, 85 of which were proprietary compounds, consisting largely of products or formulations. The ICE test method has been used to test a variety of chemical and product classes. The chemical classes tested included, but were not limited to, alcohols, acids, hydrocarbons, surfactants, inorganic chemicals, acyl halides, alkalis, solvents, esters, heterocyclics, ketones, onium compounds, and organophosphates. The proprietary compounds tested included, but were not limited to, detergents, pesticides, silicone powder, ink, toilet cleaners, and thermal paper coatings.

Some of the published *in vivo* rabbit eye test data on the substances used to evaluate the accuracy of ICE for detecting ocular corrosives and severe irritants was limited to average score data or the reported irritancy classification. However, detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal at 24, 48, and 72 hours and/or assessment of the presence or absence of lesions at 7, 14, and 21 days was necessary to calculate the appropriate EPA (1996), EU (2001), and GHS (UN 2003) ocular irritancy hazard classification. Thus, some of the test substances for which there was only limited *in vivo* data could not be used for evaluating test method accuracy and reliability.

Three of the studies received contained original study records. Summary *in vitro* data was 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.

The ability of the ICE 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 ICE was assessed separately for each *in vitro-in vivo* comparative study. In the second approach, the accuracy of ICE was assessed after pooling data across *in vitro-in vivo* comparative studies that used the same method of data collection and analysis. 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 the ICE test method performance was comparable among the three hazard classification systems. The overall accuracy of the ICE test method ranged from 83% to 87%, depending on the classification system used. Sensitivity and specificity ranged from 50% to 59% and 92% to 94%, respectively. The false positive rate ranged from 6% to 8%, while the false negative rate ranged from 41% to 50%.

According to the accuracy analysis, the chemical class with the highest false positive rate in all three classification systems was alcohols, with false positive rates ranging from 27% to 50%. The chemical class with the next highest false positive rate in all three classification systems was esters, with false positive rates ranging from 11% to 13%. No other chemical classes were consistently overpredicted by all three systems, although for most of the chemical classes tested, the number of substances in each was too few to resolve any definitive overprediction trends by the ICE test method. Alcohols were also consistently underpredicted, with false negative rates ranging from 33% to 50%. Other underpredicted chemical classes were amines/amidines (33% to 50%; GHS and EPA systems only), carboxylic acids (17% to 43%), heterocyclics (33% to 40%), inorganics (50%; EU system only), onium compounds (33% to 40%) and polyethers (100%; EU system only).

Regarding the physical form of overpredicted substances, no solids were overpredicted in any classification system, while liquids showed false positive rates ranging from 7% to 10%. Both solids and liquids were underpredicted, however, showing false negative rates ranging from 46% to 70% for solids and 39% to 44% for liquids.

Changes in the ICE test method performance statistics for substances classified according to the GHS classification system were observed when three discordant classes (alcohols, surfactants, and solids) were excluded from the data set; accuracy increased from 83% (120/144) to 92% (69/75), the false negative rate decreased from 50% (15/30) to 29% (2/7) and the false positive rate decreased from 8% (9/114) to 6% (4/68).

Test substances labeled as pesticides were not overpredicted in any classification system, but showed false negative rates ranging from 40% to 60%. Test substances labeled as surfactants were also not overpredicted, but showed false negative rates ranging from 44% to 57%.

Regarding the pH of underpredicted substances for which such information was available, substances with a pH less than 7.00 showed false negative rates of 27% to 40% (3/11 to 4/10) and substances with a pH greater than 7.0 showed false negative rates of 50% to 57% (3/6 to 4/7). However, it is noted that pH information was available for only a portion of the 27 to 32 severe irritant substances (i.e., Category 1, Category I, or R41) for each classification system in the database.

Finally, with respect to the GHS classification system only, as evidenced by an analysis of NICEATM-defined GHS Category 1 sub-groupings, the eight underpredicted substances were more likely to be classified *in vivo* based on persistent lesions (false negative rate of 60% [3/5]), rather than on severe lesions (false negative rate of 28% [5/18]).

A quantitative assessment of intralaboratory data from one study (Prinsen 2000), using scores for each endpoint (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) and the ICE Irritation Index, indicates the extent of intralaboratory reproducibility of the ICE test method. Four test substances were used in this study. When considering the results of this analysis, note that some test substances had a mean or a standard deviation equal to zero for some endpoints and that scores for corneal opacity and fluorescein retention have a small dynamic range (0 to 4 and 0 to 3, respectively). Corneal thickness measurements within experiments showed %CV values ranging from 0.9 to 6.1 and corneal opacity scores showed %CV values ranging from zero to 86.6 (the highest value was obtained for a nonirritating substance). The %CV values for fluorescein retention were zero for three of the four substances and ranged from zero to 86.6 for the nonirritating substance, although this range is based on only two experiments. Finally, the %CV values for the ICE Irritation Index for the four substances ranged from -86.6 to 41.6, with the same nonirritating substance exhibiting the outlying values (-86.6 and 41.6).

The data from Prinsen (2000) was also used to do a CV analysis on between-experiment values for each endpoint (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE Irritation Index, for each test substance. When considering the results of this analysis, note that scores for corneal opacity or fluorescein retention have a small dynamic range (0 to 4 and 0 to 3, respectively).

The %CV values for the corneal thickness measurement ranged from 1.8 to 6.3 and those for corneal swelling ranged from 13.9 to 138.7. The %CV values for the corneal opacity score ranged from 8.7 to 95.8. The %CV values for the fluorescein retention score ranged from zero to 141.4. Finally, the %CV values for the ICE Irritation Index ranged from 4.1 to 91.8. Note that for all endpoints considered except corneal thickness, the highest %CV values were obtained for the nonirritating substance.

A qualitative assessment of the data provided for multiple laboratories in one study (Balls et al. 1995) provides an indication of the extent of interlaboratory reproducibility. In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the four participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 44 to 45 (75% to 76%) of the 59 substances tested *in vitro* in the study, depending on the classification system used. All four laboratories were in 100% agreement on the classification of 60% to 70% of substances classified as corrosives/severe irritants, 85% to 88% of substances classified as nonsevere irritants/nonirritants.

Among the 15 substances classified according to the GHS scheme that exhibited interlaboratory differences in *in vitro* classification, four were classified as alcohols. Two of the 15 substances were classified as cationic surfactants, two were classified as acetates/esters, and two were classified as ketones. Solvents was the product class appearing

most frequently among these substances, with seven of the 15 substances represented. Other product classes represented by multiple substances were chemical intermediates (five substances) and synthetic flavor ingredients (four substances). In regard to physical properties, of the 15 substances with discordant results among the four laboratories, 10 were liquid (seven water soluble) and five were solid (four water insoluble).

Mean endpoint values (i.e., fluorescein retention, corneal opacity, corneal swelling) and the ICE Irritation Index for each substance were provided for each of the four laboratories participating in the study. To provide a quantitative assessment of interlaboratory variability, individual laboratory ICE test results were used to calculate a mean, standard deviation, and the %CV for corneal opacity, fluorescein retention, corneal swelling, and the Irritation Index for each substance tested. Mean and median %CV values for all 59 substances were 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 ICE, a wide range of %CV values for individual substances is evident for all endpoints. The mean/median %CV values were 39%/36% (ranging from 0 to 159%) for fluorescein retention, 47/37% for corneal opacity (ranging from 0 to 159%), 77%/75% for corneal swelling (ranging from 31 to159%), and 35%/32% (ranging from 10 to 98%) for the Irritation Index. When only severe irritants (GHS Category 1, based on *in vivo* data) are considered, the %CV values are lower for all endpoints, with corneal swelling (mean of 72%, median of 69%) the sole endpoint with a mean/median %CV value greater than 35%. Of the four liquid substances with a CV < 35% for corneal swelling (2,2-dimethylbutanoic acid, 2,6-dichlorobenzoyl chloride, benzalkonium chloride 5%, and cetylpyridinium bromide 10%), two were water insoluble. No solid substances had a CV < 35% for corneal swelling. It is noteworthy that some of the corneal swelling values reported in the data are greater than 80% and therefore above the reported historical maximum range of 60-80%. However, different depth measuring devices may have been used by the participating laboratories to determine corneal thickness, which, unless normalized, would have contributed to the increased variability and/or the excessive values calculated for this evaluation (Prinsen M, personal communication).

Common physicochemical characteristics do not appear among the substances showing the most variable responses (defined as CV >70% for any of the endpoints). Of the 37 substances with significant variability in at least one endpoint, 18 are solids (of a total of 19 solids, 12 of which are water soluble) and 19 are liquids (of a total of 40 liquids, 14 of which are water soluble). However, some chemical classes appear to predominate among the 37 substances with CV values greater than 70%, including seven surfactants (of 12 tested), five heterocyclic compounds (of six tested), four acetate/esters (of six tested), and four acids (of six tested). Therefore, the majority of substances tested from these chemical classes exhibited increased interlaboratory variability.

Balls et al. (1995) also determined the interlaboratory correlation between ICE test method endpoint data generated by each laboratory for all substances tested, as well as for subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). Interlaboratory correlation coefficients generally spanned a range of 0.6 to 0.9 depending on the specific subsets of substances being evaluated. However, the range of correlation

coefficients for some endpoints was larger (e.g., correlation coefficients for ICE-Mean Swelling ranged from 0.210 to 0.757 when testing substances that are insoluble in water).

Review of the mean *in vitro* data from this study indicates that wide ranges of corneal swelling values were recorded for the five insoluble test substances that were classified as ocular corrosives/severe irritants. For all five substances, the same laboratory produced the highest values, with mean corneal swelling percentages ranging from 1.5 to 6 times greater than the next highest mean corneal swelling value for the same substance tested by the other three laboratories. In addition, of the 14 remaining ocular corrosives/severe irritants (soluble and surfactant combined), a considerably higher value was reported for corneal swelling by the same laboratory for 12 substances. This trend was also apparent for nonsevere irritants/nonirritants.

Although the interlaboratory variability for fluorescein retention or corneal opacity was not as pronounced for the insoluble ocular corrosives/severe irritants, and could not be associated with a single laboratory, the ranges of correlation coefficients for these endpoints are also relatively high. Therefore, the apparently large interlaboratory variability noted among these substances cannot be attributed to a single laboratory or to a single endpoint.

At least one eye is traditionally included in each ICE study as a negative/vehicle control (isotonic saline). Individual eye data that could be used to perform a CV analysis on between-experiment values for each of the test method endpoints (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE Irritation Index for each test substance were obtained from negative control eyes. This analysis revealed that responses in the negative control eye remain relatively consistent.

Concurrent positive control substances have not been employed in the ICE test method, and therefore, an evaluation of historical positive control data is not possible.

As stated above, this BRD provides a comprehensive summary of the current validation status of the ICE test method, including what is known about its reliability and accuracy, and the scope of the substances tested. Raw data for the ICE 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*

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

*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 = U.S. Federal Hazardous Substances Act; OSHA = U.S. Occupational Safety and Health Administration

[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 (CPSC 1995; 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 quantitative 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**). 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. Category: I = Corrosive, corneal involvement, or irritation persisting more than 21 days II= Corneal involvement or irritation clearing in 8-21 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) 6 animals 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) 3 animals 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.

Regulatory Agency (Authorizing Act)	Number of Animals	Minimum Observation Times (after treatment)	Mean Score Taken?	Positive Response	Irritant/Nonirritant Classification
GHS-Irreversible Eye Effects	3	1, 2, 3 days (observation until Day 21)	Yes	Mean animal values (over Days 1, 2, and 3) of: Opacity ≥ 3 and/or Iritis ≥ 1.5	- At least 2 positive response animals = Eye Irritant Category 1 - At least 1 animal where Opacity, Chemosis, Redness, or Iritis > 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 Product Safety Act]), FDA (Food, Drug, and Cosmetics Act), and OSHA (Occupational Safety and Health Act)	6 (12, 18 possible)	1, 2, 3 days (observation may be extended to 7 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 Product 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 categories are based on severity of the lesions and/or the duration of persistence. The GHS, the U.S., 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.

The present BRD evaluates the ability of the ICE test method to identify ocular corrosives and severe irritants. This test method has been referenced in the published literature as the ICE (Balls et al. 1995) as well as the Chicken Enucleated Eye Test (CEET, Prinsen and Koëter 1993; Prinsen 1996; Chamberlain 1997). To maintain consistency among the isolated eye test methods, the term ICE is used throughout this BRD. The ICE protocol was first described by Prinsen and Koëter (1993) and was developed based on the IRE test developed by Burton et al. (1981). In this *in vitro* bioassay, the test substance is applied to the cornea of eyes isolated from chickens that have been processed for human consumption. Three parameters are evaluated to measure the extent of damage to the eye following exposure to a chemical substance: corneal swelling, corneal opacity, and fluorescein retention. While the latter two parameters involve a qualitative assessment, analysis of corneal swelling provides a quantitative measurement, thus potentially providing improved precision and reduced interlaboratory variability compared to the traditional *in vivo* rabbit eye test, which relies only on qualitative measurements.

For current regulatory applications, the ICE 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 ICE as ocular corrosives or severe irritants could be classified as Category 1 eye irritants without the need for animal testing. Substances that are negative in ICE 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 ICE test method also may be useful in a battery of *in vitro* eye irritation methods that that collectively predict the eye irritation potential of a substance *in vivo*. However, the predictivity of a battery approach will first require the assessment of the individual performance of each component test method.

The ICE test method is currently used in some European companies (e.g., pharmaceutical and contract testing companies) as an in-house screen to assess the ocular irritation potential of a

wide range of substances or products. Substances are tested either individually, as mixtures, or in product formulations. Materials that are considered nonirritating based on the ICE test method are tested *in vivo* to confirm the *in vitro* results (Prinsen 1996; Chamberlain et al. 1997).

Although the ICE 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 ICE has not been shown to adequately discriminate between eye irritants and non-irritants (Liebsch and Spielmann 2002; EU 2004).

1.1.2 Peer Reviews of the ICE Test Method

Studies have been conducted in recent years to assess the validity of the ICE test method as a complete replacement for the *in vivo* ocular toxicity test method (e.g., Balls et al. 1995). 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 ICE test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (EPA 1996), and the EU (EU 2001). This BRD was prepared for use by an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) expert panel review of ICE as a method to identify ocular corrosives and severe irritants. Parallel reviews of the IRE, HET-CAM, and BCOP test methods are being 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 ICE Test Method

1.2.1 Purpose and Mechanistic Basis of the ICE Test Method

The ICE is an organotypic model (i.e., isolated whole organ, or component thereof) that provides short-term maintenance of the whole eye in an isolated system (Chamberlain et al., 1997). ICE was developed as a modification to the IRE test in order to obviate the need for laboratory animals as the source for test eyes.

The endpoints evaluated in the ICE to measure the extent of damage to the eye following exposure to a chemical substance are corneal swelling, corneal opacity, and fluorescein retention. Corneal swelling is determined by calculating the increase in corneal thickness from a baseline measurement. Corneal thickness has been identified as a quantitative and reliable endpoint for the evaluation of corneal injury (Burton 1972). Fluorescein retention provides an assessment of corneal permeability, indicative of damage to the corneal surface. Finally, because it is used in both assays, corneal opacity provides a measurement of corneal damage in the ICE that can be directly correlated to the *in vivo* rabbit eye test. In addition,

morphological changes may be included in the study protocol and used in the categorization process. Histopathology may also be included on a case-by-case basis, and may be useful in resolving borderline cases (Prinsen M, personal communication).

Histopathology or confocal microscopy would allow for a more accurate assessment of 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 ICE 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.

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.

Figure 1-1 Anatomy of the Human Eye

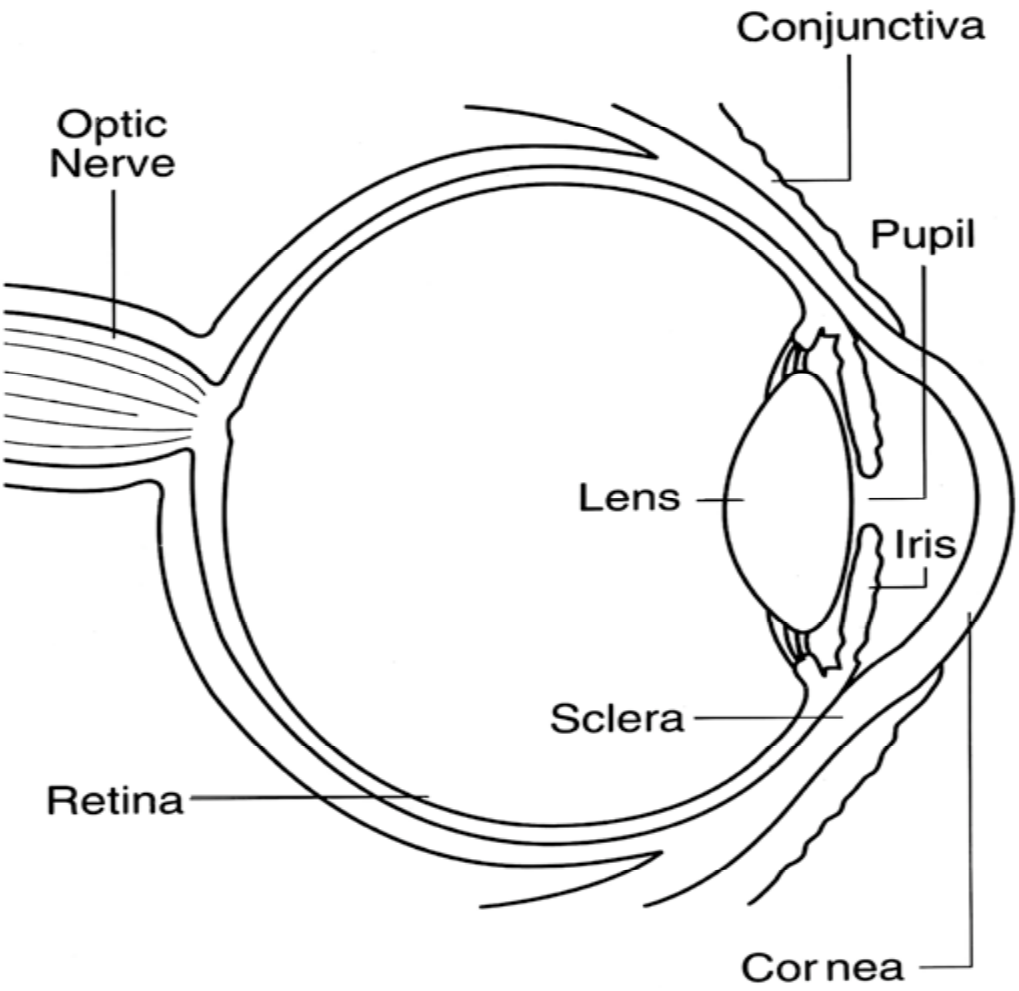


Figure obtained at http://www.nei.nih.gov/photo/eyean/index.asp

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 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 and Harbell 1998; Cooper et al. 2001). This could suggest that the rabbit's tear film is less resistant to evaporation. 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, 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 a human 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 (FHSA 1964; EPA 1996; 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 etc.) (Berta 1992; Fox and Boyes 2001; Grant 1974; Nourse et al. 1995; McCulley 1987). 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) (Pfister 2005; Unger 1992). 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; Thoft 1979; York et al. 1982; McCulley 1987; 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, 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-Blano 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 interleukin-1 (IL-1) and tumor necrosis factor (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 ethylenediaminetetraacetic acid (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, 1968b). 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 *Differences Between the Chicken and Mammalian Eye*

Similar to rabbits, but unlike humans, birds have three separate eyelids. The upper lid is fixed, the lower lid is movable, and a third lid (nictitating membrane) is extensive and movable by two muscles, which are not found in mammals. In addition, birds are devoid of eyelashes. The avian eyeball is flattened along the visual axis, while the lens is more spherical relative to mammalian eyes. This flattened shape allows the retina to be positioned at the focus of all light passing through the lens. Therefore, visual acuity is greatly increased in birds relative to mammals (mammals focus light on one section of the retina, the fovea, thus seeing acutely only one small area of vision). The lens is joined to the ciliary muscles with an annular pad. This soft pad appears to act as a cushion. A bony ring, the scleral ring, is located at the corneal-scleral junction and is made up of overlapping plates that form the avian visceral skeleton. Thin, overlapping plates of scleral cartilage continue backwards from the scleral ring to the optic nerve. The pecten is an unusual formation composed of a folded, highly vascularized tissue layer found at the junction of the optic nerve to the eye (optic papilla) and extends from the retina into the posterior chamber. Unlike mammals, the avian retina is avascular, and therefore its nutrition comes from choroidal vessels and the pecten. The pecten is also believed to be involved in other functions such as ocular fluid exchange, intraocular pressure maintenance, and as a navigational aid (i.e., for estimating the sun's angle) (Bone 1979).

1.2.2.5 *Comparison of the ICE Test Method with the In Vivo Rabbit Eye Test Method*

The ICE test method is capable of evaluating the principal ocular component damaged by severe irritants, the cornea. This test method provides both quantitative (corneal swelling) and qualitative (corneal opacity; fluorescein retention) measurements of corneal injury. In contrast, the *in vivo* rabbit eye test qualitatively evaluates corneal opacity, effects on the iris and conjunctiva, as well as the reversibility and delayed onset of any ocular effects detected. The standard *in vivo* test is carried out over three full days and can last up to 21 days if irritation persists. Thus, the ICE test method differs from the *in vivo* rabbit eye test method in the following significant ways:

- ICE evaluates only corneal effects and does not take into account effects on the iris and the conjunctiva that are evaluated in the *in vivo* rabbit eye test
- ICE does not account for the reversibility of corneal effects induced by a test substance
- ICE does not account for systemic effects following ocular instillation that may be noted with the *in vivo* rabbit eye test (e.g., toxicity or lethality as in the case of certain pesticides)
- as a short-term test, ICE may not identify slow-acting irritants (i.e., irritants with a delayed response)

In the isolated chicken eye, neurogenic components that drive tear film production are not functional. Although the cornea is constantly hydrated with a saline drip in the ICE 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 ICE test method is visually observed or may be assessed using a slit-lamp.

In the ICE 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 ICE Test Method and/or Limits of the ICE Test Method

Studies indicate that the ICE test method is amenable to use with a broad range of solid and liquid substances with few limitations. Substances amenable to testing include, but are not limited to: inorganic chemicals, hydrocarbons, heterocyclic chemicals, polymers, and mixtures/formulations.

Substances that are poorly soluble or those materials that run off corneal surfaces may not be compatible with this test method. Such substances may not be in contact with the eye for an adequate period of time, which could lead to inaccurate results and conclusions (Earl 1998). Chemicals and substances that adhere to the eye, despite rinsing, may hinder evaluation and assessment of the eye during the analysis portion of the test method. Based on studies with a limited number of surfactants or formulations containing surfactants (e.g., detergents), these substances appear to be underpredicted by the ICE test method. Similarly, a limited dataset indicates that solid substances may also be underpredicted by the ICE test method. In contrast, studies with a limited number of alcohols indicate that some of these substances may be overpredicted by the ICE test method.

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 ICE 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) 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 ICE 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 ICE 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 U.S. Consumer Product Safety Commission (CPSC), and the Chemical Safety and Hazard Investigation Board. Therefore, the ICE 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 (decreases or eliminates pain and distress), (b) reduce animal use, or (c) replace animal use.[2]

The ICE test method has the potential to refine or reduce animal use in eye irritation testing. The ICE test method was designed to use an animal species that is routinely used in the food industry (chicken) and that are routinely slaughtered for other purposes (e.g., food consumption). Substances that are identified as ocular corrosives or severe irritants would be excluded from further *in vivo* testing, which would reduce the number of rabbits used for ocular testing and 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

---

[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).

