Draft Background Review Document

Current Status of *In Vitro* Test Methods for Identifying Mild/Moderate Ocular Irritants:

The Isolated Chicken Eye (ICE) Test Method

Interagency Coordinating Committee on the Validation of Alternative Methods

National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

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

March 2009
NIH Publication No.

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

20 ***[This Page Intentionally Left Blank]***
21

Table of Contents

75

List of Tables

166 # List of Figures

169

List of Abbreviations and Acronyms

BCOP	Bovine Corneal Opacity and Permeability
BRD	Background Review Document
CASRN	Chemical Abstracts Service Registry Number
CEET	Chicken Enucleated Eye Test
CPSC	U.S. Consumer Product Safety Commission
EC	European Commission
EC/HO	European Commission/British Home Office
ECVAM	European Center for the Validation of Alternative Methods
EPA	U.S. Environmental Protection Agency
EU	European Union
FDA	U.S. Food and Drug Administration
FR	*Federal Register*
FRAME	Fund for the Replacement of Animals in Medical Experiments
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
INVITTOX	*In Vitro* Techniques in Toxicology (ERGATT FRAME ECVAM Data bank)
IRE	Isolated Rabbit Eye
MeSH	U.S. National Library of Medicine's Medical Subject Heading
MMAS	Modified maximum average score

NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NIH	National Institutes of Health
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
TNO	TNO Nutrition and Food Research
UN	United Nations
ZEBET	German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments

Interagency Coordinating Committee on the Validation of Alternative Methods: Agency Representatives

Agency for Toxic Substances and Disease Registry
• Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission
• Marilyn L. Wind, Ph.D. (Chair)
◊ Kristina Hatlelid, Ph.D.
Joanna Matheson, Ph.D.

Department of Agriculture
• Jodie Kulpa-Eddy, D.V.M. (Vice-Chair)
◊ Elizabeth Goldentyer, D.V.M.

Department of Defense
• Robert E. Foster, Ph.D.
◊ Patty Decot
Peter J. Schultheiss, D.V.M., D.A.C.L.A.M.
Harry Salem, Ph.D.

Department of Energy
• Michael Kuperberg, Ph.D.
◊ Marvin Stodolsky, Ph.D.

Department of the Interior
• Barnett A. Rattner, Ph.D.
◊ Sarah Gerould, Ph.D.

Department of Transportation
• George Cushmac, Ph.D.
◊ Steve Hwang, Ph.D.

Environmental Protection Agency
Office of Science Coordination and Policy
• Karen Hamernik, Ph.D.
Office of Research and Development
◊ Julian Preston, Ph.D.
Office of Pesticide Programs
Deborah McCall
OECD Test Guidelines Program
Jerry Smrchek, Ph.D.

• Principal agency representative
◊ Alternate principal agency representative

Food and Drug Administration
Office of Science
• Suzanne Fitzpatrick, Ph.D., D.A.B.T.
Center for Drug Evaluation and Research
◊ Abigail C. Jacobs, Ph.D.
Paul C. Brown, Ph.D.
Center for Devices and Radiological Health
Melvin E. Stratmeyer, Ph.D.
Vasant G. Malshet, Ph.D., D.A.B.T.
Center for Biologics Evaluation and Research
Richard McFarland, Ph.D., M.D.
Ying Huang, Ph.D.
Center for Food Safety and Nutrition
David G. Hattan, Ph.D.
Robert L. Bronaugh, Ph.D.
Center for Veterinary Medicine
Devaraya Jagannath, Ph.D.
M. Cecilia Aguila, D.V.M.
National Center for Toxicological Research
William T. Allaben, Ph.D.
Paul Howard, Ph.D.
Donna Mendrick, Ph.D.
Office of Regulatory Affairs
Lawrence D'Hoostelaere, Ph.D.

National Cancer Institute
• T. Kevin Howcroft, Ph.D.
◊ Alan Poland, M.D.