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 ICE test method would reduce the time needed to assess a substance, when compared to the currently accepted *in vivo* rabbit eye test method protocol. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days and can be extended for up to 21 days. Comparatively, the ICE test method can be completed in about six hours from the start of treatment. As it is currently used at TNO (TNO Nutrition and Food Research, Toxicology and Applied Pharmacology, Zeist, The Netherlands), ICE is incorporated as a prescreen for the *in vivo* rabbit test without additional costs. If the prescreen shows that severe irritancy is expected, a full ICE test is performed without further *in vivo* testing at the price of the *in vivo* test. If a full ICE test is used as a stand-alone assay (as mandated in EU countries for cosmetics/household products), depending on the number of samples tested, the cost of a test ranges from $847 to $1,694 per sample (as of 25 May 2004). However, these costs do not include the inclusion of a positive control, as is recommended in the proposed standardized protocol (**Appendix A**), which would increase the cost of the assay. By comparison, the current cost of a GLP compliant EPA OPPTS Series 870 Acute Eye Irritation (EPA 1998) or Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 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). Therefore, it would appear that the cost, based on conducting Good Laboratory Practice (GLP) compliant studies, of an ICE test is comparable to that of an *in vivo* rabbit test.

1.3.2 Intended Uses of the ICE 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, Ocular Irritant Class R41 per the EU classification system, or Ocular Irritant Class 1 per the GHS classification system).

1.3.3 Similarities and Differences in the Endpoints Measured in the ICE and the *In Vivo* Reference 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 (Draize et al. 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 1.3**, the ICE test method evaluates only corneal effects to measure the extent of an irritant response. Corneal opacity is the only common endpoint shared between the ICE and the *in vivo* rabbit eye test.

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

The ICE test method is being considered for use in the identification of ocular corrosives and severe irritants in a tiered testing strategy (e.g., 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.

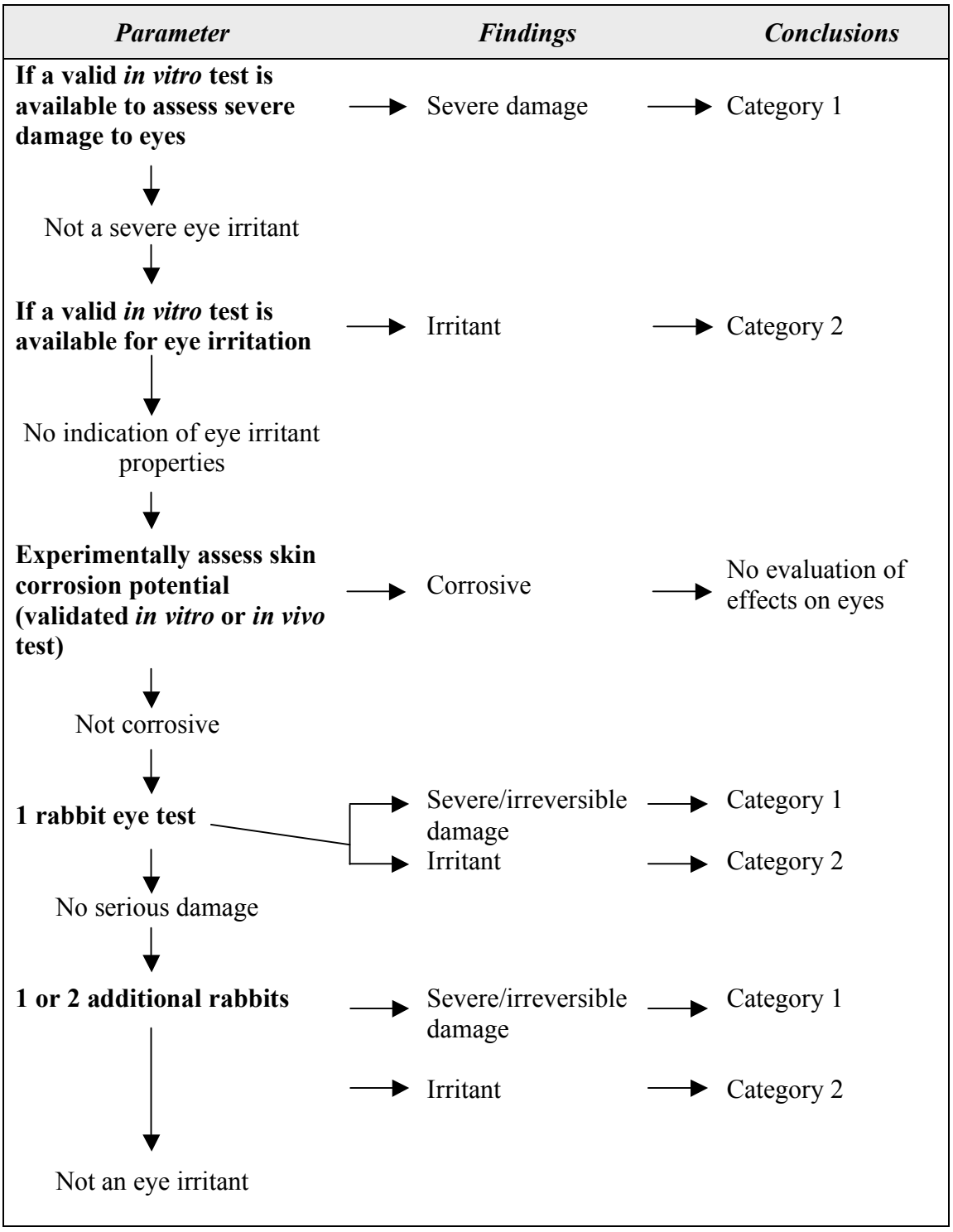
1.4 Validation of the ICE Test Method

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 [P.L.] 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 ICE 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 cover the range of responses 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 ICE 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

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



Adapted from UN (2003).

available information on the various versions of the ICE test method that have been published. Where adequate data are available, the qualitative and quantitative performances of the assay are evaluated and the reliability of each version of ICE is compared with the

reliability of the other ICE versions. If there are insufficient data to support the recommendation of a standardized protocol for ICE, 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 ICE BRD

The ICE 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, ALTBIB, and Web of Science was conducted to retrieve database records on publications reporting on *in vitro* testing of substances using the ICE test method. Specifically, records were sought using the search terms (1) "chicken AND (eye OR eyes) AND isolated AND (test OR assay OR [in AND vitro])" and (2) "chicken AND (eye OR eyes) AND enucleated." Each database record included authors, bibliographic citation, and indexing terms. Most records also included abstracts. A database of the literature citations was established using bibliographic database software. Each database record included authors, bibliographic citation, and indexing terms. Most records also included abstracts. Of the 177 records obtained from the search (last updated in January 2004), three contained results from a ICE test method. A search of the STN International database was completed in February 2004, with no additional articles containing results from an ICE test method identified.

[This Page Intentionally Left Blank]

2.0 ICE TEST METHOD PROTOCOL COMPONENTS

2.1 Overview of How the ICE Test Method is Conducted

As discussed in **Section 1.0,** the enucleated eye test using rabbit eyes (i.e., the IRE test method) was introduced by Burton et al. (1981) as a prescreen for severe eye irritants. The ICE protocol, first described by Prinsen and Koëter (1993), was developed based on the IRE test method. For this reason, most of the test method components remained the same for the ICE test method, although the dose-volume of the test substance was modified for the chicken eye. **Appendix A** provides a comparative summary of test method components for all ICE protocols reviewed in this BRD.

In the ICE test method, three parameters are evaluated to measure the extent of eye damage following exposure to a chemical substance: corneal swelling, corneal opacity, and fluorescein retention. While the latter two parameters involve a qualitative assessment done by slit-lamp microscopic examination, analysis of corneal swelling is measured quantitatively, potentially providing improved precision and reduced interlaboratory variability compared to the *in vivo* rabbit eye test, which relies only on qualitative measurements.

During an ICE study, a test substance is applied to the corneas of enucleated chicken eyes, isolated from chickens processed for human consumption. Chicken heads are transported from the slaughterhouse to the laboratory within two hours, and eyes are immediately dissected and placed in the superfusion apparatus, where isotonic saline is supplied onto the cornea through a steel tube attached to a peristaltic pump. Test substances are applied as a single dose (30 µL for liquids, or 30 mg for solids) for 10 seconds, followed by rinsing with 20 mL isotonic saline. Corneal reactions are measured at regular intervals up to four hours post-treatment, and mean values for each parameter (corneal swelling, corneal opacity, and fluorescein retention) are determined; fluorescein retention is evaluated at 30 minutes post-treatment only. Based on the maximum mean values[1] of these measurements, the irritation potential of the test substance is defined within a range from nonirritating to severely irritating.

2.2 Description and Rationale for the Test Method Components

The ICE test method protocol has remained virtually unchanged since its initial publication in 1993. The protocol was used by three other laboratories (Shell, Unilever, and Rhône-Poulenc) in the European Commission (EC)/British Home Office (HO) validation study (Balls et al. 1995). The laboratory at Rhône-Poulenc (currently Bayer CropScience) is still using the ICE test method. Many of the essential protocol components are based on historical use, and rationales for their inclusion are not known.

[1] For each endpoint, the mean of three eyes is recorded for each time point and the largest mean value is used for scoring.

2.2.1 Materials, Equipment, and Supplies Needed

2.2.1.1 *Sources of Chicken Eyes*

Historically, chickens obtained from a slaughterhouse have been used for this assay because they are killed for human consumption, eliminating the need for laboratory animals. A local source of chickens (preferably spring chickens of either sex, approximately seven weeks old, approximately 2.5-3.0 kg, breed not specified), close to the laboratory should be located, so chicken heads can be transferred to the laboratory and processed within two hours after the birds are killed. Although a controlled study to evaluate the optimum chicken age has not been done, the age and weight of the chickens specified represents that of spring chickens traditionally processed by a poultry slaughterhouse. Unpublished studies on adult chickens show no significant differences in results (Prinsen M., personal communication). The lack of an age difference is presumed to be due to the physiological properties of the cornea in general, which deviates only slightly during the lifespan of the chicken. Although a formal study to determine the optimum window of time to transport the heads to the laboratory has not been conducted, two hours appears to produce consistent results (Prinsen M, personal communication). However, given the quality control measures taken once the eyes reach the laboratory (i.e., baseline fluorescein retention and corneal thickness measurements), it appears that longer transport times could be considered.

2.2.1.2 *Preparation of the Eyes*

Because eyes can be more precisely dissected in the laboratory, intact heads are transported there from the slaughterhouse, at ambient temperature in plastic boxes humidified with tissues moistened with isotonic saline or water. Transportation of the eyes in the intact chicken head provides effective protection from external damage during transport while humidified transportation boxes prevent desiccation. The temperature range during transport is not considered critical because of quality control measures done before an eye is used in an assay (Prinsen M, personal communication). The post-mortem eyelid closure reflex provides an efficient barrier to external contaminants, desiccation and physical injury during transportation. However, the effect of hypoxia on the eye resulting from closed eyelids has not been studied. It is unclear if less elapsed time between the animal's death and study initiation would improve results.

Before inspection, the eyelids are carefully excised, taking care not to damage the cornea. Corneal integrity is quickly assessed with a drop of 2% (w/v) sodium fluorescein (British Pharmacopoeia - BP) applied to the corneal surface for a few seconds, and then rinsed with isotonic saline. Although an exact exposure duration is not used, the fluorescein should not remain on the corneal surface for more than a few (i.e., 1-3) seconds. The treated cornea is then examined for fluorescein retention by damaged corneal epithelial cells and for corneal opacity with a slit-lamp microscope. If undamaged, the eye is further dissected from the skull, taking care not to damage the cornea. Because of the firmness of the chicken eye sclera, dissection is simple. The eyeball is pulled from the orbit by holding the nictitating membrane firmly with surgical forceps, and the eye muscles are cut with a bent, blunt-tipped scissor. It is important to avoid causing corneal damage due to excessive pressure (e.g., compression artifacts). When the eye is removed from the orbit, a visible portion of the optic nerve should be left attached. If the optic nerve is detached from the eye during dissection, a hole can be created at the surface of the posterior portion of the eye, resulting in a change in

turgor pressure by exposing the posterior chamber to the external environment. Once removed from the orbit, the eye is placed on an absorbent pad and the nictitating membrane and other connective tissue are cut away.

2.2.1.3 *ICE Experimental Setup*

Each eye is mounted in a custom-built stainless steel clamp (**Figure 2-1**), with the cornea positioned vertically and then transferred to a chamber in a custom-built superfusion apparatus (**Figure 2-2**). Photographs and line drawings of the clamp and superfusion apparatus are available from M. Prinsen at TNO and can be used to craft similar clamps.

Figure 2-1 Custom-Built Stainless Steel Eye Clamp for the ICE Test Method

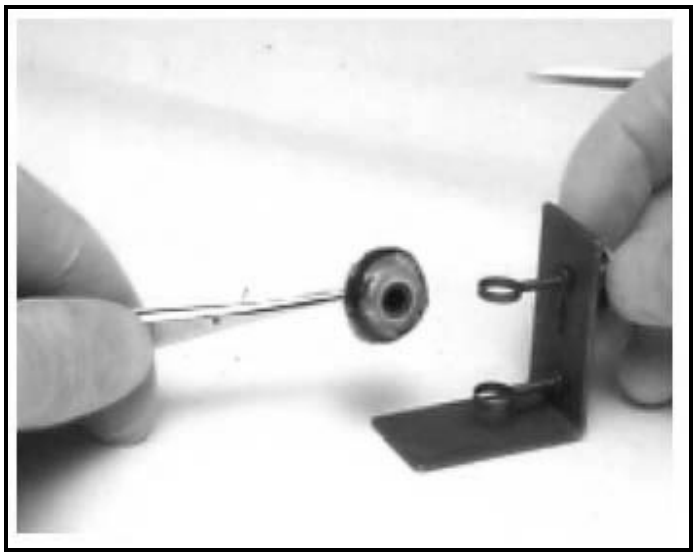
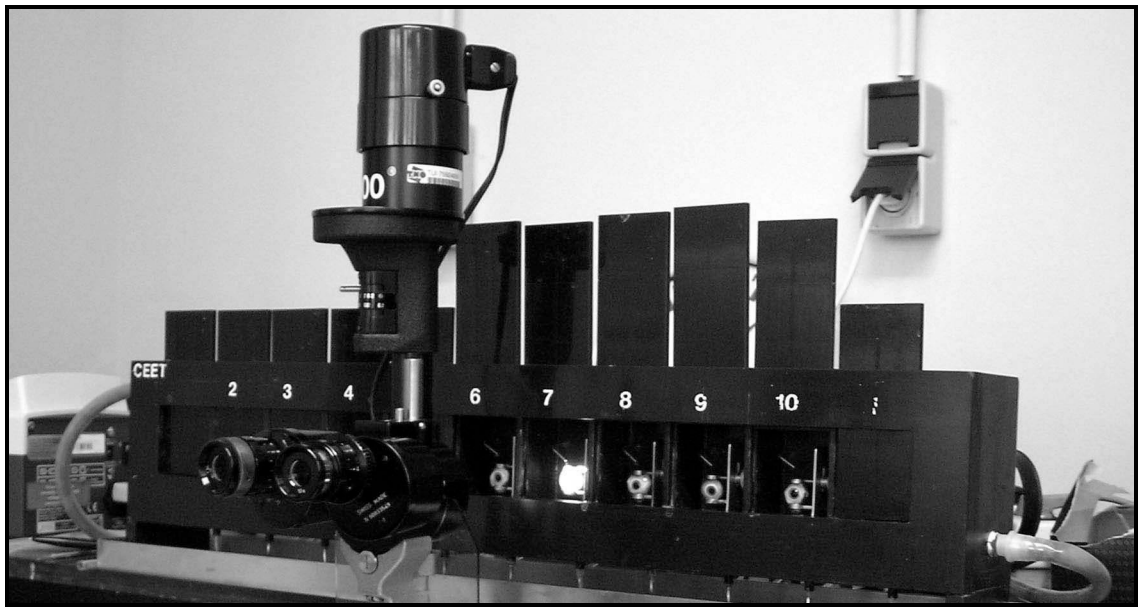


Figure 2-2 Superfusion Apparatus for the ICE Test Method



The clamp is positioned in the multi-chambered superfusion apparatus so the entire cornea is moistened with isotonic saline from a bent stainless steel tube via a peristaltic pump at a rate of approximately 0.10-0.15 mL (2 or 3 drops)/minute. This flow rate was adopted from that used in the IRE test method and has been used extensively in the test method developer's laboratory. The isotonic saline is temperature-controlled to 32 ± 1.5°C , because the steel tube passes through the warm-water mantle of the superfusion apparatus. After placement in the superfusion apparatus, the corneas are again examined with the slit-lamp microscope to ensure that they have not been damaged during the procedure.

Corneal thickness is measured at the corneal apex using an optical pachymeter on the slit-lamp microscope. A slit-lamp system is preferred to ultrasound-based instruments (e.g., ultrasonic pachymeter) because the latter requires direct contact of a probe with the cornea, risking injury to the corneal surface, especially if the test substance has already damaged it. Eyes with a corneal thickness deviating more than 10% from the mean value for the eyes, eyes with a fluorescein retention score of > 0.5 (indicating corneal permeability), or eyes showing corneal opacity or any other signs of damage are rejected and replaced. Based on historical use of the ICE in the developer's laboratory, these values appear to accurately represent the range seen with the average, untreated, and undamaged chicken eye. Once all eyes have been examined and approved, they are equilibrated for 45 to 60 minutes prior to dosing. Unpublished observations have shown that the duration of equilibration is not critical, and may be allowed to extend longer if necessary. In addition, as the chamber doors are opened frequently during the test period and the corneas are frequently rinsed with isotonic saline at ambient temperature during the assay, temperature fluctuations seem to have little effect on the integrity of the cornea, as evidenced by the results obtained with the negative control eyes (Prinsen M, personal communication).

Originally, five eyes were used per test substance (Prinsen and Köeter 1993), but later publications included as few as three eyes per test substance (Balls et al. 1995; Prinsen 1996, Prinsen 2000; Prinsen 2005). Reducing the number of treated eyes from five to three does not appear to have decreased the performance of the ICE test method (Prinsen M, personal communication).

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

Dose-selection procedures are not relevant to this *in vitro* assay. Test substances are applied as neat chemicals (solids or liquids), if possible, or as received in the case of proprietary mixtures or formulations. If dilution is required, test substances are diluted with an aqueous vehicle.

2.2.3 Endpoints Measured

The control and test eyes are examined pre-treatment and at 30, 75, 120, 180, and 240 minutes after treatment using the criteria and scoring system described in **Section 2.2.6**. These time points provide an adequate number of measurements over the four-hour treatment period while leaving sufficient time between measurements for the requisite observations to be made for all eyes.

The endpoints evaluated are corneal opacity, corneal swelling, fluorescein retention (corneal permeability) and morphological effects (e.g., pitting or loosening of the epithelium). Corneal opacity and corneal thickness are evaluated at each time point. Because fluorescein retention reflects initial damage and does not change over time, it is determined only at 30 minutes. If the test substance adheres to the cornea and precludes determination at 30 minutes, fluorescein retention may be assessed at a later time point.

After the final examination at four hours, eyes are typically preserved in 4% neutral buffered formaldehyde for histopathological examination (if necessary or requested). It is widely recognized that microscopic assessment of effects can provide additional information to be added to an overall assessment of toxicity. However, the expense of such a detailed examination may not be warranted in all cases (e.g., if the outcome of the test is clearly negative or clearly positive). Instead, histopathological effects could most efficiently be used to resolve borderline responses by determining the depth-of-injury (Maurer et al. 2002).

2.2.4 Duration of Exposure

2.2.4.1 *Quantity of Test Substance Applied*

A liquid test substance is applied at 0.03 mL with a micropipette, so that the entire corneal surface is bathed. A solid test substance is applied at 0.03 g as a fine powder (grinding may be necessary), and evenly distributed over the corneal surface. Using a fine powder ensures more uniform coverage of the corneal surface. Excess test substance could result in a "piling up" effect, which could preclude uniform coverage of the corneal surface (i.e., some test substance may not come into contact with the surface for the entire exposure period). These quantities were originally chosen because the diameter of the chicken cornea is approximately 30% that of the rabbit cornea (the standard quantities used in the IRE are 0.1 mL or 0.1 g).

2.2.4.2 *Application of the Test Substance*

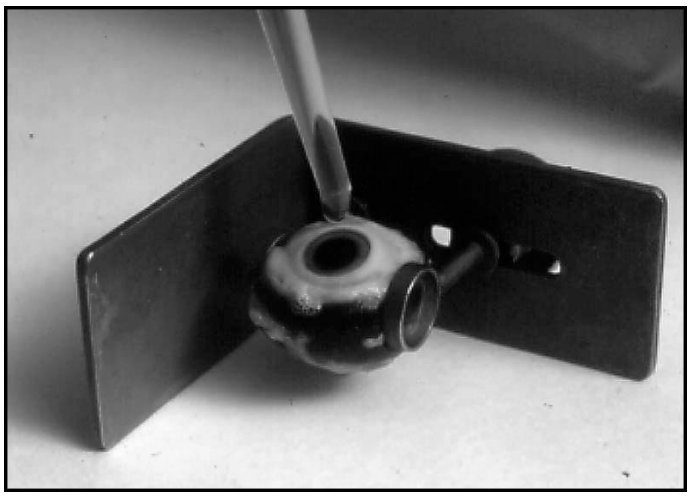
After an equilibration period, a zero reference corneal thickness measurement is taken to establish a baseline (i.e., time = 0) for comparison to post-treatment measurements. Immediately following the zero reference measurement, each eye (in its holder) is removed from the superfusion apparatus, placed in a horizontal position, and the test substance is applied to the cornea (**Figure 2-3**). The test substance is applied for 10 seconds, rinsed from the eye with 20 mL isotonic saline at ambient temperature, then the eye (in its holder) is returned to the superfusion apparatus in the original upright position. During dosing, the clamp holding the eye is removed from the superfusion apparatus and placed on tissue paper with the cornea facing upwards. This position maximizes uniform contact of the test substance with the corneal surface.

The time of application was chosen based on the IRE study design. According to Burton et al. (1981), a 10-second exposure was chosen after experimentation demonstrated that this time interval produced the best discrimination between irritant and non-irritant substances.

2.2.5 Known Limits of Use

Like the *in vivo* rabbit eye test, the ICE test method has been used to test a wide range of substances with various physicochemical characteristics (Prinsen and Koëter 1993;

Figure 2-3 Application of the Test Substance in the ICE Test Method



Prinsen 1996; Prinsen 2000; Prinsen 2005). However, some substances may require an alternative testing strategy due to their increased potential for yielding results that are not predictive of an *in vivo* response (Balls et al. 1995).

One such class of substances is hydrophobic compounds. Because of the aqueous environment under which the assay is conducted, very hydrophobic compounds may not fully contact the corneal surface, which could potentially result in an underprediction of the *in vivo* response. To ensure adequate contact with the cornea, it may be necessary to employ multiple exposures, or to remove as much isotonic saline as feasible from the corneal surface prior to application. Such measures would require additional protocol optimization to ensure that the accuracy of the results was improved.

Solid substances (e.g., powders, pastes) may also cause discordant responses in the ICE test method. In some cases, solids may adhere to the corneal surface and rinsing may fail to detach them. Residual solid substance may preclude accurate assessment of endpoints. More extensive rinsing (i.e., by using an increased rinse volume or by increasing the number of rinses) may be applied and, in general, residual test substance will be removed within one hour (Prinsen M, personal communication).

The ICE test method can only provide information on the ability of a test substance to interact with the cornea (i.e., damage to conjunctival tissue is not assessed). However, a direct relationship between conjunctival damage and corneal swelling has been reported by Burton (1972). Based on this relationship, it may be inferred that only very mildly irritating substances are capable of inducing conjunctival effects in the absence of corneal effects (Prinsen M, personal communication). Therefore, focusing solely on corneal effects does not appear to limit the effectiveness of ICE with respect to predicting corrosives and severe irritants.

Finally, the availability of a poultry abattoir close to the testing laboratory may be a limiting factor. However, because the test method is essentially portable and is a relatively quick procedure, the experimental set-up could be moved to the slaughterhouse (Balls et al. 1995).

2.2.6 Nature of the Response Assessed

2.2.6.1 *Data to be Collected*

As noted in **Section 2.2.3**, corneal endpoints observed in the ICE are opacity, swelling, fluorescein retention, and morphological changes. Response severity is graded at each time point. Numerical and descriptive data are collected. Numerical data includes scores for opacity, thickness, and fluorescein retention, while descriptive data represents morphological and histopathological results. Therefore, the responses assessed in this test method are both qualitative and quantitative.

2.2.7 Appropriate Controls and the Basis for their Selection

2.2.7.1 *Negative Controls*

The recommended negative control in all of the published ICE protocols is isotonic saline. This control appears most suitable since the test method is conducted using isotonic saline to moisten the enucleated chicken eyes as well as for the requisite rinsing steps. Treating the negative control eyes with isotonic saline instead of test substance ensures that any mechanical alterations (i.e., those not related to the test substance) and the general conditions maintained by the superfusion apparatus are properly controlled. In cases where the test substance is solubilized or diluted, the most common vehicle recommended is deionized or distilled water. Other vehicles may be used if demonstrated to be useful to the assay. However, inclusion of alternate vehicles should be adequately controlled in the experimental design.

2.2.7.2 *Positive Controls*

As discussed by Harbell and Curren (2002), the function of the positive control is to ensure the test system is operating within normal limits and each experiment is properly executed so toxic effects of interest can be properly detected. A concurrent positive control 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.

None of the published ICE protocols recommend the use of a concurrent positive control substance. The rationale for excluding such a control has been based on the historical success with the ICE method in the developer's laboratory (Prinsen M, personal communication).

2.2.7.3 *Solvent Controls*

Solvent controls are recommended when solvents other than deionized water or saline are used to dissolve test substances, in order to demonstrate that the solvent is not interfering with the test system.

2.2.7.4 *Benchmark Controls*

Benchmark controls may be useful for demonstrating that the test method is functioning properly for detecting the ocular irritancy potential of chemicals of a specific chemical class or with a specific range of responses, or for evaluating the relative irritancy potential.

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

2.2.8.1 *Negative Controls*

An acceptable range of negative control responses is an irritancy classification of nonirritating (**Section 2.2.13**). Because an aqueous medium (isotonic saline) is used to moisten and rinse the chicken eyes, isotonic saline or distilled water may be used as the vehicle when a test substance is diluted. Therefore, eyes exposed to isotonic saline typically serve as both the negative and vehicle controls.

2.2.8.2 *Positive Controls*

Because positive controls have not been traditionally used in this test method, a defined range of responses has not been previously described. However, it would seem prudent that the positive control substance produce an Irritancy Score (**Section 2.2.13**) that is appropriate based on its historical classification as a severe irritant in the *in vivo* rabbit eye test. If adequate historical ICE test method data are not available for a particular positive control, pilot studies may have to be conducted to provide this information.

2.2.8.3 *Solvent Controls*

If another solvent is used to dissolve or dilute a test substance, separate vehicle and negative (isotonic saline) controls should be included in the experiment. In this case, the vehicle control should also produce a nonirritating response.

2.2.8.4 *Benchmark Controls*

Benchmark controls 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 responses, or for evaluating the relative irritancy potential of an ocular irritant. Therefore, the response of the benchmark should be consistent across multiple experiments.

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

The severity level for each study endpoint is evaluated at each time point. The following criteria and scoring system are applied for the assessment of possible effects:

2.2.9.1 *Corneal Swelling*

Corneal swelling is determined from corneal thickness measurements made with an optical pachymeter on a slit-lamp microscope. It is expressed as a percentage and is calculated from corneal thickness measurements according to the following formula:

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

The mean percentage of swelling for all eyes exposed to the test substance is calculated for each observation time point. Based on the highest mean score for corneal swelling, as observed at any time point, an overall category score is then given for each test substance.

2.2.9.2 *Corneal Opacity*

Corneal opacity is determined with slit-lamp examination, by scoring the area of the cornea that is most densely opacified.

Score		*Observation*
0	=	No opacity
0.5	=	Very faint opacity
1	=	Scattered or diffuse areas; details of the iris are clearly visible
2	=	Easily discernible translucent area; details of the iris are slightly obscured
3	=	Severe corneal opacity; no specific details of the iris are visible; size of the pupil is barely discernible
4	=	Complete corneal opacity; iris invisible

The mean corneal opacity value for all test eyes is calculated for each observation time point. Based on the highest mean score for corneal opacity, as observed at any time point, an overall category score is then given for each test substance.

2.2.9.3 *Fluorescein Retention*

The mean fluorescein retention value for all test eyes is calculated for the 30-minute observation time point only. When test substances have adhered to the cornea, fluorescein retention can be determined whenever the test substance has been adequately removed.

Score		*Observation*
0	=	No fluorescein retention
0.5	=	Very minor single cell staining
1	=	Single cell staining scattered throughout the treated area of the cornea
2	=	Focal or confluent dense single cell staining
3	=	Confluent large areas of the cornea retaining fluorescein

2.2.9.4 *Morphological Effects*

These effects include *pitting* of corneal epithelial cells, *loosening* of the epithelium, *roughening* of the corneal surface and *sticking* of the test substance to the cornea. These findings can vary in severity and may occur simultaneously. The classification of these findings is subjective according to the investigator's interpretation. On the basis of severity of the observed findings, these effects are divided into four categories: 1 = none; 2 = slight; 3 = moderate; 4 = severe. Multiple observers within a single laboratory should ensure that consistency is maintained in assigning scores. A histopathological evaluation may also be done to elucidate more detailed evidence of damage, or lack thereof.

2.2.9.5 *Methods for Collection of Numerical Data*

Qualitative corneal opacity measurements are made with a slit-lamp microscope, and a score assigned based on the scale provided in **Section 2.2.9.2**. The most densely opacified area of the cornea is used for scoring. Fluorescein retention is also a qualitative measurement, evaluated using a slit-lamp microscope. Numerical values are assigned according to the scale

provided in **Section 2.2.9.3**. Corneal thickness is a quantitative measurement that requires either a slit-lamp microscope equipped with an optical pachymeter, or an ultrasonic pachymeter. The slit-lamp system is generally preferred to an ultrasonic pachymeter, as the probe for the latter must be in contact with the cornea during the measurement, which can increase the risk of corneal damage, especially if it is already damaged by irritant exposure. Corneal thickness measurements are used to calculate corneal swelling relative to baseline measurements (**Section 2.2.9.1**).

The severity of corneal damage has been reported to be proportional to the severity of corneal irritation (Burton et al. 1972). The scoring method for each endpoint has been correlated to the EU regulatory classification system (EU 2001) for comparison to *in vivo* results (See **Section 2.2.13**).

2.2.10 Type of Media in Which Data Are Stored

Although not specifically mentioned in published ICE protocols, it is reasonable to assume that data can be collected either by hand, or by directly entering it into a computer spreadsheet. Handwritten data requires subsequent computer entry in order to perform the requisite mathematical calculations (**Section 2.2.12**). Data should be routinely backed up on the source computer hard drive as well as an external disk drive.

Data from the test method should be stored and archived in a manner consistent with international GLP guidelines (OECD 1998; EPA 2003a, 2003b; 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 to expedite information retrieval and storage conditions should minimize document deterioration. An individual should be identified as responsible for these data archives.

GLP guidelines are nationally and internationally recognized rules designed to ensure high-quality laboratory records. They provide a standardized approach for reporting and archiving laboratory data and records, and the test protocol, in order to ensure the integrity, reliability, and accountability of a study (OECD 1998; EPA 2003a, 2003b; FDA 2003).

2.2.11 Measures of Variability

As indicated in **Section 2.2.9**, both numerical and descriptive data are generated by the ICE test method. Variability of numerical ICE test data is typically assessed through calculation of the mean along with the standard error of the mean for each numerical endpoint. Descriptive data may also provide an additional, qualitative measure of variability.

2.2.12 Statistical or Non-Statistical Methods Used to Analyze the Resulting Data

2.2.12.1 *Irritation Index*

The severity level for each study endpoint (i.e., corneal swelling, corneal opacity and fluorescein retention) recorded at each time point is used to calculate the maximum mean score for each endpoint (**Section 2.2.9**), from which an irritation index can be determined. The irritation index is derived by calculating the sum of the maximum mean scores for each

endpoint. Corneal swelling at each time point is calculated as a percentage of corneal thickness at time = zero. An overall corneal swelling score is calculated based on both (1) the mean corneal swelling of all three eyes treated with the test substance and (2) the time when swelling first occurred. Corneal opacity is qualitatively scored at each time point on a scale from zero to four, with zero representing no opacity and four representing complete corneal opacity. The overall corneal opacity score is the highest mean score obtained across the five observation times for the three eyes treated with the test substance. Additionally, the mean retention of fluorescein among the three treated eyes is determined at 30 minutes. To derive the irritation index, the opacity and fluorescein retention scores are equally weighted relative to the maximum corneal swelling obtained. Historical data from the test method developer's laboratory indicates that the maximum swelling observed is approximately 60% to 80%. Therefore, the maximum opacity (score = 4) and fluorescein retention (score = 3) scores obtained with any particular test substance are multiplied by a factor of 20 in order to increase their weighting (Chamberlain et al. 1997). The irritation index has a possible range of 0 to 200.

2.2.13 Decision Criteria and the Basis for the Prediction Model Used to Classify a Test Chemical

2.2.13.1 *Interpretation of Endpoint Scores*

Once each endpoint has been scored, irritancy categories can be assigned based on a pre-determined range. The rationale for the values selected for each range is based on a logical subdivision of these values into the ocular irritancy categories of non, slight, moderate, or severe (Prinsen M, personal communication). Interpretation of corneal thickness, corneal opacity, and fluorescein retention using four irritancy categories is done according to the following scales:

Corneal Thickness

Mean Corneal Swelling (%)	Category
0 to 5	I
> 5 to 12	II
> 12 to 18 (>75 min after treatment)	II
> 12 to 18 (≤75 min after treatment)	III
> 18 to 26	III
> 26 to 32 (>75 min after treatment)	III
> 26 to 32 (≤75 min after treatment)	IV
> 32	IV

Corneal Opacity

Mean Maximum Opacity Score	Category
0.0-0.5	I
0.6-1.5	II
1.6-2.5	III
2.6-4.0	IV

Fluorescein Retention

Mean Fluorescein Retention Score at 30 Minutes Post-treatment	Category
0.0-0.5	I
0.6-1.5	II
1.6-2.5	III
2.6-3.0	IV

2.2.13.2 *Defining the Irritancy Classification*

The overall *in vitro* irritancy classification for a test substance is assessed by reading the irritancy classification that corresponds to the combination of categories obtained for corneal swelling, corneal opacity, and fluorescein retention. For the purposes of this evaluation, which focuses specifically on the ability of ICE to identify corrosives and severe irritants, only the severe irritancy classification for a test substance is presented in the scheme below.

Classification	Combinations of the 3 Endpoints
Severely Irritating	3 x IV
	2 x IV, 1 x III
	2 x IV, 1 x II[1]
	2 x IV, 1 x I[1]
	Corneal opacity ≥ 3 at 30 min (≥ 2 eyes)
	Corneal opacity = 4 at any time point (≥ 2 eyes)
	Severe loosening of the epithelium (≥ 1 eye)

[1]Combinations less likely to occur.

Using similar combination schemes, ICE test results have also been used to predict the *in vivo* classification of substances according to both EU and GHS classification schemes (Prinsen M, personal communication). However, as indicated in **Section 2.2.6**, histopathology may be warranted in order to discriminate between effects that are on the borderline between severe and moderate irritation. If a mathematical comparison is desired, the irritation index described in **Section 2.2.12** may be calculated.

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

The test report should include the following, 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 standard deviation 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).

A standard data collection form is provided in the INVITTOX protocol (INVITTOX 1994). Although this form was generated during earlier ICE studies, in which a total of five eyes were used per test substance, this form could easily be adapted to the presently recommended three eyes per test substance, as is presented in **Figure 2-4**.

2.3 Basis for Selection of the Test System

Several aspects of the chicken eye have been cited as advantages over other species for the enucleated eye test. Chickens are used widely as a food animal species, and therefore access to chicken eyes can be obtained relatively easily. Although rabbits also are used for human consumption, their role as a food animal species is much less prominent than that of beef, pork, or poultry. Consequently, the IRE test is often performed using animals from previous laboratory studies, and therefore the need for laboratory animals is still evident. Given their widespread availability in slaughterhouses, bovine and porcine eyes have also both been investigated as sources of enucleated eyes. The limitations of the bovine cornea in the isolated eye test are its dimensions and the thickness of the cornea, which makes the screening of mild irritants difficult. Apart from the dimensions of the cornea, the same applies for the pig cornea. In addition, due to the slaughtering process, obtaining eyes from processed cows and pigs, as well as the labor required to remove the eye, is more difficult than in chickens (Prinsen and Koëter 1993). The structure of the chicken cornea is comparable to the rabbit cornea and additionally has a well-developed Bowman's membrane, as does the human cornea. Finally, because of the dark black background provided by the iris of a chicken eye, changes in opacity are more easily discriminated than in the rabbit (Prinsen M, personal communication).

Figure 2-4 Sample Scoring Form for the ICE Test Method (Modified from INVITTOX 1994)

TEST COMPOUND:														LIQUID: YES/NO		VISCOUS: YES/NO		
PROJECT NO.:														SOLID: YES/NO		GROUND: YES/NO		
DATE OF TEST:														HYDROPHILIC/HYDROPHOBIC				
SIGNATURE:														WARMED: YES/NO				
														APPEARANCE:				
Eye No.	CORNEAL THICKNESS IN INSTRUMENTAL UNITS AT t =							CORNEAL OPACITY SCORES AT t =						OTHER EFFECTS	FLUORESCEIN RETENTION			
Time	-45	0	30	75	120	180	240	0	30	75	120	180	240		0	30	I.N.	
1																		
%sw																		
2																		
%sw																		
3																		
%sw																		
initials																		
mean																		
SEM																		
category:																		
classification:																		

I.N. = if necessary
%sw = percent corneal swelling

2.4 Proprietary Components

The ICE does not employ any proprietary components. However, differences in slit-lamp systems for measurement of corneal swelling have been documented, and therefore different laboratories may have different corneal swelling values for the different irritancy categories. Therefore, when comparing ICE test data generated by different laboratories, a "correction factor" may be needed to compensate for these differences (i.e., ranking of substances according to corneal swelling figures should be similar, regardless of the apparatus) (INVITTOX 1994). However, to date, this issue has not been considered as a problem in study analysis. Therefore, the potential impact of this issue on the utility of the ICE test method cannot be addressed at this time.

2.5 Basis for Number of Replicate and Repeat Experiments

2.5.1 Within Experiment Replicates

Early studies involving the ICE test method used up to six eyes per test. Five eyes were used for the test substance-treated group, and one eye was used as a negative control (Prinsen and Koëter 1993). However, unpublished observations indicated that reducing the number of test substance-treated eyes to three did not adversely affect the assay results. Because the superfusion apparatus currently used in the developer's laboratory has a capacity for 11 eyes, it was more efficient to use three eyes per test substance, and therefore evaluate three test substances in a single experiment (Prinsen M, personal communication). Additional eyes may be included if there is a concern regarding equivocal results.

2.5.2 Experimental Replicates

None of the published reports indicated that repeating experiments is necessary.

2.6 Compliance with Good Laboratory Practice

Studies should be performed in compliance with GLP Guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). Conducting studies in compliance with GLP guidelines increases confidence in the quality and reliability of test data. Furthermore, if data using these test methods are to be submitted to the EPA or FDA in response to Federal testing requirements, then compliance with appropriate GLP guidelines is required.

2.7 Study Acceptance Criteria

According to the published reports for the ICE method, the only criteria for an acceptable study is that the negative control gives an irritancy classification of nonirritating.

3.0 SUBSTANCES USED FOR VALIDATION OF THE ICE 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 limits to the range of responses that can be evaluated by the *in vitro* test method.

Five reports contained sufficient *in vitro* and *in vivo* data for accuracy analyses[1]. These five reports are Prinsen and Koëter (1993), Balls et al. (1995), Prinsen (1996), Prinsen (2000) and Prinsen (2005).

As noted in **Section 2.2.5**, the ICE test method has been used for a wide range of test substances with different physicochemical characteristics. However, highly hydrophobic compounds and certain solids may require alternative testing strategies to ensure that contact with the corneal surface is maximized (Balls et al. 1995). There is no mention in any of the following studies of modification to the ICE protocol employed to account for this issue.

3.1.1 Prinsen and Koëter (1993)

The chemicals tested in this study were used in a previous study sponsored by the Commission of the European Communities (CEC 1991) to evaluate several *in vitro* ocular toxicity methods, including IRE and HET-CAM. These same chemicals were used by Prinsen and Koëter (1993) to provide comparative data and to determine the suitability of the chicken as an alternative to the rabbit as an eye donor for the isolated eye test.

3.1.2 Balls et al. (1995)

In the EC/HO validation study (Balls et al. 1995), 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 OECD 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 that met the established criteria were found in the ECETOC data bank. However, this selection was determined to be inadequate due to the low number of solids, the insufficient number of moderate to severe irritants, and the lack of pesticides. To

[1] The ability of the ICE test method to accurately identify test substances classified as corrosive or a severe irritant 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 a severe irritant is provided in **Section 4.0**.

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 that met the established criteria (obtained primarily from unpublished studies), 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 removing the solid material from the eye one hour after application, rather than allowing it to remain continuously. Thus, the study protocol 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 Prinsen (1996)

This report described the use of the ICE test method as a prescreen for severe eye irritants at TNO. All substances tested at TNO, from the time that the ICE was implemented as a prescreen up to the report date (1992-1994), are discussed in this report. Therefore, it appears that substances were tested as they were submitted to TNO by industrial, cosmetic, and food manufacturing companies for testing and subsequent regulatory classification. Thus, there was no specific rationale in the use of these substances.

3.1.4 Prinsen (2000)

The four substances tested for this report were siloxane polymers and surfactants, selected as part of phase II of a reference standard validation project conducted at TNO. No specific rationale was provided for the selection of any particular substance.

3.1.5 Prinsen (2005)

This report contained ICE test method data for 50 substances submitted to TNO, subsequent to those tested in Prinsen 1996. Again, no specific rationale for the use of any of these substances was provided.

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

A total of 175 test substances were evaluated in the five studies, of which 90 were individual chemicals and 85 were commercial products, formulations or other mixtures. Chemical classes tested included alcohols, acids, hydrocarbons, inorganic chemicals, acyl halides, alkalis, esters, heterocyclics, ketones, and organophosphates; commercial products or formulations tested included detergents/surfactants, pesticides, solvents, silicone powder, ink, paint, toilet cleaners, and thermal paper coatings.

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, CASRN, chemical and/or product class, concentration(s) tested, purity, supplier or source, and literature reference using the test substance are provided in **Appendix B**. However, if a product class was not assigned in the study report, this information was sought from other sources, including the National Library of Medicine's ChemID Plus database. Chemical classes were assigned to each test substance using a standard classification scheme, based on the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at http//www.nlm.nih.gov/mesh) that ensures consistency in classifying substances among all *in vitro* ocular test methods under consideration.. A substance could be in more than one chemical or product class. **Tables 3-1** and **3-2** show the chemical classes and some of the product classes of the test substances evaluated with the ICE test method. All of the product classes are included in **Appendix B**.

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

Chemical Class	# of Substances	Chemical Class	# of Substances
Acetate	**1**	**Inorganic Chloride Compound**	**1**
Acid	**5**	**Inorganic Salt**	**3**
Acyl halide	**1**	**Inorganic Silver/ Nitrogen Compound**	**1**
Alcohol	**15**	**Ketone**	**4**
Aldehyde	**2**	**Lactone**	**1**
Alkali	**3**	**Lipid**	**1**
Amide/Amidine	**7**	**Nitrile**	**1**
Amino Acid	**1**	**Nitro Compound**	**1**
Boron Compound	**1**	**Not Classified**	**85**
Carbohydrate	**2**	**Onium Compound**	**8**
Carboxylic Acid	**12**	**Organic Silicon Compound**	**2**
Ester	**10**	**Organic Sulfur Compound**	**3**
Ether	**1**	**Organometallic**	**2**
Heterocyclic	**9**	**Organophosphrous Compound**	**1**
Hydrocarbon	**5**	**Polycyclic**	**4**
Imide	**2**	**Polyether**	**3**
Inorganic Chemical	**1**	**Urea Compound**	**1**

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

Product Class	# of Substances	Product Class	# of Substances
Adhesive	**2**	**Fertilizer**	**1**
Antifungal	**2**	**Food Additive**	**1**
Antihistamine	**1**	**Fungicide/Germicide**	**1**
Anti-infective	**3**	**Industrial Chemical, Intermediate or Formulation**	**20**
Antiseptic	**2**	**Not Classified**	**23**
Caustic Agent	**4**	**Optical Resolution Agent**	**1**
Chlorination by-product	**1**	**Paint**	**4**
Cleaner	**8**	**Pesticide/Herbicide**	**15**
Copolymer	**3**	**Preservative**	**6**
Cosmetic Ingredient	**1**	**Pharmaceutical Compound**	**5**
Detergent	**8**	**Raw Material**	**9**
Developer	**1**	**Reagent**	**4**
Disinfectant	**5**	**Resin**	**2**
Dyes & Stains	**10**	**Silicone Resin**	**1**
Elastomer	**2**	**Soap**	**9**
Enzyme Inhibitor	**1**	**Surfactant**	**25**
Enzyme Solution	**3**	**Solvent**	**37**

As shown in **Table 3-1**, the chemical classes with the greatest amount of ICE data are alcohols, carboxylic acids, esters and heterocyclics. Of the 175 substances included in **Appendix B**, 85 substances, including formulations and mixtures of unidentified composition, could not be assigned a specific chemical class.

As shown in **Table 3-2**, the most common product classes tested in the ICE assay are industrial chemicals, solvents, soaps/surfactants and pesticides/herbicides . Other product classes tested include dyes and stains, and raw materials. Of the 175 substances included in **Appendix B,** 23 substances could not be assigned a product class.

3.3.1 Prinsen and Koëter (1993)

In this study, 21 substances were tested. Substances were provided by the Fund for the Replacement of Animals in Medical Experiments (FRAME) through Aldrich Chemicals. All substances were tested undiluted, except for acetic acid, silver(I)nitrate, sodium fluorescein, and sodium hydroxide, which were tested at a concentration of 10%, 3%, 20%, and 1% (w/v) in demineralized water, respectively. No explanation was provided for the dilutions tested. No chemical class or physicochemical characteristic (e.g., pH) information was provided, but this information was gathered based on the listed supplier for each test substance.

3.3.2 Balls et al. (1995)

In this study, the substances tested were classified as acids (4), an acyl halide (1), alcohols (9), an aldehyde (1), an alkali (1), esters (6), heterocyclics (3), hydrocarbons (2), inorganics (4), ketones (3), an organophosphate (1), pesticides (5), surfactants (6), and miscellaneous (6). The authors provided CASRNs, chemical class, sources, catalog numbers, purity, form tested, and concentration tested in the report

3.3.3 Prinsen (1996)

In this study, ICE test results for 44 substances were correlated to *in vivo* rabbit ocular irritation test results. The substances tested included formulations (3), pesticides (4), detergents (3), silicone powders (2), a lubricant (1), ink (4), paint (1), a liquid nylon product (1), solvents (10), thermal paper coatings (2), toilet cleaners (2), and individual chemicals (11). The composition of the products was not provided. There were 33 liquids, 9 solids, 1 paste, and 1 gel. No other information on physicochemical characteristics (e.g., pH) was provided.