National Institute of Environmental Health Sciences
• William S. Stokes, D.V.M., D.A.C.L.A.M
◊ Raymond R. Tice, Ph.D.
Rajendra S. Chhabra, Ph.D., D.A.B.T.
Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
• Paul Nicolaysen, V.M.D.
◊ K. Murali Rao, M.D., Ph.D.

National Institutes of Health
• Margaret D. Snyder, Ph.D.

National Library of Medicine
• Pertti (Bert) Hakkinen, Ph.D.
◊ Jeanne Goshorn, M.S.

Occupational Safety and Health Administration
• Surender Ahir, Ph.D.

[This Page Intentionally Left Blank]

Acknowledgements

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

U.S. Consumer Product Safety Commission
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D., (ICCVAM Chair)

Department of Defense
Harry Salem, Ph.D.

Department of Transportation
Steve Hwang, Ph.D.

U.S. Environmental Protection Agency
Office of Pesticide Programs
Meta Bonner, Ph.D.
Jonathan Chen, Ph.D.
Masih Hashim, D.V.M., Ph.D.
Karen Hicks
Marianne Lewis
Deborah McCall
Timothy McMahon, Ph.D.
Mark Perry, Ph.D.
John Redden, Ph.D.
Amy Rispin, Ph.D.
Jenny Tao, Ph.D.

Office of Research and Development
Andrew Geller, Ph.D.

Office of Science Coordination and Policy
Karen Hamernik, Ph.D. (OTWG Co-Chair)

U.S. Food and Drug Administration

Center for Drug Evaluation and Research
Paul C. Brown, Ph.D.
Abigail Jacobs, Ph.D.
Jill Merrill, Ph.D. (OTWG Co-Chair)

Center for Food Science and Nutrition
Robert Bronaugh, Ph.D.
Donnie Lowther

Office of Science and Health Coordination
Suzanne Fitzpatrick, Ph.D., D.A.B.T.

National Institute of Environmental Health Sciences
Mark Cesta, D.V.M, D.A.C.V.P.
Raymond (Buck) Grissom, Ph.D.
William S. Stokes, D.V.M., D.A.C.L.A.M. (Director, NICEATM)
Raymond R. Tice, Ph.D.

Occupational Safety and Health Administration (OSHA)
Surrender Ahir, Ph.D.

European Centre for the Validation of Alternative Methods – Liaison
João Barroso, Ph.D.
Thomas Cole, Ph.D.
Chantra Eskes, Ph.D.
Valerie Zuang, Ph.D.

Japanese Center for the Validation of Alternative Methods - Liaison
Hajime Kojima, Ph.D.

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

National Institute of Environmental Health Sciences

William Stokes, D.V.M., D.A.C.L.A.M.
Director; Project Officer

Deborah McCarley
Special Assistant; Assistant Project Officer

NICEATM Support Contract Staff (Integrated Laboratory Systems, Inc.)

David Allen, Ph.D.
Senior Toxicologist/Principal Investigator

Jonathan Hamm, Ph.D.
Senior Toxicologist

Nelson Johnson
Senior Project Coordinator/Technical Writer

Elizabeth Lipscomb, Ph.D.
Staff Toxicologist

Linda Litchfield
Meeting Coordinator/Admin. Asst.

Greg Moyer, M.B.A.
Project Manager

Catherine Sprankle
Senior Communications Specialist

James Truax
Senior Project Coordinator/Technical Writer

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

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

Accidental contact with hazardous chemicals frequently causes eye injury and visual impairment. United States and international regulatory agencies currently use the Draize rabbit eye test (Draize et al. 1944) to identify potential ocular hazards associated with chemicals. The U.S. Consumer Product Safety Commission, U.S. Environmental Protection Agency (EPA), U.S. Food and Drug Administration (FDA), and U.S. Occupational Health and Safety Administration have testing requirements and guidelines for assessing the ocular irritation potential of substances such as pesticides, household products, pharmaceuticals, cosmetics, and agricultural and industrial chemicals.