3.3.4 Prinsen (2000)

This report contained ICE test method data for four substances: cetylpyridinium bromide (6%), cyclohexylamino-functional polymethylsiloxane (PMS), dimethylcyclopentasiloxane and Triton X-500 (5%). The EU classification for each substance was provided but the corresponding rabbit eye test data were not provided. Therefore, the EPA and GHS classifications for these substances could not be determined.

3.3.5 Prinsen 2005

In this study, ICE test results for 50 substances were correlated to *in vivo* rabbit eye test results. None of these substances was classified to a particular chemical class. The substances tested included cleaners (1), copolymers (8), disinfectants (2), dyes (2), elastomers (2), enzyme solutions (3), paints (3), pesticides (1), raw materials (8), resins (2), silicone resins (1) and soaps: surfactants (6). Eleven of the substances were not classified as to product class. Of the substances tested, 28 were liquids, 13 solids, 7 emulsions and the form tested was not provided for 2 substances.

3.4 Coding Procedures Used in the Validation Studies

The coding procedures used in the reviewed literature references were evaluated only by the information provided in the published reports. No attempt was made to obtain original study records to assess these procedures.

3.4.1 Prinsen and Koëter (1993)

No specific coding mechanisms for the substances tested are detailed, and none appear to have been used. Because only one laboratory performed the ICE test method in this study (the author's laboratory), an interlaboratory evaluation was not feasible

3.4.2 Balls et al. (1995)

Test substances and participating laboratories were each assigned a numeric code in order for subsequent data analysis to be performed without knowledge of the identities of the test

substance or the laboratory. The total number of aliquots of each test substance required for the full study was determined. Computer software was then used to generate random codes for the total number of samples, so that a unique number could be assigned to each sample.

3.4.3 Prinsen (1996)

The substances used in this study were mostly proprietary compounds. While the identity of these proprietary compounds was not provided in the publication, physicochemical properties were provided for each substance, which included chemical or product class. No specific coding methods for the substances are detailed, and do not appear to have been used. Because only one laboratory performed the ICE in this study (the author's laboratory), an interlaboratory evaluation was not feasible.

3.4.4. Prinsen (2000)

The substances used in this study were surfactants and siloxane polymers. It appears that test substances were each assigned a numeric code, although the coding mechanism was not described. Because only one laboratory performed the ICE in this study (the author's laboratory), an interlaboratory evaluation was not feasible.

3.4.5 Prinsen (2005)

The substances used in this study were mostly proprietary compounds. While the identity of these proprietary compounds was not provided in the publication, physicochemical properties were provided for each substance, which included the product class. It appears that test substances were each assigned a numeric code, although the coding mechanism was not described. Because only one laboratory performed the ICE in this study (the author's laboratory), an interlaboratory evaluation was not feasible.

4.0 *IN VIVO* REFERENCE DATA USED FOR AN ASSESSMENT OF ICE 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 4 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.

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., New Zealand White). The

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.

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 are usually those described by Draize et al. (1944) in **Table 4-1**. For current uses, other lesions,

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 liquid substances directly into the 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 post-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.

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

[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.

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).

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 8-21 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-hours observation times, or (b) the score of four or more rabbits, averaged across the 24-, 48-, and 72-hours 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

Most of the ICE studies evaluated in this BRD include *in vivo* reference data generated using the basic procedures for the *in vivo* rabbit eye test method described above. For the EC/HO validation study (Balls et al. 1995), MMAS were calculated for the 59 test substances from

existing and concurrently run *in vivo* studies, all of which were done according to OECD TG 405 (see **Table 4.2**) , following GLP guidelines. The data were generated since 1981 and met the following criteria:

- Normally used at least 3 New Zealand White rabbits tested at the same time
- A volume of 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 developed from Draize scores calculated 24 hours or more after test substance instillation. Detailed *in vivo* data for each of these substances, consisting of cornea, iris and conjunctiva scores for each animal, are available in the ECETOC Reference Chemicals data bank (ECETOC 1998). These substances have been classified by the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM) according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For the Prinsen (2005) study, *in vivo* studies were conducted in the author's laboratory subsequent to the ICE test according to OECD TG 405 (see **Table 4.2**). For most substances tested, adequate information was provided to assign an irritancy classification according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For the Prinsen (2000) study, no original *in vivo* data were provided. The irritancy classification, based on the EU system (1992) only, was provided for the four substances tested.

For the Prinsen (1996) study, *in vivo* studies were conducted in the author's laboratory according to the relevant OECD (1987) and EU (1992) guidelines on irritation testing at the time that they were conducted. Draize scores were generated based on reactions in the eyes at 1, 24, 48, and 72 hours after administration. Residual effects were recorded up to 21 days after treatment. For the *in vivo* studies, summary data and the irritancy classification, based on the EU system (1992) only, were provided. Individual animal *in vivo* data were not provided, which precluded assigning a precise classification according to the EPA (1996) and GHS (UN 2003) classification systems for most test substances. However, for some test substances, adequate information was provided in the summary data to conclude that a classification of non-severe irritancy (Category 2A, 2B, non-irritant according to GHS, or Category II, III, IV according to EPA) could be assigned. Therefore, these test substances were included in the evaluation.

For the Prinsen and Koëter (1993) study, no original *in vivo* data was provided. The published report provides the irritancy classification, based on the EU system (1992) only, for 19 of 21 chemicals, as assigned by Botham et al. (1989). The remaining two chemicals were classified based on *in vivo* studies conducted in the author's laboratory (Prinsen 1991a, 1991b, data requested but not provided). Botham et al. (1989) contains toxicological summaries that provide a recommended EU classification for each of the chemicals. In three cases, there was adequate summary *in vivo* data with which to also generate irritancy

classifications for the EPA (1996) and GHS (UN 2003) classification systems. *In vivo* rabbit eye test results were available from other sources for eight substances. Therefore, *in vivo* data were obtained for 11 of 21 chemicals tested in this study.

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 ICE 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 the ICE test method, as well as for the three other *in vitro* test methods (IRE, BCOP and 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 effects fully reverse 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., 1 out of 3 or 2 out 6 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**.

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.

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.

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**.)

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.

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).

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

Original study records containing individual animal data for the substances screened with the ICE test method in Prinsen (1996) and Prinsen (2005) were kindly provided by Mr. Menck Prinsen of TNO.

An attempt was made (by contacting the authors of the validation studies and/or the organizations that provided the comparative data) to obtain the original study records containing individual animal data for substances tested in the studies in Prinsen and Koëter (1993), Balls et al. (1995) and Prinsen (2000). However, the original study records could not be obtained and are not likely to become available.

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 (OCED

1998; EPA 2003a, 2003b; FDA 2003). These 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 ICE validation studies, were compliant with GLP guidelines is based on the information provided in the provided reports. Based on the available information, all of the reports included *in vivo* data obtained according to 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 ICE 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 ICE 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 ICE test method for predicting human ocular hazard.

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

[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.

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 within 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, etc.), 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 1-3 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%-8.3% vs. 5.4%-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%-8% for Category 2A substances, 1% for Category 2B substances, and 0% for nonirritants.

[This Page Intentionally Left Blank]

5.0 ICE TEST METHOD DATA AND RESULTS

5.1 Description of the ICE Test Method Protocols Used To Generate Data

A total of five reports, three published (Prinsen and Koëter 1993; Balls et al. 1995; Prinsen 1996) and two unpublished (Prinsen 2000; Prinsen 2005) contained sufficient data to do an accuracy analysis of the ICE test method. The test method components of the ICE protocols used in these studies (discussed in **Section 2.0**) are summarized in **Appendix A**. As discussed in **Sections 1.0** and **2.0**, only one modification to the original ICE test method protocol (Prinsen and Koëter, 1993) was made; i.e., the number of chicken eyes evaluated was reduced from five to three per test substance. Reportedly, the reduction has no effect on the overall accuracy of the ICE test method (Prinsen M, personal communication). However, a formal evaluation of the effect of the number of eyes per test substance on the accuracy or the reliability of the ICE test method has not been done. Historically, positive controls have not been used in the ICE test method, and therefore do not appear in any of the previously published protocols. The only negative control used to date has been isotonic saline – an untreated negative control. **Section 2.2.7.2** describes the need for a solvent control when a test substance is dissolved in a solvent other than water or isotonic saline.

5.2 Availability of Copies of Original Data Used to Evaluate the Accuracy and Reliability

The NICEATM staff made several attempts to obtain original ICE data for substances that had also been tested *in vivo* using the standard rabbit eye test. *Federal Register* (*FR*) notices were published on March 24, 2004 (Vol. 69. No. 57, pp. 13589-12861; available at http://iccvam.niehs.nih.gov/methods/ eyeirrit.htm) and February 28, 2005 (Vol. 70, No. 38, pp. 9661-9662; http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original ICE data, comparative *in vivo* rabbit data, as well as any human exposure data (either via ethical human studies or accidental exposure). In addition, the NICEATM staff contacted authors of published ICE studies to request original ICE data used to support the authors' conclusions. In response to these efforts, summaries of ICE results (i.e., total scores) but not original data were obtained for the 60 substances evaluated by Balls et al. (1995). NICEATM also received original study records, containing data for the substances screened with the ICE test method in Prinsen (1996), Prinsen (2000), and Prinsen (2005), kindly provided by Mr. Menk Prinsen of TNO (TNO Nutrition and Food Research, Toxicology and Applied Pharmacology, Zeist, The Netherlands).

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

As noted in **Section 2.2.12**, statistical analyses to compare ICE test method results to those from the *in vivo* reference test method have been done predominantly by comparing the ICE Irritation Index and the maximum mean scores of its individual components (i.e., corneal swelling, corneal opacity, fluorescein retention) to a numerical *in vivo* rabbit eye score (e.g., MMAS). However, because this BRD is concerned with the regulatory applicability of the ICE test method and MMAS scores are not used for regulatory classification, this approach was not taken in the analyses done for this BRD. Rather, the *in vitro* classification system

described in **Section 2.2.13** was used to assign an *in vitro* ocular irritation classification for each test substance. This approach entails calculation of mean corneal opacity, corneal swelling, and fluorescein retention scores at each time point for each test substance (see **Section 2.2.9**) and relating the maximum scores for each endpoint to an *in vitro* irritancy category. Interpretation of corneal thickness, corneal opacity, and fluorescein retention using four irritancy categories is done according to the scales shown below, provided by endpoint.

Corneal Thickness

Mean Corneal Swelling (%)	Category
0 to 5	I
> 5 to 12	II
> 12 to 18 (>75 min after treatment)	II
> 12 to 18 (≤75 min after treatment)	III
> 18 to 26	III
> 26 to 32 (>75 min after treatment)	III
> 26 to 32 (≤75 min after treatment)	IV
> 32	IV

Corneal Opacity

Mean Maximum Opacity Score	Category
0.0-0.5	I
0.6-1.5	II
1.6-2.5	III
2.6-4.0	IV

Fluorescein Retention

Mean Fluorescein Retention Score at 30 Minutes Post-treatment	Category
0.0-0.5	I
0.6-1.5	II
1.6-2.5	III
2.6-3.0	IV

The categories for each individual ICE test method endpoint can then be combined into an overall *in vitro* ocular irritancy classification for comparison to the *in vivo* ocular irritancy classification. For assigning the classification of severe irritant to a test substance, the combinations of results shown below was used (Prinsen and Koëter 1993).

Classification	**Combinations of the 3 Endpoints**
Severely Irritating	3 x IV
	2 x IV, 1 x III
	2 x IV, 1 x II[1]
	2 x IV, 1 x I[1]
	Corneal opacity ≥ 3 at 30 min (≥ 2 eyes)
	Corneal opacity = 4 at any time point (≥ 2 eyes)
	Severe loosening of the epithelium (≥ 1 eye)

[1]Combinations less likely to occur.

To date, this method has been published only as an application to the EU classification system. However, using the same classification system, ICE results have also reportedly been used to predict the *in vivo* classification of substances according to the GHS classification system (Prinsen M, personal communication). For this BRD, the *in vitro* classification was compared to the *in vivo* classification based on the EU, GHS, and EPA classification systems (EPA 1996; EU 2001; UN 2003), when feasible; i.e., when adequate *in vivo* data were available to assign a classification. To conduct this analysis, no modifications to the *in vitro* classification system were made.

Four of the five studies considered for this BRD (Prinsen and Koëter 1993; Prinsen 1996; Prinsen 2000; Prinsen 2005) assigned the *in vitro* classification of test substances based on this system. However, because one study (Balls et al. 1995) did not use this approach, the data generated in this study was used to assign an *in vitro* classification (as directed in **Section 2.2.13.1**). Once the *in vitro* classification was established for the substances tested in all relevant studies, an accuracy assessment was done for each parameter investigated (i.e., ICE classification versus the *in vivo* classification according to the rules applied by each regulatory agency, if adequate *in vivo* data were available to assign each classification).

5.4 Summary of Results

When provided, the specific information extracted for each substance included its name, CASRN (if available), chemical class, product class, concentration tested, form tested, ICE test method endpoint values (maximum mean), *in vitro* classification, and reference. No attempt was made to identify the source and purity of a test substance if the authors did not provide such information. If not provided, the CASRN was obtained from various sources, including the National Library of Medicine's ChemID database (available at http://chem2.sis.nlm.nih.gov/chemidplus). All substances with the same CASRN were listed under the same name, regardless of the synonym used in the original report. Chemical and product classes were assigned based on the classification of the National Library of Medicine's Medical Subject Heading (MeSH; available at http://www.nlm.nih.gov/mesh). **Appendix B** provides information on the names, synonyms, CASRN, and chemical/product

class, where available, for each substance while **Appendix C** contains the *in vitro* ICE test method data sorted by reference and alphabetically by substance name. The type of data contained in each study evaluated for this BRD varied, as discussed in Sections 5.4.1 to 5.4.5.

5.4.1 Prinsen and Koëter (1993)

The mean percentage corneal swelling at each time point, mean corneal opacity at each time point, and mean fluorescein retention at 30 minutes were generated for all 21 test substances. However, individual scores for each eye were not provided. No i*n vivo* scores were provided, but an irritation classification (according to the EU classification system [EU 2001]) was provided for all test substances.

5.4.2 Balls et al. (1995)

Neither the scores for each ICE test method endpoint nor the Irritation Index are included in the published report. Rather, the study report includes scatter plots showing the relationship between mean corneal swelling, mean opacity score, mean fluorescein retention score, and ICE Index score, as obtained in the lead laboratory, to the MMAS for the entire set of test substances. However, the maximum mean percentage corneal swelling and corneal opacity and the mean fluorescein retention at 30 minutes, along with the Irritation Index, was provided for all 59 test substances following a request to the European Centre for the Validation of Alternative Methods (ECVAM) by NICEATM.

5.4.3 Prinsen (1996)

Forty-four test substances were assayed in the ICE test method. Thirty-nine of the 44 test substances were evaluated in both the ICE test method and the *in vivo* rabbit eye test method. Five of the test substance were labeled as corrosive to skin and thus were not evaluated in the rabbit eye test, but rather presumed to be severely irritating to the eye (i.e., EU classification of R41 [EU 2001]). Seven substances were evaluated that had an *in vivo* classification of R41. For the *in vitro* test method, the mean percentage corneal swelling at each time point, mean corneal opacity at each time point, and mean fluorescein retention at 30 minutes were provided on all test substances, although individual eye scores were not. However, Mr. Menk Prinsen (TNO) subsequently provided this information.

5.4.4 Prinsen (2000)

This report contained ICE test method data for four substances. For the *in vitro* test method, individual eye scores for corneal thickness and corneal opacity were provided for each time point, and mean fluorescein retention at 30 minutes was provided for all test substances. The EU classification for each substance was provided, but the corresponding *in vivo* rabbit eye test data were not.

5.4.5 Prinsen (2005)

This report contained ICE test method data for 50 substances. For the *in vitro* test method, individual eye scores for corneal thickness and corneal opacity were provided for each time point, and mean fluorescein retention at 30 minutes was provided for all test substances. Corresponding *in vivo* data were also provided for each test substance, although, in some

cases, this data was inadequate to assign an irritancy classification in a particular classification system.

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 and the submitted reports. The data quality presented in the reviewed literature references can only be evaluated to the extent such information was provided in the published reports. Based on the available information, all ICE test method studies evaluated were conducted according to GLP guidelines.

Based on the information in the five studies evaluated, Balls et al. (1995) was the only study that employed specific mechanisms to code the chemicals that were tested (See **Section 3.4.2**).

5.6 Lot-to-lot Consistency of Test Substances

Ideally, a single lot of each substance is used during the validation of a test method. In situations where multiple lots of a chemical must be used, the lot-to-lot consistency of a test substance must be evaluated to ensure the consistency of the substance evaluated over the course of the study. A description of the procedures used to evaluate lot-to-lot consistency was provided in the reports. No attempt was made to review original records to assess the procedures used to evaluate different batches of substances.

One selection criterion for reference chemicals selected for the ECETOC evaluation was known high consistency and purity. Test substances for the Balls et al. (1995) evaluation were selected from the ECETOC database, and where feasible, the same source and specification was used. If obtaining the test substance from the same source and/or specification was not feasible, a test substance with a specification as close to that included in the *in vivo* testing was selected.

Based on the limited chemical information provided in the remaining reports (Prinsen and Koëter 1993; Prinsen 1996; Prinsen 2000; Prinsen 2005), and the absence of specifically cited selection criteria in these studies, an accurate assessment of lot-to-lot consistency of the test substances evaluated was not feasible. Prinsen (1996) and Prinsen (2005) appear to have used the same batch of test substances in both the ICE and *in vivo* test methods, thus ensuring an optimum level of consistency for both test methods used in these studies.

5.7 Availability of Data for External Audit

The availability of the original study records, for the reports included in the accuracy and reliability evaluation of the ICE test method, for external audit was not determined.

[This Page Intentionally Left Blank]

6.0 ICE TEST METHOD ACCURACY

6.1 Accuracy of the ICE 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 ICE test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS, EPA, and EU classification systems (EPA 1996; EU 2001; UN 2003)[1], was evaluated using two approaches. In the first approach, the performance of ICE was assessed separately for each *in vitro-in vivo* comparative study (i.e., publication) reviewed in **Sections 4.0** and **5.0**. In the second approach, an overall analysis of ICE test method accuracy was conducted by combining results from each study, and then an overall ocular irritancy classification was assigned for each substance. When the same substance was evaluated in multiple laboratories, the overall ICE ocular irritancy classification was based on the majority of calls among all of the studies. When there was an equal 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).

The three regulatory ocular hazard classification systems considered during this analysis use different decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results (see **Section 1.0**). All three classification systems are based on individual animal data in terms of the magnitude of the response and, for the EPA and GHS,

[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), as 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**).

on the extent to which induced ocular lesions fail to reverse by day 21. Thus, to evaluate the accuracy of the ICE test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at the different observation times are needed for each substance. However, these data were not consistently available in the studies considered, which limited the number of results that could be used to assess test method accuracy. 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.

This evaluation of ICE test method performance included substances evaluated in Prinsen and Koëter (1993), Balls et al. (1995), Prinsen (1996), Prinsen (2000) and Prinsen (2005). Two studies (Prinsen and Koëter 1993; Prinsen 2000) provided, for each substance tested, summary *in vivo* rabbit eye data and the corresponding ocular irritancy classification according to the EU classification system (i.e., R41, R36, nonirritating [EU 2001]). The authors did not provide the individual rabbit *in vivo* data on which this classification was based (these data were requested but not provided). Thus, irritancy classification for some of the substances tested in these studies according to the EPA and GHS systems was not possible. However, for some nonsevere irritating substances, the summary information provided by the authors could be used to assign a nonsevere irritancy classification according to the GHS (Category 2A, 2B, non-irritant [UN 2003]) or EPA (Category II, III, IV [EPA 1996]) systems. Although not helpful for assessing sensitivity or the false negative rate, inclusion of these substances in the performance evaluation did increase the numbers of nonsevere substances included in calculating specificity and the false positive rate of the ICE test method.

For the remaining studies considered (Balls et al. 1995, Prinsen 1996, and Prinsen 2005), individual animal data for the substances screened with the ICE test method were available, so most of the test substances could be assigned an irritancy classification in each of the three regulatory ocular hazard classification systems. The number of substances analyzed for each classification system is noted in the section discussing the accuracy analysis for that system.

Accuracy of ICE for Individual Studies: For the *per study* accuracy analysis, two different analyses were used. For the first analysis, the ICE ocular irritancy potential of each substance in each study under consideration was determined (**Appendix C**). For the one study where the same substance was evaluated in more than one laboratory (see Balls et al. 1995 in **Appendix C**), the ICE ocular irritancy potential for each independent test result was determined. Subsequently, an overall ICE ocular irritancy classification was assigned for each substance in this study based on the majority of ocular irritancy classification calls, (e.g., if two tests classified a substance as a nonirritant and three tests classified a substance as a severe irritant; the overall *in vitro* irritancy classification for the substance was 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 each study under consideration, the ability of the ICE test method to identify ocular corrosives and severe irritants, as defined

by the three different classification systems, was determined for each study. The *in vitro* and *in vivo* classifications assigned to each substance are provided in **Appendix D**.

In the second analysis used in the *per study* evaluation, each classification obtained when the same substance was evaluated in more than one laboratory (Balls et al. 1995) was used separately to assess test method accuracy (i.e., results were not combined across multiple tests to develop an overall ICE ocular irritancy classification). The ability of the ICE 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 tested substances.

Accuracy of ICE for Pooled Studies: For an overall analysis of ICE test method accuracy, results from all studies under consideration were combined and an ocular irritancy classification was determined for each substance. When the same substance was evaluated in more than one laboratory, the overall ICE ocular irritancy classification was based on the majority of calls among all of the laboratories in all studies under consideration (see **Appendix C**).

6.1.1 GHS Classification System: ICE Test Method Accuracy

The four studies Prinsen and Koëter (1993), Balls et al. (1995), Prinsen (1996), Prinsen (2005) contained ICE test method data on 171 substances, 144 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the GHS classification system (UN [2003])[2] (see **Appendix C**). Based on results from *in vivo* rabbit eye experiments, 30[3] of the 144 substances were classified as severe irritants (i.e., Category 1), the other 114 substances were classified as nonsevere irritants (either Category 2A, 2B) or nonirritants. The 27 substances that could not be classified according to the GHS classification system due to the lack of adequate animal data are so noted in **Appendix C**.

6.1.1.1 *Prinsen and Koëter (1993)*

Based on the available *in vivo* rabbit eye data, 10 of the 21 substances tested in this study could be assigned a GHS classification (**Table 6-1**). The remaining 11 substances had insufficient *in vivo* data for assigning a classification according to the GHS system (UN 2003). For the 10 substances that could be evaluated, the ICE test method has an accuracy of 80% (8/10), a sensitivity of 100% (2/2), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 0% (0/2)

6.1.1.2 *Balls et al (1995)*

Based on the available *in vivo* rabbit eye data, 54 of the 59 substances tested in this study could be assigned a GHS classification (**Table 6-1**). The remaining five substances had

---

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

[3] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2B classification. The accuracy analysis was performed with the substance classified as Category 1.