Although ocular safety assessment has clearly helped to protect consumers and workers, concerns have been raised about the humane aspects of the Draize rabbit eye test. Regulatory authorities have adopted various modifications that reduce the number of animals used and the potential pain and distress associated with the procedure. Significant progress has been made during the last decade. Now only one to three rabbits are required per test, compared to six rabbits in the original protocol. Provisions have been added that allow for animals with severe lesions or discomfort to be humanely euthanized.

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) previously evaluated the validation status of the bovine corneal opacity and permeability (BCOP), isolated chicken eye (ICE), isolated rabbit eye (IRE), and hen's egg test–chorioallantoic membrane (HET-CAM) assays for the identification of severe (irreversible) ocular irritants/corrosives using the EPA, United Nations Globally Harmonized System of Classification and Labeling of Chemicals (GHS), and European Union regulatory hazard classification systems. In ICCVAM's assessment, the performance of the BCOP and ICE assays substantiated their use in testing some substances for regulatory hazard classification. The IRE and HET-CAM assays lacked sufficient performance and/or sufficient data to substantiate their use for regulatory hazard classification.

ICCVAM recommended that the BCOP and ICE should be used in a tiered-testing strategy in which positive substances can be classified as ocular corrosives or severe irritants without animal testing. In accordance with the ICCVAM Authorization Act of 2000 (Public Law 106-545), these recommendations were made available to the public and provided to

U.S. Federal agencies for consideration in the *ICCVAM Test Method Evaluation Report* – In Vitro *Ocular Toxicity Test Methods for Identifying Severe Irritants and Corrosives* (NIH Publication No: 07-4517, available at http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmer.htm). The ICCVAM recommendations were accepted by U.S. Federal agencies, and *in vitro* test methods may now be used instead of the Draize rabbit eye test for certain regulatory testing.

ICCVAM is now reviewing the validation status of these *in vitro* test methods for identification of nonsevere ocular irritants (that is, those that induce reversible ocular damage) and substances not labeled as irritants. Accordingly, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the ICCVAM Ocular Toxicity Working Group (OTWG) prepared draft background review documents (BRDs) that summarize the current validation status of each test method based on published studies and other data and information submitted in response to a June 7, 2007, *Federal Register* request **(**72 FR 31582, available at http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_10966.pdf). The BRDs form the basis for draft ICCVAM test method recommendations, which are provided in separate documents. Liaisons from the European Centre for the Validation of Alternative Methods (ECVAM) and the Japanese Centre for the Validation of Alternative Methods (JaCVAM) will provide input and contribute to the OTWG throughout the evaluation process.

An international independent scientific peer review panel (Panel) will convene in public forum on May 19–21, 2009, to develop conclusions and recommendations on the *in vitro* BCOP, ICE, IRE, and HET-CAM test methods. The Panel includes expert scientists nominated by ECVAM and JaCVAM. We anticipate that these organizations can use the subsequent independent Panel report to deliberate and develop their own test method recommendations. The Panel will consider these BRDs and evaluate the extent to which the available information supports the draft ICCVAM test method recommendations. ICCVAM will consider the conclusions and recommendations of the Panel, along with comments from the public and the Scientific Advisory Committee on Alternative Toxicological Methods, and then finalize the BRD and test method recommendations. These will be forwarded to Federal agencies for their consideration and acceptance decisions where appropriate.

We gratefully acknowledge the organizations and scientists who provided data and information for this document. We also acknowledge the efforts of those individuals contributing to the preparation of this summary review document, including the following staff from the NICEATM Support Contractor, Integrated Laboratory Systems, Inc.: David Allen, Jon Hamm, Nelson Johnson, Elizabeth Lipscomb, Linda Litchfield, Gregory Moyer, Catherine Sprankle, and Jim Truax. We also thank the members of the ICCVAM Ocular Toxicity Working Group, chaired by Karen Hamernik, Ph.D. (EPA) and Jill Merrill, Ph.D. (FDA), and ICCVAM representatives who reviewed and commented on draft versions. We also thank Valerie Zuang, Ph.D., and Dr. Hajime Kojima, Ph.D., the liaisons to the OTWG from ECVAM and JaCVAM, respectively, for their participation.