Table 6-1. Evaluation of the Performance of the ICE 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 and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	10/21	80	8/10	100	2/2	75	6/8	50/2/4	3/4	100	6/6	25	2/8	0	0/2
Balls et al. (1995)[4,5]	54/59	69	37/54	50	11/22	81	26/32	65	11/17	70	26/37	19	6/32	50	11/22
Balls et al. (1995)[4]	215/235	70	150/215	46	40/87	86	110/128	69	40/58	70	110/157	14	18/128	54	47/87
Prinsen (1996)	36/44	97	35/36	50	1/2	100	34/34	100	1/1	97	34/35	0	0/34	50	1/2
Prinsen (2005)	46/50	89	41/46	0	0/4	98	41/42	0	0/1	91	41/45	2	1/42	100	4/4
Entire Data Set[5,6]	144/171	83	120/144	50	15/30	92	105/114	63	15/24	88	105/120	8	9/114	50	15/30

[1]GHS = Globally Harmonized System (UN 2003).
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]No. = Data used to calculate the percentage.
[4]One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to GHS classification; the analysis was performed assuming Category 1 classification.
[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[6]Includes the data from Balls et al. (1995) using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories

inadequate *in vivo* data for assigning a classification according to the GHS system (UN [2003]). Using the first accuracy analysis approach (single call per test substance), for the 54 substances assigned a GHS classification, the ICE test method has an accuracy of 69% (37/54), a sensitivity of 50% (11/22), a specificity of 81% (26/32), a false positive rate of 19% (6/32), and a false negative rate of 50% (11/22). Using the second accuracy analysis approach (results not combined across multiple tests to develop an overall ICE ocular irritancy classification) for the 215 substances considered, the ICE test method has an accuracy of 70% (150/215), a sensitivity of 46% (40/87), a specificity of 86% (110/128), a false positive rate of 14% (18/128), and a false negative rate of 54% (47/87).

6.1.1.3 *Prinsen (1996)*

Based on the *in vivo* rabbit eye data, 36 of the 44 substances tested in this study could be assigned a GHS classification (**Table 6-1**). The remaining eight substances had inadequate *in vivo* data for assigning a classification according to the GHS system (UN 2003). For the 36 substances that could be evaluated, the ICE test method has an accuracy of 97% (35/36), a sensitivity of 50% (1/2), a specificity of 100% (34/34), a false positive rate of 0% (0/34), and a false negative rate of 50% (1/2).

6.1.1.4 *Prinsen (2005)*

Based on the available *in vivo* rabbit eye data provided in this submission, 46 of the 50 substances tested in this study could be assigned a GHS classification (**Table 6-1**). The remaining four substances had inadequate *in vivo* data for assigning a classification according to the GHS system. For the 46 substances that could be evaluated, the ICE test method has an accuracy of 89% (41/46), a sensitivity of 0% (0/4), a specificity of 98% (41/42), a false positive rate of 2% (1/42), and a false negative rate of 100% (4/4).

6.1.1.5 *Entire Data Set*

A total of 144 substances had sufficient *in vivo* data among the four studies to perform an accuracy analysis, based on the GHS classification system (**Table 6-1**). Twenty-two substances lacked sufficient *in vivo* information on which to assign a GHS classification. Based on these 144 substances, the ICE test method has an accuracy of 83% (120/144), a sensitivity of 50% (15/30), a specificity of 92% (105/114), a false positive rate of 8% (9/114), and a false negative rate of 50% (15/30).

6.1.1.6 *Discordant Results According to the GHS Classification System*

In order to evaluate discordant responses of the ICE test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. 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).

As indicated in **Table 6-2**, there were some notable trends in the performance of the ICE test method. According to the GHS classification system, the most consistently overpredicted

Table 6-2. False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the GHS[1] Classification System

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	144	8	9/114	50	15/30
Chemical Class[6]					
Alcohol	12	50	5/10	50	1/2
Amine/Amidine	5	0	0/2	33	1/3
Carboxylic acid	10	0	0/3	43	3/7
Ester	9	13	1/8	0	0/1
Heterocyclic	9	0	0/3	33	2/6
Onium compound	8	0	0/2	33	2/6
Properties of Interest					
Liquids	108	10	9/90	44	8/18
Solids	36	0	0/24	58	7/12
Pesticide	11	0	0/6	60	3/5
Surfactant – Total	21	0	0/12	56	5/9
-nonionic	4	0	0/3	100	1/1
-anionic	2	0	0/1	100	1/1
-cationic	7	0	0/1	33	2/6
pH – Total[7]	20	-	-	40	8/20
- acidic (pH < 7.0)	12	-	-	33	4/12
- basic (pH > 7.0)	8	-	-	50	4/8
Category 1 Subgroup[8]					
- Total	23[10]	-	-	35	8/23
- 4 (CO=4 at any time)	12	-	-	33	4/12
- 3 (severity/persistence)	2	-	-	50	1/2
- 2 (severity)	4	-	-	0	0/4
- 2-4 combined[9]	18	-	-	28	5/18
- 1 (persistence)	5	-	-	60	3/5

[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 ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as defined in **Appendix B**.
[7]Total number of GHS Category 1 substances for which pH information was obtained.
[8]NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: corneal opacity (CO) = 4 at any time.
[9]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.
[10]The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

(i.e., false positive[4]) substances were alcohols, which accounted for five out of nine overpredicted substances overall. Other chemical classes represented among overpredicted

[4] False positive in this context refers to a substance classified as a nonsevere (mild or moderate) irritant or nonirritant based on *in vivo* data, but as a severe irritant by the ICE test method.

substances were one each of alkalis, ketones, esters, and an unclassified substance. Regarding the physical form of overpredicted substances, eight were liquids and one (the unclassified substance) was an emulsion (which was counted as a liquid in this analysis). No solid test substances were overpredicted by the ICE test method.

According to the GHS classification system, the most consistently underpredicted (i.e., false negative[5]) substances were carboxylic acids (3), followed closely by heterocyclics (2) and onium compounds (2). Other chemical classes represented among underpredicted substances included one each of alcohols, amines/amidines/polycyclics, imides/organic sulfur compounds, inorganic salts/boron compounds and five unclassified substances. Underpredicted substances were evenly distributed regarding physical form, with seven each of solids and liquids, along with one emulsion (which was counted as a liquid in this analysis). For eight underpredicted substances for which pH data was available, four had a pH less than 7.00, ranging from 3.34 to 5.72 and four had a pH greater than 7.00, ranging from 7.18 to 9.98. Finally, for the eight underpredicted substances classified as severe irritants (GHS Category 1) for which such information was available, three were classified as severe irritants based on persistent lesions (3/5; 60%) while four were classified as severe irritants based on severe lesions (5/18; 28%).

Table 6.3 shows the effects on the ICE test method performance characteristics of excluding from the data set problematic classes (i.e., that gave the most discordant results, according to the GHS classification system). In general, exclusion of alcohols, surfactants or solids individually resulted in small changes in the performance statistics, with the exception that the exclusion of alcohols from the data set caused a two-fold decrease in the false positive rate from 8% (9/114) to 4% (4/104). Similarly, when both alcohols and surfactants were excluded from the data set, changes in the performance statistics were small, again with the exception of the effect on the false positive rate, which decreased two-fold, from 8% (9/114) to 4% (4/92). The largest changes in almost all of the performance statistics were observed when all three discordant classes were excluded from the data set; accuracy increased from 83% (120/144) to 92% (69/75), and the false negative rate decreased from 50% (15/30) to 29% (2/7). The false positive rate decreased from 8% (9/114) to 6% (4/68), but the decrease was not as large as that observed when alcohols alone or alcohols plus surfactants were removed from the data set.

6.1.2 EPA Classification System: ICE Test Method Accuracy

The four studies (Prinsen and Koëter 1993; Balls et al. 1995; Prinsen 1996; Prinsen 2005) contained ICE test method data on 171 substances, 145 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the EPA classification system (EPA 1996)[6] (see **Appendix C**). Based on results from the *in vivo* rabbit eye test, 29 of these 145 substances were classified as severe irritants (i.e., Category I), while the other 116

[5] False negative in this context refers to a substance classified as a nonsevere (mild or moderate) irritant or nonirritant by the ICE test method, but as a severe irritant based on *in vivo* data.

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

Table 6-3. Effect of Exclusion of Discordant Classes on False Negative and False Positive Rates of the ICE Test Method, for the GHS[1] Classification System

Data Set	Accuracy		False Positive Rate[2]		False Negative Rate[3]	
	%	No.[4]	%	No.	%	No.
Overall	83	120/144	8	9/114	50	15/30
w/o Alcohols	86	114/132	4	4/104	50	14/28
w/o Surfactants	85	104/123	9	9/102	48	8/18
w/o Solids	84	91/108	10	9/90	44	8/18
w/o Alcohols & Surfactants	86	96/111	4	4/92	47	9/19
w/o Alcohols & Surfactants & Solids	92	69/75	6	4/68	29	2/7

[1]GHS =- Globally Harmonized System (UN 2003).
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*
[3]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*
[4]Data used to calculate the percentage.

substances were classified as nonsevere irritants or nonirritants (Categories II, III, or IV). The 26 substances that could not be classified according to the EPA classification system are so noted in **Appendix C**.

6.1.2.1 *Prinsen and Koëter (1993)*

Based on the available *in vivo* rabbit eye data, 10 of the 21 substances tested in this study could be assigned an EPA classification (**Table 6-4**). The remaining 11 substances had inadequate *in vivo* data for assigning a classification according to the EPA system (EPA 1996). For the 10 substances that could be evaluated, the ICE test method has an accuracy of 80% (8/10), a sensitivity of 100% (2/2), a specificity of 75% (6/8), a false positive rate of 25% (2/8), and a false negative rate of 0% (0/2).

6.1.2.2 *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**). The remaining six substances had inadequate *in vivo* data for assigning a classification according to the EPA system (1996).

Table 6-4. Evaluation of the Performance of the ICE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA[1] Classification System, by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	10/21	80	8/10	100	2/2	75	6/8	50	2/4	100	6/6	25	2/8	0	0/2
Balls et al. (1995)[4,5]	53/59	72	38/53	53	10/19	82	28/34	63	10/16	76	28/37	18	6/34	47	9/19
Balls et al. (1995)[4]	211/235	74	156/211	51	38/75	87	118/136	68	38/56	76	118/155	13	18/136	49	37/75
Prinsen (1996)	36/44	97	35/36	50	1/2	100	34/34	100	1/1	97	34/35	0	0/34	50	1/2
Prinsen (2005)	46/50	89	41/46	0	0/4	98	41/42	0	0/1	91	41/45	2	1/42	100	4/4
Entire Data Set[5,6]	145/171	84	122/145	52	15/29	92	107/116	63	13/24	89	107/121	8	9/116	48	14/29

[1]EPA = U.S. Environmental Protection Agency (EPA 1996).
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]Data used to calculate the percentage.
[4]One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to EPA classification; the analysis was performed assuming Category I classification.
[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[6]Includes the data from Balls et al. (1995) using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories

Using the first accuracy analysis approach (single call per test substance), for the 53 substances assigned an EPA classification, the ICE test method has an accuracy of 72% (38/53), sensitivity of 53% (10/19), a specificity of 82% (28/34), a false positive rate of 18% (6/34), and a false negative rate of 47% (9/19). Using the second accuracy analysis approach (results not combined across multiple tests to develop an overall ICE ocular irritancy classification), for the 211 substances considered, the ICE test method has an accuracy of 74% (156/211), a sensitivity of 51% (38/75), a specificity of 87% (118/136), a false positive rate of 13% (18/136), and a false negative rate of 49% (37/75).

6.1.2.3 *Prinsen (1996)*

Based on the *in vivo* rabbit eye data, 36 of the 44 substances tested in this study could be assigned an EPA classification (**Table 6-4**). The remaining eight substances had inadequate *in vivo* data for assigning a classification according to the EPA system (1996). For the 36 substances that could be evaluated, the ICE test method has an accuracy of 97% (35/36), a sensitivity of 50% (1/2), a specificity of 100% (34/34), a false positive rate of 0% (0/34), and a false negative rate of 50% (1/2).

6.1.2.4 *Prinsen (2005)*

Based on the available *in vivo* rabbit eye data, 46 of the 50 substances tested in this study could be assigned an EPA classification (**Table 6-4**). The remaining four substances had inadequate *in vivo* data for assigning a classification according to the EPA system (1996). For the 46 substances that could be evaluated, the ICE test method has an accuracy of 89% (41/46), a sensitivity of 0% (0/4), a specificity of 98% (41/42), a false positive rate of 2% (1/42), and a false negative rate of 100% (4/4).

6.1.2.5 *Entire Data Set*

A total of 145 substances had sufficient *in vivo* data among the four studies to perform an accuracy analysis, based on the EPA classification system (**Table 6-4**). Twenty-six substances lacked sufficient *in vivo* information on which to assign an EPA classification (EPA [1996]). Based on these 145 substances, the ICE test method has an accuracy of 84% (122/145), a sensitivity of 52% (15/29), a specificity of 92% (107/116), a false positive rate of 8% (9/116) and a false negative rate of 48% (14/29).

6.1.2.6 *Discordant Results According to the EPA Classification System*

In order to evaluate discordant responses of the ICE test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. 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).

As indicated in **Table 6-5**, there were some notable trends in the performance of the ICE test method. According to the EPA classification system, the most consistently overpredicted (i.e., false positive) substances were alcohols, which accounted for five out of nine overpredicted substances overall. Other chemical classes represented among overpredicted substances, with one instance each, were alkalis, esters, ketones and one unclassified

Table 6-5. False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the EPA[1] Classification System

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	143	8	9/116	52	14/27
Chemical Class[6]					
Alcohol	12	50	5/10	50	1/2
Amine/Amidine	5	0	0/3	50	1/2
Carboxylic acid	10	0	0/3	43	3/7
Ester	9	11	1/9	0	0/0
Heterocyclic	8	0	0/3	40	2/5
Onium compound	7	0	0/2	40	2/5
Properties of Interest					
Liquids	109	10	9/92	41	7/17
Solids	34	0	0/24	70	7/10
Pesticide	11	0	0/7	50	2/4
Surfactant – Total	20	0	0/13	57	4/7
-nonionic	4	0	0/4	0	0/0
-anionic	2	0	0/1	100	1/1
-cationic	6	0	0/1	40	2/5
pH – Total[7]	16	-	-	44	7/16
- acidic (pH < 7.0)	10	-	-	40	4/10
- basic (pH > 7.0)	6	-	-	50	3/6

[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 ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as defined in **Appendix B**.
[7]Total number of EPA Category I substances for which pH information was obtained.

substance. Regarding the physical form of overpredicted substances, nine were liquids and none were solids.

According to the EPA classification system, the most consistently underpredicted (i.e., false negative) substances were carboxylic acids, which accounted for three out of 14 overpredicted substances overall. Other chemical classes represented among overpredicted substances included heterocyclics (2), onium compounds (2), imides (1), inorganic boron compounds (1), and polycyclics (1). Regarding the physical form of underpredicted substances, seven were liquids and seven were solids. For the seven underpredicted substances classified as severe irritants (EPA Category I) for which pH data was available, four had a pH less than 7.00, ranging from 3.34 to 5.72 and three had a pH greater than 7.00, ranging from 7.95 to 9.98.

6.1.3 EU Classification System: ICE Test Method Accuracy

The five studies (Prinsen and Koëter 1993; Balls et al. 1995; Prinsen 1996; Prinsen 2000; Prinsen 2005) contained ICE test method data on 175 substances, 154 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according the EU classification

system (EU 2001)[7] (see **Appendix C**). Based on results from the *in vivo* rabbit eye test, 32[8] of the 154 substances were classified as severe irritants (i.e., R41) and the other 122 substances were classified as nonsevere irritants (i.e., R36) or nonirritants. The 21 substances that could not be classified according to the EU classification system are so noted in **Appendix C**.

6.1.3.1 *Prinsen and Koëter (1993)*

All 21 substances tested in this study were included in an analysis of accuracy (**Table 6-6**). Based on the available *in vivo* rabbit eye data or the EU ocular irritancy classification for each substance provided in the published study (individual rabbit eye test data was not available for all of the substances) and using the first accuracy analysis approach (single call per test substance), the ICE test method has an accuracy of 95% (20/21), a sensitivity of 100% (7/7), a specificity of 93% (13/14), a false positive rate of 7% (1/14), and a false negative rate of 0% (0/7).

6.1.3.2 *Balls et al. (1995)*

Based on the available *in vivo* rabbit eye data, 50 of the 59 substances tested in this study could be assigned an EU classification (**Table 6-6**). Nine substances lacked sufficient *in vivo* information on which to assign an EU classification (EU 2001). For the 50 substances assigned an EU classification, the ICE test method has an accuracy of 72% (36/50), sensitivity of 53% (10/19), a specificity of 84% (26/31), a false positive rate of 16% (5/31), and a false negative rate of 47% (9/19). Using the second accuracy analysis approach (results not combined across multiple tests to develop an overall ICE ocular irritancy classification), for the 199 substances considered, the ICE test method has an accuracy of 73% (145/199), a sensitivity of 48% (36/75), a specificity of 88% (109/124), a false positive rate of 12% (15/124), and a false negative rate of 52% (39/75).

6.1.3.3 *Prinsen (1996)*

Based on the *in vivo* rabbit eye data, 36 of the 44 substances tested in this study could be assigned an EU classification (**Table 6-6**). Eight substances lacked sufficient *in vivo* information on which to assign an EU classification (EU 2001). For the 36 substances that could be evaluated, the ICE test method has an accuracy of 97% (35/36), a sensitivity of 50% (1/2), a specificity of 100% (34/34), a false positive rate of 0% (0/34), and a false negative rate of 50% (1/2).

6.1.3.4 *Prinsen (2000)*

The EU classifications were provided by the author for the four substances tested in this study that were used for the accuracy analysis (**Table 6-6**). For these substances, the ICE test method has an accuracy (4/4), sensitivity (1/1), and specificity (3/3) of 100%, and false positive (0/3) and false negative (0/1) rates of 0%.

---

[7] 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.

[8] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to EU classification. According to one test, the classification was R41, while results from the other test yielded an R36 classification. The accuracy analysis was performed with the substance classified as R41.

Table 6-6. Evaluation of the Performance of the ICE Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU[1] Classification System, by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	21/21	95	20/21	100	7/7	93	13/14	88	7/8	100	13/13	7	1/14	0	0/7
Balls et al. (1995)[4,5]	50/59	72	36/50	53	10/19	84	26/31	67	10/15	74	26/35	16	5/31	47	9/19
Balls et al. (1995)[4]	199/235	73	145/199	48	36/75	88	109/124	71	36/51	74	109/148	12	15/124	52	39/75
Prinsen (1996)	36/44	97	35/36	50	1/2	100	34/34	100	1/1	97	34/35	0	0/34	50	1/2
Prinsen (2000)	4/4	100	4/4	100	1/1	100	3/3	100	1/1	100	3/3	0	0/3	0	0/1
Prinsen (2005)	46/50	89	41/46	0	0/4	98	41/42	0	0/1	91	41/45	2	1/42	100	4/4
Entire Data Set[5,6]	154/175	87	134/154	59	19/32	94	115/122	73	19/26	90	115/128	6	7/122	41	13/32

[1]EU =- European Union System (EU 2001).
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]Data used to calculate the percentage.
[4]One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to EU classification; the analysis was performed assuming Category 1 classification.
[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.
[6]Includes the data from Balls et al. (1995) using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories

6.1.3.5 *Prinsen (2005)*

Based on the available *in vivo* rabbit eye data, 46 of the 50 substances tested in this study could be assigned an EU classification (**Table 6-6**). The remaining four substances had inadequate *in vivo* data for assigning a classification according to the EU system. For the 46 substances that could be evaluated, the ICE test method has an accuracy of 89% (41/46), a sensitivity of 0% (0/4), a specificity of 98% (41/42), a false positive rate of 2% (1/42), and a false negative rate of 100% (4/4).

6.1.3.6 *Entire Data Set*

A total of 154 substances had sufficient *in vivo* data among the five studies to perform an accuracy analysis, based on the EU classification system (**Table 6-6**). For these 154 substances, the ICE test method has an accuracy of 87% (134/154), a sensitivity of 59% (19/32), a specificity of 94% (115/122), a false positive rate of 6% (7/122), and a false negative rate of 41% (13/32).

6.1.3.7 *Discordant Results According to the EU Classification System*

As indicated in **Table 6-7**, there were some notable trends in the performance of the ICE test method. According to the EU classification system, the most consistently overpredicted (i.e., false positive) substances were alcohols, which accounted for three out of seven overpredicted substances overall. Other chemical classes represented among overpredicted substances, with one instance each, were alkalis, esters, ketones and one unclassified substance. Regarding the physical form of overpredicted substances, seven were liquids and none were solids.

According to the EU classification system, the most consistently underpredicted (i.e., false negative) substances were heterocyclics and onium compounds, with two representatives each out of 13 total underpredicted substances. Other chemical classes represented among underpredicted substances included one each of alcohols, amines/amidines, carboxylic acids, imides/organic sulfur compounds, polycyclics and polyethers. Underpredicted substances were evenly distributed with regard to physical form with six each of liquids and solids and one emulsion (counted as a liquid in this analysis). For the seven underpredicted substances classified as severe irritants (EU Category R41) for which pH data was available, three had a pH less than 7.00, ranging from 3.77 to 5.72 and four greater than 7.00, ranging from 7.18 to 9.98.

6.2 Accuracy of the ICE Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Results

While differences in results among the three hazard classification systems evaluated occurred (i.e., EPA [1996], EU [2001], and GHS [UN 2003]), the accuracy analysis revealed that the ICE test method performance was comparable among the three systems. As can be seen in **Tables 6-1**, **6-4**, and **6-6**, depending on the classification system, the overall accuracy of the ICE test method ranged from 83% to 87%. Sensitivity ranged from 50% to 59% and specificity ranged from 92% to 94%. The false positive rate ranged from 6% to 8%, while the false negative rate ranged from 41% to 50%. Given the relatively homogeneous performance of the ICE test method among the three classification systems, the discussion below encompasses all three of them, unless otherwise indicated.

Table 6-7. False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the EU[1] Classification System

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	154	6	7/122	41	13/32
Chemical Class[6]					
Alcohol	14	27	3/11	33	1/3
Carboxylic acid	10	0	0/4	17	1/6
Ester	9	13	1/8	0	0/1
Heterocyclic	9	0	0/3	33	2/6
Inorganics	5	0	0/3	50	1/2
Onium compound	8	0	0/2	33	2/6
Polyether	5	0	0/4	100	1/1
Properties of Interest					
Liquids	116	7	7/97	39	7/18
Solids	38	0	0/25	46	6/13
Pesticide	13	0	0/8	40	2/5
Surfactant – Total	24	0	0/15	44	4/9
-nonionic	5	0	0/5	0	0/0
-anionic	3	0	0/2	0	0/1
-cationic	7	0	0/1	33	2/6
pH – Total[7]	18	-	-	39	7/18
- acidic (pH < 7.0)	11	-	-	27	3/11
- basic (pH > 7.0)	7	-	-	57	4/7

[1]EU =- European Union System (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 ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as defined in **Appendix B**.
[7]Total number of EU Category R41 substances for which pH information was obtained.

6.2.1 Discordance Among Chemical Classes

According to the accuracy analysis, the chemical class with the highest false positive rate in all three classification systems was alcohols, with false positive rates ranging from 27% to 50%. The chemical class with the next highest false positive rate in all three classification systems was esters, with false positive rates ranging from 11% to 13%. No other chemical classes were consistently overpredicted by all three systems, although for most of the chemical classes tested, the number of substances in each was too few to resolve any definitive overprediction trends by the ICE test method. For the purposes of these analyses, NICEATM considered five substances per chemical class to be the threshold number for consideration, and thus classes represented by fewer than five substances were not considered.

Alcohols were also consistently underpredicted, with false negative rates ranging from 33% to 50%. Other underpredicted chemical classes were amines/amidines (33% to 50%; GHS and EPA systems only), carboxylic acids (17% to 43%), heterocyclics (33% to 40%), inorganics (50%; EU system only), onium compounds (33% to 40%) and polyethers (100%; EU system only).

6.2.2 Discordance Among Physical or Chemical Properties of Interest

Regarding the physical form of overpredicted substances, no solids were overpredicted in any classification system, while liquids showed false positive rates ranging from 7% to 10%. Both solids and liquids were underpredicted, however, showing false negative rates ranging from 46% to 70% for solids and 39% to 44% for liquids.