Marilyn Wind, Ph.D.
Deputy Associate Executive Director
Directorate for Health Sciences
U.S. Consumer Product Safety Commission
Chair, ICCVAM

William S. Stokes, D.V.M., D.A.C.L.A.M.
Rear Admiral, U.S. Public Heath Service
Director, NICEATM
Executive Director, ICCVAM
March 2009

Executive Summary

Background

In October 2003, the U.S. Environmental Protection Agency (EPA) submitted to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) a nomination requesting evaluation of several activities related to reducing, replacing, and refining the use of rabbits in the current *in vivo* eye irritation test method (announced in *Federal Register [FR]* notice 69 FR 13859, March 24, 2004). In response to this nomination, ICCVAM evaluated the validation status of the bovine corneal opacity and permeability (BCOP), Isolated Chicken Eye (ICE), Isolated Rabbit Eye (IRE), and Hen's Egg Test–Chorioallantoic Membrane (HET-CAM) assays. ICCVAM evaluated the test methods' ability to identify severe (irreversible) ocular irritants/corrosives using the EPA, United Nations Globally Harmonized System of Classification and Labeling of Chemicals (GHS), and European Union (EU) regulatory classification systems. ICCVAM considered two of the alternative test methods, the BCOP assay and (ICE) assay, to have sufficient performance to substantiate their use for regulatory hazard classification testing of some types of substances. The IRE and HET-CAM assays lacked sufficient performance and/or sufficient data to substantiate their use for regulatory hazard classification. ICCVAM subsequently recommended that the BCOP and ICE should be used in a tiered-testing strategy, where positive substances can be classified as ocular corrosives or severe irritants without the need for animal testing. These recommendations were forwarded to U.S. Federal agencies for consideration, and as a result, *in vitro* test methods may now be used instead of conventional tests for certain regulatory testing purposes.

ICCVAM is now reviewing the validation status of these *in vitro* test methods for identifying nonsevere ocular irritants (i.e., those that induce reversible ocular damage) and and substances not labeled as irritants (i.e., EPA Category IV, EU Not Labeled, GHS Not Classified). Accordingly, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), in conjunction with an ICCVAM Ocular Toxicity Working Group (OTWG) prepared draft background review documents (BRDs) that summarize the available data and information regarding the validity

(usefulness and limitations) of each test method. This BRD summarizes the available information for the ICE test method.

ICE Test Method Protocol

The ICE test method is an organotypic model that provides short-term maintenance of the chicken eye *in vitro*. In this test method, damage by the test substance is assessed by determination of corneal swelling, opacity, and fluorescein retention. While the latter two parameters involve a qualitative assessment, analysis of corneal swelling provides for a quantitative assessment. 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* ocular irritancy classification. Either of these outcomes can then be used to predict the *in vivo* ocular irritation potential of a test substance.

Validation Database

No new ICE data were obtained since the ICCVAM evaluation of ICE for identifying ocular corrosives and severe irritants (ICCVAM 2006a). Therefore, the same database (n = 175 substances) was used in the current evaluation (i.e., Prinsen and Koëter [1993], Balls et al. [1995], Prinsen [1996], Prinsen [2000] and Prinsen [2005]). The most common chemical classes tested in ICE are alcohols (n = 15), carboxylic acids (n = 12), esters (n = 10) and heterocyclics (n = 9). Of the 175 total substances 48% (85/175, including formulations of unidentified composition) could not be assigned a specific chemical class. The most common product classes tested in ICE are solvents (n = 37), soaps/surfactants (n = 34), industrial chemicals (n = 20), and pesticides/herbicides (n = 15). Of the 175 total substances, 13% (23/175) could not be assigned a product class.

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.

ICE Test Method Accuracy

Identification of All Ocular Hazard Categories

The ability of the ICE test method to identify all categories of ocular irritation potential, as defined by the GHS, EPA, and EU classification systems (EPA 1996; EU 2001; UN 2003), was evaluated. This analysis was also performed with specific chemical classes and/or physical properties excluded based on them previously being identified as discordant in ICE (ICCVAM 2006a). In order to verify that these were also the most discordant types of substances when all hazard categories are evaluated, separate analyses were also conducted for all chemical classes and specific physical properties of interest (e.g., physical form, surfactants) represented by at least five substances. The results indicate that alcohols, solids, and surfactants continue to be most problematic.

As indicated in **Table 1**, the overall correct classification for the ICE test method ranged from 59% to 77%, depending on the classification system used, and these ranges improved slightly when "discordant classes" noted in the ICCVAM BRD (2006a) were removed from the database (overall correct classification increased to 64% to 80%). In either case, the best performance was achieved when using the EU hazard classification system (EU 2001) and poorest performance was achieved when using the GHS hazard classification system (UN 2007). This trend is also apparent when evaluating the correct classification for the corrosive/severe irritant, moderate irritant, and not labeled categories (i.e., actual correct classifications ranged from 48% to 59%, 36% to 57%, and 66% to 89% for the corrosive/severe irritant, moderate irritant, and not labeled categories, respectively for the entire database; and 63% to 67%, 46% to 65%, and 68% to 87% for the corrosive/severe irritant, moderate irritant, and not labeled categories, respectively when discordant classes are removed).

However, while the EU hazard classification system does not discriminate between mild and moderate irritants, correct classification of mild irritants was greater when using the GHS classification rather than the EPA hazard classification system (i.e., correct classification of mild irritants was 53% and 73% for the GHS and EPA systems, respectively for the entire database, and 67% for either system when discordant classes are removed).

Table 1 Evaluation of the Performance of the ICE Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA, EU, or GHS Classification Systems

Data Source	Overall Correct Classification	Severe[2]		Moderate[3]			Mild[4]			Nonirritant[5]	
		actual	under	over	actual	under	over	actual	under	over	actual
Overall (EPA)	62% (87/140)	48% (13/27)	52% (14/27)	31% (5/16)	50% (8/16)	19% (3/16)	29% (11/38)	53% (20/38)	18% (7/38)	22% (13/59)	78% (46/59)
w/o Alcohols, Surfactants, and Solids[1] (EPA)	67% (52/78)	67% (6/9)	33% (3/9)	20% (2/10)	60% (6/10)	20% (2/10)	17% (1/6)	67% (4/6)	17% (1/6)	21% (8/39)	79% (31/39)
Overall (EU)	77% (118/153)	59% (19/32)	41% (13/32)	18% (5/28)	57% (16/28)	25% (7/28)	NA	NA	NA	11% (10/93)	89% (83/93)
w/o Alcohols, Surfactants, and Solids[1] (EU)	80% (66/82)	67% (6/9)	33% (3/9)	18% (3/17)	65% (11/17)	18% (3/17)	NA	NA	NA	13% (7/56)	87% (49/56)
Overall (GHS)	59% (83/141)	52% (15/29)	48% (14/29)	36% (8/22)	36% (8/22)	28% (6/22)	18% (2/11)	73% (8/11)	9% (1/11)	34% (27/79)	66% (52/79)
w/o Alcohols, Surfactants, and Solids[1] (GHS)	64% (49/77)	63% (5/8)	37% (3/8)	23% (3/13)	46% (6/13)	31% (4/13)	17% (1/6)	67% (4/6)	17% (1/6)	32% (16/50)	68% (34/50)

Abbreviations: EPA = Environmental Protection Agency Hazard Classification System (EPA 1998); EU = European Union Hazard Classification System (EU 2001); GHS = United Nations Globally Harmonized System for Classification and Labelling of Chemicals (UN 2007); ICE = Isolated Chicken Eye

[1]Physical properties and chemical classes highlighted in ICCVAM (2006a) as "discordant classes" based on high false positive or false negative rates when using ICE to identify ocular corrosives/severe irritants.