Exclusion of three discordant classes (i.e., alcohols, surfactants and solids) from the data set resulted in an increased accuracy (from 83% to 92%), a decreased false positive rate (from 8% to 6%) and a decreased false negative rate (from 50% to 29%).

Test substances labeled as pesticides were not overpredicted in any classification system, but showed false negative rates ranging from 40% to 60%. Test substances labeled as surfactants were also not overpredicted, but showed false negative rates ranging from 44% to 57%.

Regarding the pH of underpredicted substances for which such information was available, substances with a pH less than 7.00 showed false negative rates of 27% to 40% (3/11 to 4/10) and substances with a pH greater than 7.0 showed false negative rates of 50% to 57% (3/6 to 4/7). However, it is noted that pH information was available only a portion of the 27 to 32 severe irritant substances (i.e., Category 1, Category I, or R41) for each classification system in the database.

Finally, with respect to the GHS classification system only, as evidenced by an analysis of NICEATM-defined GHS Category 1 sub-groupings, the eight underpredicted substances were more likely to be classified *in vivo* based on persistent lesions (false negative rate of 60% [3/5]), rather than on severe lesions (false negative rate of 28% [5/18]) (**Table 6-2**)

7.0 ICE 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, positive), where applicable.

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the ICE 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 data source for conducting an assessment of ICE test method reliability was Balls et al. (1995). This study evaluated the performance and reproducibility of the ICE 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 different 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 must 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 ICE assay but, due to its excessive toxicity *in vivo*, was excluded from the comparison of *in vitro* and *in vivo* test results (see **Section 3.1.2**).

An unpublished study (Prinsen 2000) provided data from a single laboratory that tested four substances (two surfactants and two siloxanes) in four to five separate experiments, which allowed for evaluation of intralaboratory repeatability and reproducibility. Each experiment used three eyes. One of these substances was classified as a *non-irritant* (EU classification NI), two substances were classified as *irritating to the eyes* (EU classification R36) and one was classified as *severely irritating to the eyes* (EU classification R41).

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).

A CV analysis was done on within-experiment data from Prinsen (2000), using scores for each endpoint (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) and the ICE Irritation Index, for each test substance (**Table 7-1**). When considering the results of this analysis, note that some test substances had a mean or a standard deviation equal to zero for some endpoints and that scores for corneal opacity and fluorescein retention have a small dynamic range (0 to 4 and 0 to 3, respectively).

Corneal thickness measurements within experiments showed %CV values ranging from 0.9 to 6.1 and corneal opacity scores showed %CV values ranging from zero to 86.6 (the highest value was obtained for the nonirritating substance). The %CV values for fluorescein retention were zero for three of the four substances and ranged from zero to 86.6 for the non-irritating substance, although this range is based on only two experiments. Finally, the %CV values for the ICE Irritation Index for the four substances ranged from -86.6 to 41.6, with the nonirritating substance exhibiting the outlying values (-86.6 and 41.6).

7.2.2 Quantitative and Qualitative Assessment of Intralaboratory Reproducibility

Generally, analyses of intralaboratory reproducibility 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).

The data from Prinsen (2000) was also used to do a CV analysis on between-experiment values for each endpoint (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE Irritation Index, for each test substance (**Table 7-2**). When considering the results of this analysis, note that scores for corneal opacity or fluorescein retention have a small dynamic range (0 to 4 and 0 to 3, respectively).

The %CV values for the corneal thickness measurement ranged from 1.8 to 6.3 and those for corneal swelling ranged from 13.9 to 138.7. The %CV values for the corneal opacity score ranged from 8.7 to 95.8. The %CV values for the fluorescein retention score ranged from zero to 141.4. Finally, the %CV values for the ICE Irritation Index ranged from 4.1 to 91.8. Note that for all endpoints considered except corneal thickness, the highest %CV values were obtained for the non-irritating substance.

Table 7-1 Intralaboratory Repeatability of ICE Test Method Endpoints – Prinsen (2000)

Substance (Experiment No.[1])	EU[2] Class[3]	CT[4] (mean[5])	CT (%CV[6])	CS[7] (mean)	CS (%CV)	CO[8] (mean)	CO (%CV)	FR[9] (mean)	FR (%CV)	Index[10] (mean)	Index (%CV)
SP-1 (1)[11]	NI	60	3.3	0.7	346.4	0.3	86.6	0.3	86.6	15	41.6
SP-1 (2)	NI	63.3	3.3	1.7	91.6	0.3	86.6	0.5	0	18.3	39.4
SP-1 (3)	NI	62.3	2.4	2.3	24.7	0.5	0	0	-	12.3	4.7
SP-1 (4)	NI	61.7	0.9	-1.3	-86.6	0	-	0	-	-1.3	-86.6
SP-1 (5)	NI	63.3	0.9	2	0	0	-	0	-	2	0
SP-4 (1)	R36	68.7	3.0	14.3	24.5	3	0	2	0	114.3	3.1
SP-4 (2)	R36	69.3	3.0	13.3	40.0	2	0	2	0	93.3	5.3
SP-4 (3)	R36	75.7	3.3	21	23.8	2.7	21.6	2	0	114.3	14.0
SP-4 (4)	R36	69.7	4.4	14	49.5	2.7	21.6	2	0	107.3	15.1
SP-5 (5)	R36	70	3.8	12.7	27.7	2	0	2	0	92.7	3.8
SU-4 (1)	R36	72	2.4	13.7	18.4	0.7	43.3	1	0	47	16.9
SU-4 (2)	R36	68.7	3.4	14	12.4	0.7	43.3	1	0	47.3	8.5
SU-4 (3)	R36	67.7	6.0	13	15.4	0.7	43.3	1	0	46.3	9.0
SU-4 (4)	R36	66.7	3.5	11	31.5	0.8	34.6	1	0	47.7	10.6
SU-4 (5)	R36	67.7	2.2	9.7	15.8	0.7	43.3	1	0	43	16.3
SU-5 (1)	R41	77.7	1.5	23	24.2	2	0	2	0	103	5.4
SU-5 (2)	R41	74.7	4.7	20.7	19.6	2	0	2	0	100.7	4.0
SU-5 (3)	R41	75.3	6.1	21	9.5	2	0	2	0	101	2.0
SU-5 (4)	R41	76.7	2.0	16.3	25.5	1.7	34.6	2	0	89.7	16.4

[1]No. = Number.
[2]EU = European Union (EU 2001).
[3]Class. = Classification (EU 2001).
[4]CT = Corneal thickness.
[5]Mean values calculated with scores from three eyes.
[6]%CV = % Coefficient of variation.
[7]CS = Corneal swelling.
[8]CO = Corneal opacity.
[9]FR = Fluorescein retention.
[10]Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20]); No. = number.
[11]*In vivo* animal data were not provided for these substances, and therefore the EU classification that was provided by testing laboratory is presented here.

Table 7-2 Intralaboratory Reproducibility of ICE Test Method Endpoints – Prinsen (2000)

Substance (Experimental Replicates)	EU[1] Class[2]	CT[3] (mean[4])	CT (%CV[5])	CS[6] (mean)	CS (%CV)	CO[7] (mean)	CO (%CV)	FR[8] (mean)	FR (%CV)	Index[9] (mean)	Index (%CV)
SP-1 (5)[10]	NI	62.1	2.2	1.1	138.7	0.2	95.8	0.2	141.4	9.3	91.8
SP-4 (5)	R36	70.7	4.0	15.1	22.4	2.5	18.1	2	0	104.4	10.3
SU-4 (5)	R36	70.5	6.3	12.3	15.2	0.7	10.6	1	0	46.3	4.1
SU-5 (4)	R41	76.1	1.8	20.2	13.9	1.9	8.7	2	0	98.6	6.1

[1]EU = European Union (EU 2001).
[2]Class. = Classification (EU 2001).
[3]CT = Corneal thickness.
[4]Mean values calculated with scores from three eyes.
[5]%CV = % Coefficient of variation.
[6]CS = Corneal swelling.
[7]CO = Corneal opacity.
[8]FR = Fluorescein retention.
[9]Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20]); No. = Number.
[10]*In vivo* animal data were not provided for these substances, and therefore the EU classification that was provided by testing laboratory is presented here.

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 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)
- bivariant scatter diagrams/correlation analyses for pairs of laboratories to assess the extent possibility of divergence (e.g., Holzhütter et al. 1996)

7.2.3.1 *Interlaboratory Reproducibility of Hazard Classification Based on In Vitro Irritancy Classification*

In the EC/HO study reported on by Balls et al. (1995), ICE test data for an assessment of interlaboratory reproducibility was provided for four laboratories. As described in **Section 2.0**, a categorization scheme was developed that enables the assignment of a test substance, based on its activity in the ICE assay, to an ocular irritancy category that corresponds to the EU *in vivo* rabbit ocular irritancy classification system (EU 2001). This categorization scheme was used to classify the ocular irritancy potential of the 59 substances with corresponding *in vivo* rabbit eye study data tested in the ICE assay for each of the four EC/HO participating laboratories. A similar scheme was used to classify the same 59 substances according to the EPA and GHS classification systems (EPA 1996; UN 2003) for each of the four participating laboratories. The resulting *in vitro* ocular irritation classifications were used to evaluate the extent of agreement among the laboratories.

For the Balls et al. (1995) study, 19 of the 59 substances tested were assigned an overall *in vitro* classification of corrosive/severe irritant and 40 substances were assigned an overall classification of nonsevere irritant (i.e., irritants other than severe/nonirritant). For an assessment of interlaboratory reproducibility, substances classified as corrosive/severe irritants or nonsevere irritants/nonirritants were also classified (within the GHS, EPA, and EU classification schemes [EPA 1996; EU 2001; UN 2003]) by their *in vivo* rabbit eye test results. Because the focus of this assessment is on the interlaboratory reproducibility of the ICE test method 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 (e.g., three laboratories could classify a substance as a nonirritant and one laboratory could classify the same substance as a moderate irritant; for the purpose of the analysis conducted for this BRD, this would be considered 100% agreement between laboratories).

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

The four participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 (75%) of the 59 substances tested. As shown in **Table 7-3:**

Table 7-3 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 (%)
+/+	11	4[2]	7 (64)	3 (27)	1 (9)
+/-	11	4	9 (82)	2 (18)	0 (0)
-/+	6	4	1 (17)	0 (0)	5 (83)
-/-	26	4	22 (85)	4 (15)	0 (0)
?/-	3	4	3 (100)	0 (0)	0 (0)
?/+	2	4	2 (100)	0 (0)	0 (0)
TOTAL	59	4[2]	44 (75)	9 (15)	6 (10)

[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., studies were terminated too early to assess reversibility of effects), a GHS classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[2]Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three laboratories.

- All four participating laboratories agreed on the classification of seven (64%) of the 11 substances that were GHS corrosives/severe irritants[1]. Three of the four laboratories were in agreement for the three (27%) substances with discordant *in vitro* classification results among the four participating laboratories (5% benzalkonium chloride, cyclohexanol, promethazine HCl). The discordant laboratory was never the same for these three substances. In addition, two of the four laboratories were in agreement for one (9%) substance (dibenzoyl-L-tartaric acid).
- Nine (82%) of the 11 substances classified according to the GHS based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants (i.e., Category 2A and 2B irritants) or nonirritants. Of the two substances (18%) with discordant *in vitro* classification results among the four laboratories, three of the four laboratories were in agreement for both substances (10% cetylpyridinium bromide, 2,5-dimethylohexanediol). The discordant laboratory for these two substances was not the same laboratory

[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. For one chemical (trichloroacetic acid, 30%), scores for fluorescein retention and corneal swelling were not provided from one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

- One (17%) of the six substances (isobutanol) classified according to the GHS based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was incorrectly classified by the four laboratories as a corrosive/severe irritant. Of the five substances (83%) with discordant *in vitro* classification results among the four laboratories, two of the four laboratories were in agreement for all five substances (ethanol, n-hexanol, isopropanol, methyl acetate, methyl ethyl ketone). The discordant laboratories for these five substances were not consistently the same two laboratories.
- All four laboratories agreed on the classification of 22 (85%) of the 26 substances classified as GHS nonsevere irritants/nonirritants. Three of the four laboratories were in agreement for the four substances (15%) with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde, dibenzyl phosphate, methyl isobutyl ketone). The discordant laboratory for three of these four substances was always the same laboratory.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), five (8%) of the 59 test substances could not be classified according to the GHS classification scheme. Among these five substances, all four laboratories were in agreement with the classification of three substances as nonsevere irritants/nonirritants by and two substances as corrosives/severe irritants.

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

The four participating laboratories were in 100% agreement for the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 (75%) of the 59 substances tested. As shown in **Table 7-4:**

- All four participating laboratories agreed on the classification of seven (70%) of the 10 substances that were EPA corrosives/severe irritants[2]. Three of the four laboratories were in agreement for the three (30%) substances with discordant *in vitro* classification results among the four participating laboratories (benzalkonium chloride, 5%, cyclohexanol, promethazine HCl). The discordant laboratory was never the same for these three substances.
- Seven (78%) of the nine substances classified according to the EPA based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants/nonirritants. Of the two substances (22%) with discordant *in vitro* classification results among the four participating laboratories, both substances (10% cetylpyridinium bromide, 2,5-dimethylhexanediol) were incorrectly classified by three of the four laboratories. The discordant laboratory for these two substances was not the same laboratory.

---

[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. For one chemical (trichloroacetic acid, 30%), scores for fluorescein retention and corneal swelling were not provided from one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

Table 7-4 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 (%)
+/+	10	4[2]	7 (70)	3 (30)	0 (0)
+/-	9	4	7 (78)	2 (22)	0 (0)
-/+	6	4	1 (17)	0 (0)	5 (83)
-/-	28	4	24 (86)	4 (14)	0 (0)
?/-	3	4	3 (100)	0 (0)	0 (0)
?/+	3	4	2 (67)	0 (0)	1 (33)
TOTAL	59	4[2]	44 (75)	9 (15)	6 (10)

[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), an EPA classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[2]Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three laboratories.

- One (17%) of the six substances (isobutanol) classified according to the EPA based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was incorrectly classified by the four participating laboratories as a corrosive/severe irritant. Of the five substances (83%) with discordant *in vitro* classification results among the four participating laboratories, all five substances (ethanol, n-hexanol, isopropanol, methyl acetate, methyl ethyl ketone) were incorrectly classified by two of the four laboratories. The discordant laboratories for these five substances were not consistently the same two laboratories.
- All four laboratories agreed on the classification of 24 (86%) of the 28 substances that were EPA nonsevere irritants/nonirritants. Three of the four laboratories were in agreement for the four substances (14%) with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde, dibenzyl phosphate, methyl isobutyl ketone). The discordant laboratory for three of these four substances was always the same laboratory.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), six (10%) of the 59 test substances could not be classified according to the EPA classification scheme. Among these six substances, three substances were classified as nonsevere irritants/nonirritants by all four laboratories. In addition, two substances were classified as corrosives/severe irritants by all four laboratories and one substance was classified as a corrosive/severe irritant by two of the four laboratories.

7.2.3.4 *Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System*

The participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 45 (76%) of the 59 substances tested. As shown in **Table 7-5:**

Table 7-5 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 (%)
+/+	10	4[2]	6 (60)	3 (30)	1 (10)
+/-	9	4	7 (78)	2 (22)	0 (0)
-/+	5	4	1 (20)	0 (0)	4 (80)
-/-	26	4	23 (88)	3 (12)	0 (0)
?/-	5	4	5 (100)	0 (0)	0 (0)
?/+	4	4	3 (75)	0 (0)	1 (25)
TOTAL	59	4[3]	45 (76)	8 (14)	6 (10)

[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 nonsevere irritant (Category R36) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data, an EU classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[2]Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three laboratories.

- All four participating laboratories agreed on the classification of six (60%) of the 12 substances that were EU corrosives/severe irritants[3]. Three of the four laboratories were in agreement for the three (30%) substances with discordant *in vitro* classification results among the four participating laboratories (5% benzalkonium chloride, cyclohexanol, promethazine HCl). The discordant laboratory was never the same for these three substances. In addition, one (10%) substance (dibenzoyl-L-tartaric acid) was correctly classified by two of the four laboratories.

[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. For one chemical (trichloroacetic acid, 30%), scores for fluorescein retention and corneal swelling were not provided from one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

- Seven (78%) of the nine substances classified according to the EU based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants/nonirritants. Of the two substances (22%) with discordant *in vitro* classification results among the four participating laboratories, both substances (10% cetylpyridinium bromide, 2,5-dimethylhexanediol) were incorrectly classified by three of the four laboratories. The discordant laboratory for these two substances was not the same laboratory
- One (20%) of the seven substances classified according to the EU based on *in vivo* rabbit eye data as nonsevere irritants/nonirritants were incorrectly classified by the four participating laboratories as corrosives/severe irritants. Of the four substances (80%) with discordant *in vitro* classification results among the four participating laboratories, all four substances (ethanol, n-hexanol, methyl acetate, methyl ethyl ketone) were incorrectly classified by two of the four laboratories. The discordant laboratories for these five substances were not consistently the same two laboratories.
- All four laboratories agreed on the classification of 23 (88%) of the 26 substances classified as EU nonsevere irritants/nonirritants the four participating laboratories. Three of the four laboratories were in agreement for the three substances (12%) with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde, methyl isobutyl ketone). The discordant laboratory for these three substances was always the same laboratory.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), nine (15%) of the 59 test substances could not be classified according to the EU classification scheme. Among these nine substances, five substances were classified as nonsevere irritants/nonirritants by all four laboratories. In addition, three substances were classified as corrosives/severe irritants by all four laboratories and one substance was classified as a corrosive/severe irritant by two of the four laboratories.

7.2.3.5 *Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results Using the GHS Classification System*

Among the 15 substances classified according to the GHS scheme that exhibited interlaboratory differences in *in vitro* classification in the Balls et al. (1995) study, four were classified as alcohols. Two of the 15 substances were classified as cationic surfactants two were classified as acetates/esters, and two were classified as ketones. Solvents was the product class appearing most frequently among these substances, with seven of the 15 substances represented. Other product classes represented by multiple substances were chemical intermediates (five substances) and synthetic flavor ingredients (four substances). In regard to physical properties, of the 15 substances with discordant results among the four laboratories, 10 were liquid (seven water soluble) and five were solid (four water insoluble).

7.2.4 Coefficient of Variation Analysis

Mean endpoint values (i.e., fluorescein retention, corneal opacity, corneal swelling) and the ICE Irritation Index for each substance were provided for each of the four laboratories

participating in the EC/HO study. As detailed in **Section 2.2.12**, the Irritation Index is derived by calculating the sum of the maximum mean scores of each of the numerical endpoints. The opacity and fluorescein retention scores are equally weighted relative to the maximum corneal swelling obtained. Historical data from the laboratory of the developer of the ICE test method indicates that the maximum swelling observed is approximately 60-80%. Therefore, the maximum opacity (score = 4) and fluorescein retention (score = 3) scores obtained for a test substance are multiplied by a factor of 20 in order to increase their weighting (Chamberlain et al. 1997).

To provide a quantitative assessment of interlaboratory variability, individual laboratory ICE test results were used to calculate a mean, standard deviation, and the %CV for corneal opacity, fluorescein retention, corneal swelling, and the irritation index for each substance tested (**Table 7-6**). Mean and median %CV values for all 59 substances were 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 ICE, a wide range of %CV values for individual substances is evident for all endpoints. The mean/median %CV values were 39%/36% (ranging from 0-159%) for fluorescein retention, 47/37% for corneal opacity (ranging from 0-159%), 77%/75% for corneal swelling (ranging from 31-159%), and 35%/32% (ranging from 10-98%) for the Irritation Index. When only severe irritants (GHS Category 1[4], based on *in vivo* data) are considered, the %CV values are lower for all endpoints, with corneal swelling (mean of 72%, median of 69%) the sole endpoint with a mean/median %CV value greater than 35%. Of the four liquid substances with a CV < 35% for corneal swelling (2,2-dimethylbutanoic acid, 2,6-dichlorobenzoyl chloride, benzalkonium chloride 5%, and cetylpyridinium bromide 10%), two were water insoluble. No solid substances had a CV < 35% for corneal swelling. It is noteworthy that some of the corneal swelling values reported in the EC/HO data are greater than 80% (**Table 7-6**), and therefore above the reported historical maximum range of 60-80%. However, different depth measuring devices may have been used by the participating laboratories to determine corneal thickness, which, unless normalized, would have contributed to the increased variability and/or the excessive values calculated for this evaluation (Prinsen M, personal communication).

Common physicochemical characteristics do not appear among the substances showing the most variable responses (defined as CV >70% for any of the endpoints). Of the 37 substances with significant variability in at least one ICE endpoint, 18 are solids (of a total of 19 solids, 12 of which are water soluble) and 19 are liquids (of a total of 40 liquids, 14 of which are water soluble). However, some chemical classes appear to predominate among the 37 substances with CV values greater than 70%; these include seven surfactants (of 12 tested), five heterocyclic compounds (of six tested), four acetate/esters (of six tested), and four acids (of six tested). Therefore, the majority of substances tested from these chemical classes exhibited increased interlaboratory variability.