Distinguishing Substances Not Labeled as Irritants from All Other Hazard Categories

The ability of the ICE test method to distinguish substances not labeled as irritants (i.e., EPA Category IV, EU Not Labeled, GHS Not Classified) from all other ocular hazard categories (i.e., EPA Category I, II, or III; EU R41 or R36; GHS Category 1, 2A, or 2B) was also evaluated. Again, this same analysis was performed with specific chemical classes and/or physical properties excluded based on them previously being identified as discordant in ICE (ICCVAM 2006a).

As indicated in **Table 2**, overall accuracy ranged from 78% to 85% depending on the hazard classification system used. However, in contrast to the evaluation of the ICE for identifying all hazard categories, when the discordant classes were removed overall accuracy was either unaffected (remained 85% for the EU system), or slightly reduced (from 83% to 82% and from 78% to 75% for the EPA and GHS systems, respectively). Similarly, the false positive and false negative rates ranged from approximately 11% to 34% and 6% to 22%, respectively whether or not discordant classes were included in the evaluation. The lowest false negative rate (6% [4/62]) was noted for the GHS system, followed by 14% (11/81) for the EPA system, and 22% (13/60) for the EU system. However, among these false negatives, at least one substance is classified as an ocular corrosive/severe irritant based on Draize data (n = 1 each for the EPA and GHS systems, and n = 6 for the EU system).

Conversely, the lowest false positive rate (11% [10/93]) was noted for the EU system, followed by 22% (13/59) for the EPA system, and 34% (27/79) for the GHS system. One of the false positives for the EU system was classified as an ocular corrosive/severe irritant based on ICE results; this did not occur for the EPA or GHS systems.

633

634 **Table 2 Accuracy of the ICE Test Method for Distinguishing Substances Not Labeled as Irritants from All Other**
635 **Irritant Classes as Defined by the EPA, EU, and GHS Classification Systems**

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall (EPA)	140	83	116/140	86	70/81	78	46/59	22	13/59	14	11/81
w/o Alcohols, Surfactants and Solids (EPA)	78	82	69/78	85	33/39	79	31/39	21	8/39	15	6/39
Overall (EU)	153	85	130/153	78	47/60	89	83/93	11	10/93	22	13/60
w/o Alcohols, Surfactants and Solids (EU)	82	85	70/82	81	51/26	88	49/56	12	7/56	19	5/26
Overall (GHS)	141	78	110/141	94	58/62	66	52/79	34	27/79	6	4/62
w/o Alcohols, Surfactants and Solids (GHS)	77	75	58/77	89	24/27	68	34/50	32	16/50	11	3/27

636 Abbreviations: EPA = Environmental Protection Agency Hazard Classification System (EPA 1998); EU = European Union Hazard Classification System (EU
637 2001); GHS = United Nations Globally Harmonized System for Classification and Labelling of Chemicals (UN 2007); ICE = Isolated Chicken Eye
638 [1]Physical properties and chemical classes highlighted in ICCVAM (2006) as "discordant classes" based on high false positive or false negative rates when using
639 ICE to identify ocular corrosives/severe irritants.

ICE Test Method Reliability

Quantitative and qualitative evaluations of ICE test method reliability have been conducted previously (ICCVAM 2006a). Since the database used for the current evaluation of the ICE test method has not changed, the quantitative evaluation of test method reliability remains unchanged.

Interlaboratory Reproducibility

However, additional qualitative analyses of interlaboratory reproducibility were conducted to evaluate the extent of agreement of ICE hazard classifications among the four participating laboratories from the interlaboratory validation study (Balls et al. 1995). As was done for the accuracy evaluation, these qualitative evaluations of reproducibility were conducted based on 1) the use of the ICE test method for identifying all ocular hazard categories according to the EPA, EU, or GHS systems, and 2) the use of the ICE test method to distinguish substances not labeled as irritants (i.e., EPA Category IV, EU Not Labeled, GHS Not Classified) from all other ocular hazard categories (i.e., EPA Category I, II, or III; EU R41 or R36; GHS Category 1, 2A, or 2B).