---

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

Table 7-6 Coefficient of Variation Analysis of the Interlaboratory Variability of the ICE Test Method

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
1-Naphthalene acetic acid[1]	***1.3***	***40.0***	***1.0***	***5.1***	***21.3***	***88.0***	***65.4***	***28.1***
1-Naphthalene acetic acid, Na salt	***3.0***	***0.0***	***2.8***	***8.7***	***69.2***	***52.4***	***185.3***	***19.4***
2,2-Dimethylbutanoic acid	***3.0***	***0.0***	***2.7***	***10.0***	***54.1***	***32.1***	***167.2***	***12.7***
2,5-Dimethylhexanediol	***2.1***	***33.1***	***1.7***	***56.5***	***23.5***	***115.0***	***98.7***	***57.7***
2,6-Dichlorobenzoyl chloride	2.0	11.6	1.1	37.1	18.2	30.8	80.2	14.3
2-Ethyl-1-hexanol	1.8	54.7	2.0	17.8	42.7	49.3	117.6	10.3
4-Carboxybenzaldehyde	1.2	45.6	1.3	95.2	26.4	159.4	76.2	91.0
Acetone	1.8	47.8	1.1	50.2	20.0	97.4	77.7	33.2
Ammonium nitrate	1.6	26.2	1.1	42.2	16.8	101.8	70.5	37.4
Benzalkonium chloride (1 %)	***1.9***	***29.2***	***1.9***	***55.9***	***21.6***	***80.2***	***97.5***	***23.6***
Benzalkonium chloride (10%)	***3.0***	***0.0***	***2.4***	***17.5***	***53.6***	***51.5***	***161.8***	***16.7***
Benzalkonium chloride (5%)	***2.5***	***40.0***	***2.4***	***20.4***	***45.2***	***35.7***	***143.2***	***25.2***
Dibenzoyl-L-tartaric acid	***1.7***	***51.3***	***2.3***	***29.5***	***25.2***	***132.2***	***105.7***	***38.3***
Captan 90 concentrate	***0.3***	***158.7***	***0.9***	***41.6***	***17.0***	***63.5***	***41.8***	***52.6***
Cetylpyridinium bromide (0.1%)	0.8	21.6	0.5	115.5	12.0	65.6	38.9	35.5
Cetylpyridinium bromide (10%)	***2.3***	***25.2***	***1.9***	***43.4***	***28.0***	***34.5***	***113.0***	***16.1***
Cetylpyridinium bromide (6%)	***2.4***	***20.9***	***1.3***	***46.1***	***28.6***	***52.0***	***98.1***	***21.0***

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
Chlorhexidine	***3.0***	***0.0***	***3.8***	***13.3***	***78.4***	***80.3***	***218.8***	***26.1***
Cyclohexanol	***2.8***	***14.3***	***2.3***	***10.6***	***52.2***	***66.9***	***154.3***	***24.1***
Dibenzyl phosphate	1.9	34.9	1.4	68.8	23.0	54.5	88.5	38.4
Ethanol	2.3	17.0	2.6	13.3	43.8	46.3	142.4	16.1
Ethyl acetate	2.2	26.1	2.1	7.2	36.5	72.2	121.5	20.4
Ethyl trimethyl acetate	1.1	78.3	0.7	121.0	11.6	114.0	46.9	98.3
Ethyl-2-methylacetoacetate	0.8	37.8	0.5	93.4	5.1	118.5	30.1	43.3
Fomesafen	0.7	67.0	0.7	63.1	5.9	61.1	35.1	42.3
Gammabutyrolactone	1.6	42.3	1.7	18.6	22.2	74.5	89.8	20.6
Glycerol	1.1	75.4	0.6	76.6	13.4	81.4	47.9	58.8
Imidazole	***3.0***	***0.0***	***3.1***	***20.1***	***99.8***	***87.6***	***222.2***	***38.8***
Isobutanol	3.0	3.4	2.4	12.3	61.4	40.7	168.8	16.1
Isopropanol	2.0	48.1	1.8	55.0	35.5	76.2	112.0	34.9
L-aspartic acid	1.7	28.0	1.3	44.5	21.0	113.6	82.0	50.3
Maneb	0.5	115.5	1.0	70.7	12.6	108.9	42.7	86.0
Methyl acetate	1.9	47.0	2.5	16.9	38.3	95.3	126.0	26.1
Methyl cyanoacetate	0.5	67.6	0.6	47.8	16.0	117.6	38.9	63.3
Methyl ethyl ketone	2.7	17.6	2.4	18.2	42.4	91.6	143.4	34.0

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
Methyl isobutyl ketone	2.4	20.4	2.3	20.9	31.1	72.9	125.1	32.6
Methylcyclopentane	0.6	81.6	0.4	66.7	7.5	131.9	27.0	43.0
n-Butyl acetate	1.1	50.0	2.1	18.6	25.7	53.6	89.8	14.4
n-Hexanol	2.3	60.9	2.3	32.2	46.8	62.6	137.2	23.5
n-Octanol	1.6	27.5	1.7	35.2	45.1	96.0	112.2	43.1
Parafluoraniline	3.0	0.0	2.1	4.9	46.6	46.7	147.4	14.4
Polyethylene glycol 400	1.2	65.7	0.6	76.6	14.1	61.8	49.7	31.7
Potassium cyanate	1.0	59.9	0.9	82.9	17.2	53.8	55.1	45.0
Promethazine HCl	***2.7***	***17.8***	***2.4***	***29.7***	***56.9***	***101.1***	***157.9***	***44.5***
Pyridine	***3.0***	***0.0***	***2.6***	***18.2***	***60.9***	***50.3***	***173.4***	***22.8***
Quinacrine	***1.1***	***67.2***	***0.8***	***97.2***	***8.6***	***44.2***	***47.0***	***65.3***
Sodium hydroxide (1%)	1.5	51.8	1.9	46.4	33.2	50.7	100.3	36.6
Sodium hydroxide (10%)	***3.0***	***0.0***	***3.6***	***12.3***	***111.6***	***66.6***	***243.9***	***29.9***
Sodium lauryl sulfate (3 %)	0.8	66.7	0.3	158.7	15.4	109.4	36.5	40.0
Sodium lauryl sulfate (15 %)	***1.1***	***41.6***	***0.7***	***63.4***	***15.4***	***77.1***	***49.8***	***31.8***
Sodium oxalate	***0.6***	***56.1***	***0.3***	***118.6***	***8.8***	***116.3***	***26.9***	***20.1***
Sodium perborate	***0.8***	***62.6***	***0.7***	***35.0***	***12.1***	***72.1***	***41.2***	***29.4***
Tetraaminopyrimidine sulfate	1.2	20.4	1.4	34.8	13.7	84.6	65.2	28.8

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
Toluene	1.4	29.1	1.6	32.1	26.6	87.6	86.6	31.6
Trichloroacetic acid (3%)	2.0	22.4	1.9	27.9	26.4	38.6	104.4	15.5
Trichloroacetic acid (30%)	***3.0***	***0.0***	***4.0***	***0.0***	***92.5***	***92.5***	***226.3***	***23.2***
Triton X-100 (10 %)	1.7	42.1	0.8	61.2	17.6	50.6	66.8	35.0
Triton X-100 (5 %)	1.3	35.6	0.7	145.1	22.8	81.4	62.4	42.8
Tween 20	1.2	81.7	0.6	76.6	11.7	110.0	47.9	39.9
Mean for All Substances	**1.8**	**38.8**	**1.6**	**46.8**	**32.4**	**77.2**	**100.5**	**34.8**
Median for All Substances	**1.7**	**35.6**	**1.7**	**37.1**	**25.2**	**74.5**	**89.8**	**31.8**
Range for All Substances	**0.3-3.0**	**0-158.7**	**0.3-4.0**	**0-158.7**	**5.1-111.6**	**30.8-159.4**	**26.9-243.9**	**10.3-98.3**
Mean for Severe Irritants (GHS)	**2.2**	**29.9**	**2.1**	**34.2**	**44.8**	**72.4**	**129.0**	**30.3**
Median for Severe Irritants	**2.5**	**23.0**	**2.3**	**25.0**	**36.9**	**69.5**	**128.1**	**25.6**
Range for Severe Irritants	**0.3-3.0**	**0-158.7**	**0.3-4.0**	**0-118.6**	**8.6-111.6**	**32.2-132.2**	**26.9-243.9**	**12.7-65.3**

FR = Fluorescein retention; CO = Corneal opacity; CS = Corneal swelling; Index = ICE Irritation Index; %CV = Coefficient of variation expressed as a percentage

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

7.2.5 Additional Analysis of Interlaboratory Reproducibility

In the EC/HO validation study, Balls et al. (1995) determined the interlaboratory correlation between ICE test method endpoint data generated by each laboratory for all 60 substances tested, as well as for subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances as shown in **Table 7-7** (see **Appendix E** for all correlation coefficients derived from comparing each laboratory with every other laboratory). Interlaboratory correlation coefficients generally spanned a range of 0.6 to 0.9 depending on the specific subsets of substances being evaluated. However, the range in correlation coefficients for some endpoints was larger (e.g., correlation coefficients for ICE-Mean Swelling ranged from 0.210 to 0.757 when testing substances that are insoluble in water).

Review of the mean *in vitro* data from this study indicates that wide ranges of corneal swelling values were recorded for the five insoluble test substances that were classified as ocular corrosives/severe irritants. For all five substances, the same laboratory produced the highest values, with mean corneal swelling percentages ranging from 1.5 to 6 times greater than the next highest mean corneal swelling value for the same substance tested by the other three laboratories. In addition, of the 14 remaining ocular corrosives/severe irritants (soluble and surfactant combined), a considerably higher value was reported for corneal swelling by the same laboratory for 12 substances. This trend was also apparent for nonsevere irritants/nonirritants.

Although the interlaboratory variability for fluorescein retention or corneal opacity was not as pronounced for the insoluble ocular corrosives/severe irritants and could not be associated with a single laboratory, the ranges of correlation coefficients for these endpoints are also relatively high. Therefore, the apparently large interlaboratory variability noted among these substances cannot be attributed to a single laboratory or to a single endpoint.

7.3 Historical Positive and Negative Control Data

As noted in **Section 2.0**, concurrent positive control substances have not been employed in the ICE test method, and therefore, an evaluation of historical positive control data is not possible.

At least one eye is traditionally included in each ICE study as a negative/vehicle control (isotonic saline). Individual eye data that could be used to perform a CV analysis on between-experiment values for each of the test method endpoints (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE Irritation Index for each test substance were obtained from negative control eyes. This analysis revealed that responses in the negative control eye remain relatively consistent (**Table 7-8**).

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

Index Score	Interlaboratory Pearson's Correlation Coefficient (r) of the *In Vitro* Data
Full set of test substances[1] (58-60 depending on endpoint)	
ICE-Mean Swelling	0.627-0.750
ICE-Mean Opacity Score	0.679-0.759
ICE-Mean Fluorescein Retention	0.525-0.768
ICE Index Score	0.759-0.801
Chemicals soluble in water (29-30 depending on endpoint)	
ICE-Mean Swelling	0.691-0.808
ICE-Mean Opacity Score	0.771-0.847
ICE-Mean Fluorescein Retention	0.676-0.692
ICE Index Score	0.858-0.881
Chemicals insoluble in water (17-18 depending on endpoint)	
ICE-Mean Swelling	0.210-0.757
ICE-Mean Opacity Score	0.414-0.851
ICE-Mean Fluorescein Retention	0.371-0.847
ICE Index Score	0.569-0.905
Surfactants (12)	
ICE-Mean Swelling	0.392-0.920
ICE-Mean Opacity Score	0.438-0.759
ICE-Mean Fluorescein Retention	0.546-0.950
ICE Index Score	0.724-0.854
Solids (19-20 depending on endpoint)	
ICE-Mean Swelling	0.722-0.869
ICE-Mean Opacity Score	0.595-0.868
ICE-Mean Fluorescein Retention	0.674-0.869
ICE- Index Score	0.752-0.883
Solutions (13-14 depending on endpoint)	
ICE-Mean Swelling	0.539-0.889
ICE-Mean Opacity Score	0.717-0.907
ICE-Mean Fluorescein Retention	0.543-0.901
ICE- Index Score	0.464-0.914
Liquids (26)	
ICE-Mean Swelling	0.461-0.779
ICE-Mean Opacity Score	0.692-0.770
ICE-Mean Fluorescein Retention	0.394-0.748
ICE Index Score	0.745-0.856

[1]As noted in **Section 7.1**, one substance (thiourea) was tested *in vitro* in the ICE assay but, due to its excessive toxicity *in vivo*, excluded from the comparison of *in vitro* and *in vivo* test results, and thus excluded from the evaluation in **Section 7.2.1**. However, *in vitro* data for this substance was included in the original Balls et al. (1995) analysis.

Table 7-8 Intralaboratory Reproducibility of ICE Test Method Endpoints – Negative Control (Isotonic Saline) Data

Substance (Experiment No.[1])	Max[2] Corneal Thickness	Max Corneal Swelling (%)	Max Corneal Opacity	Max Fluorescein Retention	Irritation Index[3]
Negative Control[4] (1)	63	0	0	0	0
Negative Control (2)	61	-2	0	0	-2
Negative Control (3)	63	-2	0	0	-2
Negative Control (4)	60	0	0	0	0
Negative Control (5)	62	0	0	0	0
Negative Control (6)	61	-2	0	0	-2
Negative Control (7)	62	0	0	0	0
Negative Control (8)	65	0	0	0	0
Negative Control (9)	62	-2	0	0	-2
Negative Control (10)	62	0	0	0	0
Negative Control (11)	64	2	0	0	2
Negative Control (12)	61	0	0	0	0
Negative Control (13)	64	0	0	0	0
Negative Control (14)	64	0	0	0	0
Negative Control (15)	67	2	0	0	2
Negative Control (16)	60	2	0	0	2
Mean	62.6	-0.1	0	0	-0.1
SD[5]	1.9	1.4	0	0	1.4
%CV[6]	3.0	-1088.1	-	-	-1088.1

[1]No. = Number.
[2]Max = Maximum.
[3]Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20]).
[4]Isotonic saline.
[5]SD = Standard deviation.
[6]CV = Coefficient of variation (%CV = [Standard deviation/Mean] x 100); FR = Fluorescein retention

7.4 Summary of Results

The range of %CV values for the corneal thickness measurement, when results were compared within experiments, was 0.9 to 6.1. The other endpoints evaluated produced ranges of %CV values that were larger, with variability most prominent with the nonirritating substance (SP-1). However, this could be an exaggeration of variability given the relatively small values that were produced from the nonirritating substance relative to the irritating and

corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher % CV than values of 11, 14, and 18). A similar discussion can also be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively).

The range of %CV values for the corneal thickness measurement, when results were compared across labs, was from 1.8 to 6.3. The %CV values for the remaining endpoints had a larger range (e.g., corneal swelling %CV = 13.9 to 138.7). However, if the nonirritating substance is removed, the range of %CV values is reduced (e.g., corneal swelling %CV = 13.9 to 22.4).

A qualitative assessment of the data for the EC/HO validation study (Balls et al. 1995) revealed that all four laboratories were in 100% agreement on the classification of 60% to 70% of substances classified as corrosives/severe irritants, 85% to 88% of substances classified as nonsevere irritants/nonirritants and 75% to 76% of all 59 substances considered in the study, regardless of the system used to classify the substances, when using the ICE test method.

A quantitative assessment of the data for the EC/HO validation study (Balls et al. 1995) was also done by conducting a %CV analysis for each endpoint and for the ICE Irritation Index, for each substance tested. For all substances tested, the mean/median %CV for the ICE Irritation Index was 34.8%/31.8% and 30.3%/25.6% when only substances classified as severe irritants according to the GHS classification system were considered. Historically, mean/median %CV values of <35% have been considered as satisfactory for interlaboratory reproducibility (Fentem et al. 1998).

Also, in the EC/HO validation study (Balls et al. 1995) determined interlaboratory correlation between ICE test method endpoints and the ICE Irritation Index for all substances tested and for various subsets. For all substances, the correlation coefficient for the ICE Irritation Index ranged from 0.759 to 0.801.

Analysis of the responses of negative control eyes in 16 different experiments revealed that responses were relatively consistent.

[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. GLP 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 (OECD 1998; EPA 2003a, 2003b; FDA 2003).

8.1.1 Prinsen and Koëter (1993)

The majority of chemicals used in this study (19 out of 21) were tested *in vivo* in a previous study sponsored by the Commission of the European Communities (Botham et al. 1989). The remaining two chemicals were tested at TNO Toxicology and Nutrition Institute (Prinsen 1991a, 1991b). The extent of compliance of the *in vivo* studies with GLP guidelines is not stated. However, these same chemicals were tested by Prinsen and Koëter (1993) with the ICE test method, which was reportedly conducted in accordance with GLP guidelines as outlined by OECD (1991). As noted in **Section 3.4.1**, no specific coding mechanisms for the chemicals are detailed, and none appear to have been used.

8.1.2 Balls et al. (1995)

Much of the *in vivo* reference data for this study (38 out of 60 test substances) was obtained from the ECETOC Eye Irritation Reference Data Bank (ECETOC 1992). This *in vivo* data was generated in studies carried out according to OECD TG 405 (OECD 1987) and following the principles of GLPs. *In vivo* data for an additional eight test substances was retrieved from other sources of unpublished data that met the ECETOC selection criteria (which includes GLP compliance). Therefore, it is presumed that these studies were conducted according to GLPs. The remaining 14 substances were tested *in vivo* after the ICE test method studies had begun. Again, although not specifically stated in the report, it is presumed that these studies were conducted according to GLPs in order to meet the ECETOC selection criteria.

As noted in **Section 3.4.2**, test substances and participating laboratories were each assigned a numeric code in order for subsequent data analysis to be performed without knowledge of the identities of the test substance or laboratory. The total number of aliquots of each test substance required for the full study was determined. Computer software was then used to generate random codes for the total number of samples, so that a unique number could be assigned to each sample.

8.1.3 Prinsen (1996)

All tests (both *in vivo* and *in vitro*) performed for this evaluation were reportedly conducted according to GLP guidelines as outlined by OECD (1991). As noted in **Section 3.4.3**, no coding mechanisms were employed.

8.1.4 Prinsen (2000)

All tests (both *in vivo* and *in vitro*) performed for this evaluation were reportedly conducted according to GLP guidelines as outlined by OECD (1991). As noted in **Section 3.4.4,** test substances were each assigned a numeric code, although the coding mechanism was not described.

8.1.5 Prinsen (2005)

All tests (both *in vivo* and *in vitro*) performed for this evaluation were reportedly conducted according to GLP guidelines as outlined by OECD (1991). As noted in **Section 3.4.5,** test substances were each assigned a numeric code, although the coding mechanism was not described.

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* ICE test method data included in this BRD or to obtain information about data quality audits from the authors of the ICE test method study reports. Auditing the reported endpoint values would require obtaining the original data for each ICE test method experiment, which, in most cases, is not readily available.

An informal assessment of the ICE study reports publications revealed limitations that complicate interpretation of the ICE data:

- *Incomplete substance information:* Some ICE 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 substance purity.
- *Data reporting:* A majority of the ICE studies reported only the mean *in vitro* score with no accompanying standard deviation to indicate the variability of the data.
- *Criteria for an acceptable test*: None of the reviewed reports discussed the criteria used to determine whether a test was acceptable. No information on positive control irritancy scores was provided.

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

As no reports from data quality audits have been obtained, information on GLP deviations or their impact on the study results is not available.

8.4 Availability of Laboratory Notebooks or Other Records

As noted in **Section 5.2**, original data were used for this evaluation in some cases. However, with the exception of Prinsen (1996), original data for the published studies used for this evaluation were not available for review.

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).

[This Page Intentionally Left Blank]

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

This section contains summaries of the available data from other published or unpublished studies conducted using the ICE 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 ICE data; group mean but not individual *in vivo* animal scores) precluded an assessment of the performance characteristics of ICE. However, using additional data received from the authors of the reports or from alternative sources (e.g., ECVAM), the test results on some of the substances in some of these reports were used to assess the performance of ICE. The results of these analyses are provided in **Sections 6.0** and **7.0**. This section provides a summary of reports (presented in alphabetical order by lead author) and the conclusions presented by the investigators, where such information was not available. An explanation why the data presented in a report could not be used to independently assess the performance of ICE is provided. In addition, where applicable, an explanation why some data could be used as part of the performance evaluation is provided.

9.1 Reports in the Peer Reviewed Literature

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 ICE 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 potentials 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), alkali (1), esters (6), heterocyclics (3), hydrocarbons (2), inorganic chemicals (4), ketones (3), organophosphate (1), pesticides (5), surfactants (6), and miscellaneous (6). *In vivo* data for 46 of the test substances, generated in compliance with OECD TG 405, were 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. *In vivo* data in the report were presented as MMAS. Comparison of the ICE test results to the GHS, EPA, or EU classification systems was not conducted.

As noted in **Section 5.4.2**, neither the individual substance scores for each ICE test method endpoint nor the overall Irritation Index was included in the published report. Rather, the study reports on the correlation between each ICE test method endpoint and the MMAS 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. Information about the 59 substances representing a wide-range of chemical classes and irritancy ranges tested in this study can be found in **Appendix B**.

In this study, the authors first generated X/Y scatterplots to visualize the relationship between the ICE test method results and the *in vivo* MMAS values. Spearman's rank correlation test and linear regression analysis were used to compare *in vivo* MMAS values with mean corneal swelling, mean opacity score, mean fluorescein retention score, and the ICE Index score. 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**.

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 (58-60 depending on endpoint)		
ICE-Mean Swelling	0.433-0.567	0.372-0.510
ICE-Mean Opacity Score	0.346-00.529	0.341-0.493
ICE-Mean Fluorescein Retention	0.380-0.568	0.357-0.576
ICE Index Score	0.490-0.599	0.416-0.552
Chemicals soluble in water (29-30 depending on endpoint)		
ICE-Mean Swelling	0.417-0.572	0.294-0.509
ICE-Mean Opacity Score	0.379-0.508	0.311-0.401
ICE-Mean Fluorescein Retention	0.329-0.408	0.291-0.453
ICE Index Score	0.451-0.558	0.334-0.450
Chemicals insoluble in water (17-18 depending on endpoint		
ICE-Mean Swelling	0.539-0.751	0.501-0.680
ICE-Mean Opacity Score	0.353-0.584	0.255-0.549
ICE-Mean Fluorescein Retention	0.233-0.779	0.197-0.736
ICE Index Score	0.603-0.748	0.510-0.664
Surfactants (n = 12)		
ICE-Mean Swelling	0.428-0.889	0.350-1.811
ICE-Mean Opacity Score	0.601-0.730	0.526-0.808
ICE-Mean Fluorescein Retention	0.638-0.879	0.640-0.873
ICE Index Score	0.724-0.833	0.657-0.872
Solids (19-20 depending on endpoint		
ICE-Mean Swelling	0.331-0.545	0.160-0.464
ICE-Mean Opacity Score	0.220-0.516	0.026-0.429
ICE-Mean Fluorescein Retention	0.223-0.345	0.193-0.364
ICE Index Score	0.335-0.492	0.060-0.424
Solutions (13-14 depending on endpoint		
ICE-Mean Swelling	0.471-0.853	0.342-0.823
ICE-Mean Opacity Score	0.549-0.751	0.503-0.725
ICE-Mean Fluorescein Retention	0.672-0.833	0.705-0.824
ICE Index Score	0.692-0.777	0.617-0.761
Liquids (n = 26)		
ICE-Mean Swelling	0.484-0.703	0.511-0.725
ICE-Mean Opacity Score	0.442-0.528	0.379-0.606
ICE-Mean Fluorescein Retention	0.401-0.676	0.421-0.657
ICE Index Score	0.557-0.666	0.583-0.676

The resulting analysis showed that overall, the ICE test method (based on Index Score) was not highly predictive of the MMAS (Pearson's Correlation Coefficient: 0.49 to 0.60 for the full set of test substances). Correlations with individual *in vitro* endpoints (corneal opacity, corneal swelling, and fluorescein retention) versus the MMAS also were relatively low (r = 0.35 to 0.57). Subset analyses revealed some differences among specific groups of test substances with Pearson's correlation coefficients ranging from 0.33 to 0.56 for water-soluble test substances, 0.23 to 0.78 for water insoluble test substances, 0.43 to 0.89 for surfactants, 0.22 to 0.55 for solids, 0.47 to 0.85 for solutions, and 0.40 to 0.70 for liquids.

To evaluate interlaboratory reproducibility of the ICE test method, Spearman's rank correlation coefficients and Pearson's correlation coefficients were calculated for each pair of participating laboratories for the entire test substance set, as well as for five subsets of test substances (water-soluble substances, surfactants, solids, solutions, and liquids). This analysis has been included in **Section 7.2.3**.

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 ICE for detecting ocular corrosives and severe irritants according to the GHS, EPA, or EU classification systems (EPA 1996; EU 2001; UN 2003). However, using data provided by ECVAM, NICEATM was able to evaluate the ability of the ICE 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**).

9.1.2 Chamberlain et al. (1997)

This report describes a retrospective study of various alternative ocular irritation toxicity test methods that was conducted by the U.S. Interagency Regulatory Alternatives Group (IRAG). In response to a request by IRAG to the scientific community, one ICE test method submission was received for consideration. For reasons of confidentiality, information (substances tested, sponsors) submitted to the working group was not provided in the report. The report indicated that the ICE test method protocol used by Prinsen and Koëter (1993) was used to generate ICE test method data in this study. ICE test method data on 20 substances were provided. These substances included industrial chemicals, pesticides, detergents, commercial formulations, and foodstuffs. The 20 substances included 12 liquids, six solids, one gel, and one paste. The number of substances in each chemical class and other physicochemical characteristics (e.g., pH) were provided. Since the confidential data reviewed by IRAG may have overlapped with data provided in the reports already reviewed, this evaluation was not included in the main sections of this BRD.

In vivo rabbit eye reference data were provided for 15 of the 20 substances. The remaining five substances were found to be severe irritants in the ICE test method and therefore not evaluated *in vivo*. The *in vivo* ocular MAS values for the 15 tested compounds ranged from 0 to 68. The *in vivo* rabbit ocular tests were stated to have been conducted according to OECD TG 405. The protocol used to generate the *in vivo* reference data and information on the number of substances that were identified as non-irritants, irritants, and severe irritants are not provided in the published report. However, the *in vitro* ICE and *in vivo* rabbit ocular irritation data reportedly met the guidelines developed by a separate IRAG working group for

acceptance and evaluation of data submitted for comparing *in vitro* and *in vivo* test results (Scala and Springer 1997). This guideline provides general requirements for data acceptance, criteria for acceptable *in vitro* and *in vivo* data, and criteria for the consistent review and evaluation of data. According to this guideline, GLP compliant data are assigned greater significance, but submitted data need not be collected in compliance with these guidelines. It is unknown if these data were obtained from studies conducted in compliance with GLP guidelines. The original study data has not been made available.

Individual *in vitro* endpoint scores, ICE Index scores, individual animal results, or *in vivo* MAS scores were not provided in the report. However, the *in vivo*/*in vitro* correlation between the ICE Index scores and 10 different *in vivo* endpoints (**Table 9-2**) were calculated using Pearson's correlation coefficients.

Table 9-2 *In Vitro*/*In Vivo* Correlations in Chamerberlain et al. (1997)

In Vivo Endpoint	*In Vitro*/*In Vivo* Correlation (Pearson's Correlation Coefficient; r)
MAS	0.94
Total Opacity Score	0.94
Total Area Score	0.89
Total Iris Score	0.96
Total Redness Score	0.95
Total Swelling Score	0.93
Total Score for Discharge of the Conjunctivae	0.97
Number Days to Recover Score	0.96
Total Score for All the Effects of the Conjunctivae	0.97
Total Score for All the Effects of the Cornea	0.92

MAS = Maximum Average Score

Based on the Pearson's correlation coefficients, the ICE Index Score and the *in vivo* MAS values for the 15 test substances evaluated by the IRAG working group were highly correlated (r = 0.94). Correlations with other *in vivo* endpoints (corneal opacity, swelling, etc.) also were relatively high (r = 0.89 to 0.97). No other assessments of accuracy (e.g., concordance, sensitivity, specificity, false negative and false positive rates) were conducted and could not be evaluated since original data were not provided.

9.1.3 Prinsen (1996)

The author used a similar statistical approach to that of Balls et al. (1995) to calculate Pearson's correlation coefficient. However, this study included a comparison of the ICE Irritation Index and its individual components to 14 different *in vivo* scores (including MAS). A correlation analysis of the Irritation Index Score and the *in vivo* MAS for the 39 test substances evaluated *in vitro* by Prinsen (1996) resulted in a Pearson's correlation coefficient (r) of 0.91 (**Table 9-3**), a much higher correlation than that reported by Balls et al. (1995).

Table 9-3. *In Vitro*/*In Vivo* Correlation Coefficient from Prinsen (1996)

In Vivo Endpoint	*In Vitro*/*In Vivo* correlation (Pearson's Correlation Coefficient; r)
MAS	0.91
Total Opacity Score	0.87
Total Area Score	0.86
Total Iris Score	0.92
Total Redness Score	0.88
Total Swelling Score	0.90
Total Score for Discharge of the Conjunctivae	0.92
Number Days to Recover Score	0.88
Total Score for All the Effects of the Conjunctivae	0.92

MAS = Maximum Average Score

Correlations with the remaining 13 individual *in vivo* endpoints were also relatively high (r = 0.86 to 0.92), as were the correlations of individual *in vitro* endpoints (corneal swelling, opacity, and fluorescein retention) to the MAS (0.83 to 0.92). A list of the substances tested in this study is provided in **Appendix C**.

The data also showed that all of the substances defined as corrosive were classified as having a risk of causing serious eye damage (EU classification R41 [EU 2001]) by the ICE test method. However, because the MAS is not used for regulatory classification, this evaluation was not included in the main sections of this BRD.

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 ICE test method data and *in vivo* reference data, was published on March 24, 2004. In addition, authors of published ICE studies were contacted to request original ICE data and *in vivo* reference data. In response to the *FR* notice, Procter and Gamble submitted ICE test method data and *in vivo* rabbit eye test data. Original data for the Prinsen and Koëter (1993) and the Prinsen (1996) studies could not be obtained.

9.2.1 Procter and Gamble (P&G) Submission from Drs. Daniel Marsman and Karen Acuff

On behalf of P&G, Drs. Daniel Marsman and Karen Acuff submitted sets of ICE test method data from studies performed to evaluate the ability of the ICE test method to discriminate between chemicals and benchmark proprietary formulations representing several different consumer laundry and cleaning products with varying eye irritation potential. The report notes that the ICE test method studies were conducted at TNO Nutrition and Food Institute, and provides the TNO protocol. This ICE test method protocol is the same as that used to

generate the proposed standardized protocol. ICE test method data on 28 substances were provided. These substances included surfactant raw materials, light duty dishwashing liquids, heavy-duty liquid laundry detergents, bleach containing laundry additives, and fabric enhancers. All of the formulations tested were liquid or surfactant solutions. The quantitative composition of each formulation was provided using both generalized and specific chemical information (e.g., total nonionic surfactants = 5 to 10%, sodium xylene sulfonate = 1 to 4%). The number of substances in each chemical class and other physicochemical characteristics (e.g., pH) were provided. The *in vivo* reference data used to compare the ICE test method was obtained using the low volume eye test (LVET) or available human data. The LVET varies from the traditional *in vivo* rabbit eye test by using an application of only 10 μL of a test substance, rather than the traditional 100 μL, in an attempt to reduce the amount of pain and suffering potentially experienced by test animals. The MAS and days to clearing (DTC) were provided for each test substance along with the EU classification. No individual animal data was provided for any of the test substances. Because reference data were not generated with the standard *in vivo* protocol, this evaluation was not included in the main sections of this BRD.

Mean maximum *in vitro* endpoint scores (along with the time of occurrence) were provided in the report, along with histopathology findings (when performed), and the predicted EU classification, but no individual eye scores were included. No statistical analysis was performed. Rather, a simple numerical assessment of the extent to which the ICE test method accurately predicted the *in vivo* classification was performed. **Table 9-4** provides the comparative results of each test substance. Several of the test substances were not assigned an EC classification based on the *in vivo* test, but rather were included in the evaluation based on accidental human exposure.

As demonstrated in **Table 9-4**, the ICE test method was able to accurately discriminate between surfactant raw materials (Adogen 444 [50%] and benzalkonium chloride [10%]) classified as severe irritants, by correctly determining the EU classification (based on LVET results) for both substances. However, results with various formulations were somewhat less predictive. With regard to light duty dishwashing liquids, the ICE test method accurately predicted the EU classification of three of the four test substances evaluated. The one remaining test substance reportedly would have been correctly predicted if histopathological findings had been included in the evaluation.

For the heavy-duty liquid laundry detergents, only one test substance was assigned an *in vivo* EU classification, which was marginally predicted by the ICE test method (i.e., AISE B5 was classified by the LVET as a borderline NI/R36 and the ICE test method classified it as an NI). The remaining six test substances were included based on mild, reversible effects noted in humans. The ICE test method predicted that all would be either NI or NI/R36.

For the bleach-containing laundry additives, the ICE test method correctly predicted the EU classification of only one of the five substances tested, underpredicting one severe irritant, and overpredicting three nonirritants. However, P&G stated that some of these errors might be corrected by including histopathology.

Table 9-4 EU Classification of P&G Consumer Laundry/Cleaning Products Based on the LVET and the ICE Test Methods

Test Material	*In Vivo* Classification (EU)	*In Vitro* Classification (EU)
Adogen 444 (50%)	R41	R41
Benzalkonium chloride (10%)	R41	R41
LDL645	R41	R36(R41[1])
Peroxi695*	R41	R36
LDL298*	R36/41	NI
Neodol 45-7	R36	R36
Peroxi694*	R36	R36
FE1828	NP	R36
FE2586	NP	NI(R36[6])
FE2587	NP	NI
FE2588	NP	R36
FE2589	NP	NI
FE2592*	NP	NI
HDL1813[4]	NP	NI
HDL1814[4]	NP	NI
HDL1815*[4]	NP	NI/R36
HDL2209[4]	NP	NI/R36
HDL2591*[4]	NP	NI
HDL809*[4]	NP	NI
AISE B5*[3]	NI/R36	NI
AISE C16[2]	NI/R36	NI
LDL659	NI/R36	R36
5% Sodium lauryl sulfate	NI	NI
FE2590[6]	NI	R41
Hypo580[6]	NI	R41
Hypo686[6]	NI	R41
Peroxi696[6]	NI	R36

Abbreviations: FE = Fabric enhancer; HDL = Heavy duty liquid laundry detergent; Hypo = Hypochlorite-containing bleach; LDL = Light duty dishwashing liquid; LVET = Low Volume Eye Test; Peroxi = Hydrogen peroxide-containing bleach; NP = Not provided

[1]Classification could be upgraded to R41 based on histopathology

[2]Formulation administered to 10 human volunteers. Corneal and conjunctival effects were observed that cleared within 24 hours

[3]Formulation administered to 10 human volunteers. Corneal and conjunctival effects were observed that cleared within 48 hours

[4]Corneal effects following accidental exposures to the human eye cleared within 1-2 days, with an occasional case taking up to 2 weeks to clear.

[5]Classification could be upgraded to R36 based on histopathology.

[6]Designated as a benchmark formulation for the particular category.

Finally, the ICE test method overpredicted the classification of the one fabric enhancer for which such data were provided. The authors state that these test substances are non- to very slightly irritating to the eye. However, the basis for this statement is not provided.

Therefore, the ICE test method provided variable results in this study when compared to the classification based on the LVET, particularly with respect to consumer formulations.

Although a total of 27 substances were tested in the ICE test method, only 15 were presented with *in vivo* EU classifications. Of these 15 test substances, the ICE test method accurately predicted the EC classification of eight. In addition, eight of these 15 test substances were designated as benchmark formulations for their respective category, of which the ICE test method accurately predicted the EU classification of only two. However, P&G commented on several occasions that the predictivity of the ICE test method could be enhanced if histopathological findings were included in the evaluation.

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

10.1 How the ICE 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 ICE test method refines animal use. Since these animals are being humanely killed for non-laboratory purposes, there is no additional infliction of animal pain or distress caused by the testing procedure. Furthermore, substances that are identified as corrosive or severe irritants *in vitro* would be excluded from *in vivo* testing, thus sparing rabbits from the pain associated with these types of substances.

The ICE test method can also reduce animal use through two different mechanisms. The ICE test method was adapted from the IRE test method in order to use an animal species routinely raised as a food source in large numbers to replace the need for laboratory animals. Additionally, with the acceptance of a positive outcome (i.e., classification of a substance as a severe ocular irritant) from the *in vitro* method, the animals that would have been used in the *in vivo* rabbit eye test would be spared.

10.2 Requirement for the Use of Animals

Although chickens are required as a source of corneas for this organotypic *in vitro* assay, only chickens humanely killed for food or other non-laboratory purposes are 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, the level of personnel training, costs of the *in vitro* test method, and time to complete the method, are necessary. This consideration provides additional information as to whether the time, personnel cost, and effort required to conduct the test method are considered reasonable.

11.1 Transferability of the ICE Test Method

Test method transferability addresses the capacity of a method to be accurately and reliably performed by multiple 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 Required to Conduct the ICE Test Method

The capital requirements needed to outfit a laboratory to conduct the ICE test method are relatively minor, with the exception of the need for a slit-lamp biomicroscope equipped with an optical pachymeter. Along with the superfusion apparatus and eye clamps, these are the major items required in setting up an ICE test method-capable facility. It is quite possible that a facility that is presently involved in ocular toxicology would already possess or have access to a slit-lamp microscope. If necessary, such a set-up could be purchased through used equipment vendors for as little as $2,500. Due to their novelty, the superfusion apparatus and eye clamps would most likely require fabrication based on diagrams and or photographs provided by the test method developer. The actual cost of these components is not readily available. While the requisite peristaltic and vacuum pumps are typically commonplace in the laboratory, if necessary they could be purchased commercially for less than $1,000. There are no specific requirements regarding the facility at which the test is conducted (e.g., sterile environment). However, it would seem appropriate to conduct the assay under controlled temperature and humidity conditions. Should histopathology be included as a component of the ICE test method, standard tissue processing, sectioning, and staining equipment would be required at a significant additional cost. Most likely, if a facility is not already equipped to perform such tasks, this portion of the test method could be outsourced to an appropriate contractor.

The *in vivo* test, in contrast, requires a facility that is approved to house live laboratory animals; one that maintains constant, tightly regulated atmospheric conditions (i.e., temperature and humidity). In fact, the primary expense for equipping a facility to conduct the *in vivo* rabbit test would be the acquisition of an adequate animal room and associated housing (e.g., cages, bedding, food, water, etc.) for boarding animals during the study. There are no additional major equipment requirements as the remaining equipment and supplies necessary for conducting the *in vivo* test are readily available in most laboratories.

None of the equipment used for the ICE test method is fixed. Therefore, it is essentially a portable assay, provided adequate space is available to set up the slit-lamp microscope and the superfusion apparatus (a small table would suffice), and water and electrical access are available. All of the components of the assay can be readily transported to another facility if necessary. A sufficiently stable tabletop surface, that is free of major vibrations, is required for accurate use of the slit-lamp microscope (i.e., to avoid blurring the image under the microscope). Therefore, it has been suggested that, if a poultry slaughterhouse is not in sufficient proximity to laboratory, the ICE test method could be conducted at the slaughterhouse.

11.1.2 Availability of Other Necessary Equipment and Supplies

The remaining equipment and supplies necessary for conduct of the ICE test method are readily available in most scientific laboratories, or can be obtained from several scientific laboratory equipment vendors.

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 ICE Test Method Training Considerations

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

Conducting the ICE test method (i.e., set-up and dosing of the eyes) appears to involve minimal training of technical staff, and could likely be mastered in a short period of time. However, mastering the evaluation of results at the requisite time points may require additional training. Both the *in vivo* rabbit eye test and ICE test methods incorporate the qualitative assessment of corneal opacity as an endpoint in the evaluation of ocular irritancy. The ICE test method also includes a qualitative measurement of fluorescein retention. Therefore, it is essential in both cases that laboratory personnel be adequately trained to accurately and consistently identify these endpoints. Lastly, the use of a slit-lamp microscope is necessary to evaluate corneal thickness. Accurate recording of this quantitative measurement requires that the technician be trained in the proper use of this instrument. There is no precise level of training that defines when a technician is adequately trained. Rather, this must be demonstrated through experience with the oversight of an experienced supervisor. Once the technician has demonstrated competence in identifying the study endpoints, it would seem appropriate for routine assessments of observations among trained personnel using benchmark control test substances to ensure consistency. A training video or other visual media to provide guidance on the development of endpoints may be considered for use.

11.3 Cost Considerations

As it is currently used at TNO (TNO Nutrition and Food Research, Toxicology and Applied Pharmacology, Zeist, The Netherlands), the ICE test is incorporated as a prescreen for the *in vivo* rabbit test without additional costs. If the prescreen shows that severe irritancy (as defined by the EU classification system) is expected, a full ICE test is performed without

further *in vivo* testing at the price of the *in vivo* test. If a full ICE test is used as a stand-alone assay (as mandated in EU countries for cosmetics/household products), depending on the number of samples tested, the 2004 cost of the ICE ranges from $847 to $1,694 (Prinsen M, personal communication). However, these costs do not include the inclusion of a positive control, which would increase the cost of the assay. In comparison, a GLP-compliant EPA OPPTS Series 870 Acute Eye Irritation test in the rabbit ranges from $765 for a 3 day/3 animal study up to $1665 for a 21 day/3 animal study at MB Research Laboratories (MB Research laboratories, personal communication). Therefore, it would appear that the cost, based on conducting Good Laboratory Practice (GLP) compliant studies, of an ICE test is comparable to, if not less expensive than, that of an *in vivo* rabbit test.

11.4 Time Considerations

Use of the ICE 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 (although it is recognized that a corrosive response could be determined in less than 24 hr). 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 ICE 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:American Society for Testing and Materials.

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

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].

Bone J. 1979. Anatomy and Physiology of the Fowl: The Organs of Special Sense. In: Animal Anatomy and Physiology. Reston Publishing Company, Inc., 533-534.

Botham JW, McKillop CM, Purchase IFH. 1989. Acute and topical toxicity profiles for substances involved in the in vitro validation of the in vivo ocular irritation model. In: Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test. EC Document XI/632/91, V/E/1/131/91. Commission of the European Communities.

Brevard TA, Calvert GM, Blondell JM, Mehler LN. 2003. Acute occupational disinfection-related illness among youth, 1993-1998. Environ Health Perspect 111:1654-1659.

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. Toxicol In Vitro 12:669-690.

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

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

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. Wiley Interscience: New York.

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.

Commission of the European Communities 1991. Collaborative study on the evaluation of alternative methods to the eye irritation test. EC Document XI/632/91, V/E/1/131/91.

Chamberlain M, Gad SC, Gautheron P, Prinsen MK. 1997. IRAG Working Group I: Organotypic models for the assessment/prediction of ocular irritation. Food Chem Toxicol 35:23-37.

Clean Air Act Amendments of 1990. Public Law 101-549. Available: http://www.ehso.com/caa.htm [accessed 26 August 2004].

Cooper KJ, Earl LK, Harbell J, Raabe H. 2001. Prediction of ocular irritancy of prototype shampoo formulations by the isolated rabbit eye (IRE) test and bovine corneal opacity and permeability (BCOP) assay. Toxicol 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 Exper 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.

Earl L. 1998. Experience with other isolated eye models: Isolated rabbit eye (IRE). BCOP Technical Workshop. 11:342-345.

ECETOC. 1992. Eye Irritation – Reference Chemicals Data Bank. Technical Report No. 48. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

ECETOC. 1998. Eye Irritation – Reference Chemicals Data Bank. Technical Report No. 48(2). Brussels:European Centre for Ecotoxicology and Toxicology of Chemicals.

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

EPA. 1998. Health Effects Test Guideline, OPPTS 870.2400 Acute Eye Irritation. EPA 712-C-98-195. Washington:U.S. 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. 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 labelling of dangerous substances. Official Journal of the European Communities L255:1-333.

EU. 2004. Commission Directive 2004/73/EC of 29 April 2004 adapting to technical progress for the 29th 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 Union L152:1-316.

FDA. 2003. Good laboratory practice for nonclinical laboratory studies. 21CFR58.

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

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

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

Fentem J, Archer GEB, Balls M, Botham PA, Curren R, Earl LK, Esdaile DJ, Holzhütter H-G, Liebsch M. 1998. The ECVAM international validation study on in vitro tests for skin corrosivity: 2. Results and evaluation by the Management Team. Toxicol In Vitro, 12:483-524.

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

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.

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.

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

Harbell JW, Curren RD. 2002. In vitro methods for the prediction of ocular and dermal toxicity. In Handbook of Toxicology (Derelanko MJ and Hollinger MA, eds). CRC Press: Boca Raton. 835-866.

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjöström 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 No. 80: Chicken enucleated eye test (CEET). Available: http://ecvam-sis.jrc.it/invittox/static/index.html. [accessed 18 February 2004].

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, 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.

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, Volume 3, New York:Mary Ann Liebert, Inc., 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.

Liebsch M, Spielmann H. 2002. Currently available in vitro methods used in the regulatory toxicology. Toxicol Letters. 127:127-134

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, 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, 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. 36:106-17

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, Seabaugh V, Shadduck JA, Edelhauser HF. 1987. Eye irritation. In: Dermatotoxicology (Marzulli FN, Maibach HI, eds). Washington:Hemisphere Publishing Corporation, 641-696.

McLaughlin RS. 1946. Chemical burns of the human cornea. Am J Opthalmol. 29:1355-1362.

MeSH Medical Subject Heading. Available: http://www.nlm.nih.gov/mesh [accessed 12 July 2005].

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].

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

OECD. 1987. Test guideline 405, Acute eye irritation/corrosion, adopted February 24, 1987. In OECD Guidelines for Testing of Chemicals. OECD, Paris.

OECD. 1991. Series on Principles of Good Laboratory Practice and Compliance Monitoring, Number 1: The OECD Principles of Good Laboratory Practice. 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.

Oksala O, Stjernschantz J. 1988. Effects of calcitonin gene-related peptide in the eye. Inv Opth Vis Sci. 29(7):1006-1011.

Parish W. 1985. Ability of in vitro (corneal injury-eye organ-and chorioallantoic membrane) tests to represent histopathological features of acute eye inflammation. Food Chem Toxicol 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.

Prinsen MK. 1991a. Acute eye irritation/corrosion study with 1-butanol in albino rabbits. TNO-report V91.187/210069, November 1991. Dept. of Biological Toxicology, TNO Toxicology and Nutrition Institute, Zeist, The Netherlands. (Available on request)

Prinsen MK. 1991b. Acute eye irritation/corrosion study with 2-methoxyethanol in albino rabbits. TNO-report V91.188/210069, November 1991. Dept. of Biological Toxicology, TNO Toxicology and Nutrition Institute, Zeist, The Netherlands. (Available on request)

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

Prinsen MK. 2000. Chicken encleated eye test with reference surfactants and siloxane polymers in Phase II of the reference standard validation project; an *ex vivo* alternative to the Draize eye irritation test with rabbits. TNO Report V99.521b. Unpublished report provided directly to NICEATM by M Prinsen, TNO Nutrition and Food Research Institute.

Prinsen MK. 2005. *In vitro* and *in vivo* data for 94 substances tested in the isolated chicken eye test. Unpublished data provided directly to NICEATM by MK Prinsen, TNO Nutrition and Food Research Institute.

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

Russell WMS, Burch RL. 1992. The Principles of Humane Experimental Technique. 14th ed. South Mimms, Potters Bar, Herts:Universities Federation for Animal Welfare.

Scala RA, Springer J. 1997. IRAG working group 6. Guidelines for the evaluation of eye irritation alternative tests: criteria for data submission. Interagency Regulatory Alternatives Group. Food Chem Toxicol 35:13-22.

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 Toxicol 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.

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.

Wilhelmus K. 2001. The Draize eye test. Survey of Ophthalmology. 45:493-515.

Wilkie DA and Wyman M. 1991. Comparative anatomy and physiology of the mammalian eye. In Dermal and Ocular Toxicology, Fundamentals and Methods (Ed. DW Hobson). BocaRaton:CRC Press, Inc., 433-491.

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.

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.

[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.

CEET: Chicken Enucleated Eye Test; the original name of the test method referred to in this BRD as ICE.

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 ICE test method.

[2] Definition used by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM 2003).

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: A subjective 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.

Corneal Swelling: An objective measurement in the ICE 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 ICE test. Increased corneal swelling is indicative of damage to the corneal epithelium.

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.

***Ex vivo*:** Outside of the living organism. Refers to assays conducted on a component(s) of a living organism in an artificial environment outside of the living organism (e.g., an enucleated eye).

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; called also *tunica fibrosa oculi*.

Fluorescein retention: A subjective measurement in the ICE 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.

Irritation Index: A value calculated by summing the maximum mean scores of each of the ICE test method endpoints (corneal opacity, corneal swelling, and fluorescein retention). In order to increase their weighting relative to the corneal swelling value, the maximum corneal opacity and fluorescein retention scores obtained are multiplied by a factor of 20. Therefore, the irritation index has a possible range of 0 to 200.

Negative control: An untreated sample containing all components of a test system, except the test substance solvent, which is replaced with a known non-reactive 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.

Nicititating 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 that 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.

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. A pH of 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, including the listing of all necessary reagents, criteria and procedures for the 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.

Secondary bacterial keratitis: Inflammation of the cornea that occurs secondary to another insult that compromised the integrity 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).

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.

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 ICE test method, this instrument is used to view the anterior structures of the chicken eye as well as to objectively measure corneal thickness with a depth-measuring device attachment.

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: A custom-built experimental setup for the ICE test that provides a controlled environment for short-term maintenance of the metabolic and physiological activity of the isolated chicken eye and a continuous flow of isotonic saline over the ocular surface.

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.