Using the first approach (i.e., identifying all ocular hazard categories) there was 100% agreement among the four laboratories for most of the Draize ocular corrosives/severe irritants based on all three classification systems, whether they were correctly identified or underclassified by the ICE test method (e.g., for the EPA system, there was 100% agreement for 70% [7/10] of the correctly identified Category I substances and for 78% [7/9] of the Category I substances underclassified by ICE). There was also 100% agreement among the four laboratories for at least 50% of the correctly identified moderate ocular irritants (i.e., EPA Category II, GHS Category 2A, or EU R36). However, for the mild ocular irritants (i.e., EPA Category III, GHS Category 2B), there was 100% agreement for 0% (0/2) to 13% (1/8) of the correctly identified substances, with only 50% agreement among the four laboratories for 50% (4/8 or 1/2 for the EPA or GHS system, respectively) of these substances. For the Balls et al. (1995) database, most of the Draize not labeled substances (based on the EPA [Category IV] or GHS [Not Classified] systems) were overclassified by ICE, and there was at least 75% agreement among the four laboratories for all but two of these substances (e.g., there was at least 75% agreement for 85% [11/13] of the GHS Not Labeled substances that

were overclassified by ICE). For the EU system, there was at least 75% agreement among the four laboratories for 76% (13/17) of the Not Labeled substances whether they were correctly identified or overclassified by the ICE test method (e.g., there was at least 75% agreement for 75% [6/8] of the EU Not Labeled substances that were correctly identified by ICE, and 78% [7/9] overclassified by ICE).

Using the second approach (i.e., distinguishing substances not labeled as irritants from all other ocular hazard categories), there was 100% agreement among the four laboratories for 61% (36/59) to 75% (44/59) of the substances included in the Balls et al. (1995) database for the EU, or EPA and GHS classification systems, respectively.

There was 100% agreement among the four laboratories for 81% (38/47) of the substances correctly identified as an irritant according to the EPA system (i.e., Category I, II, or III). While none of the EPA Category IV substances was correctly identified by ICE, there was 75% agreement among the four laboratories for both of the Category IV substances that were overpredicted by ICE. There was at least 75% agreement among the four laboratories for 77% (7/9) of the substances correctly identified as Category IV.

There was 100% agreement among the four laboratories for 85% (22/26) of the substances correctly identified as an irritant according to the EU system (i.e., R36 or R41). There was at least 75% agreement among the four laboratories for 77% (7/9) of the substances correctly identified as Not Labeled.

There was 100% agreement among the four laboratories for 87% (33/38) of the substances correctly identified as an irritant according to the GHS system (i.e., Category 1, 2A, or 2B). While only one of the GHS Not Labeled substances was correctly identified by ICE (for which there was 75% agreement among the laboratories), there was at least 75% agreement among the four laboratories for 85% (11/13) of the Not Labeled substances that were overpredicted by ICE.

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

1.1 Background

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 not labeled as irritant responses are observed (UN 2003).

In 2006, ICCVAM completed an evaluation of the isolated chicken eye (ICE) test method for its ability to identify ocular corrosives and severe irritants (ICCVAM 2006a). Following this review, ICCVAM concluded that the ICE test method could be used, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) (ICCVAM 2006b). While it was not considered valid as a complete replacement for the *in vivo* rabbit eye test, the ICE test method was recommended for use as part of a tiered testing strategy for regulatory classification and labeling within a specific applicability domain. Accordingly, substances that are positive in this assay can be classified as ocular corrosives or severe irritants without further testing in rabbits, while a substance that tests negative would need to be tested in rabbits using a sequential testing strategy, as outlined in OECD Test Guideline 405 (OECD 2002).

ICCVAM is now conducting an evaluation to further characterize the usefulness and limitations of the ICE test method for identifying non-severe irritants and substances not labeled as irritants. As part of this evaluation process, this Background Review Document (BRD) has been prepared to describe the current validation status of the ICE test method, including what is known about its reliability and accuracy, its applicability domain, the

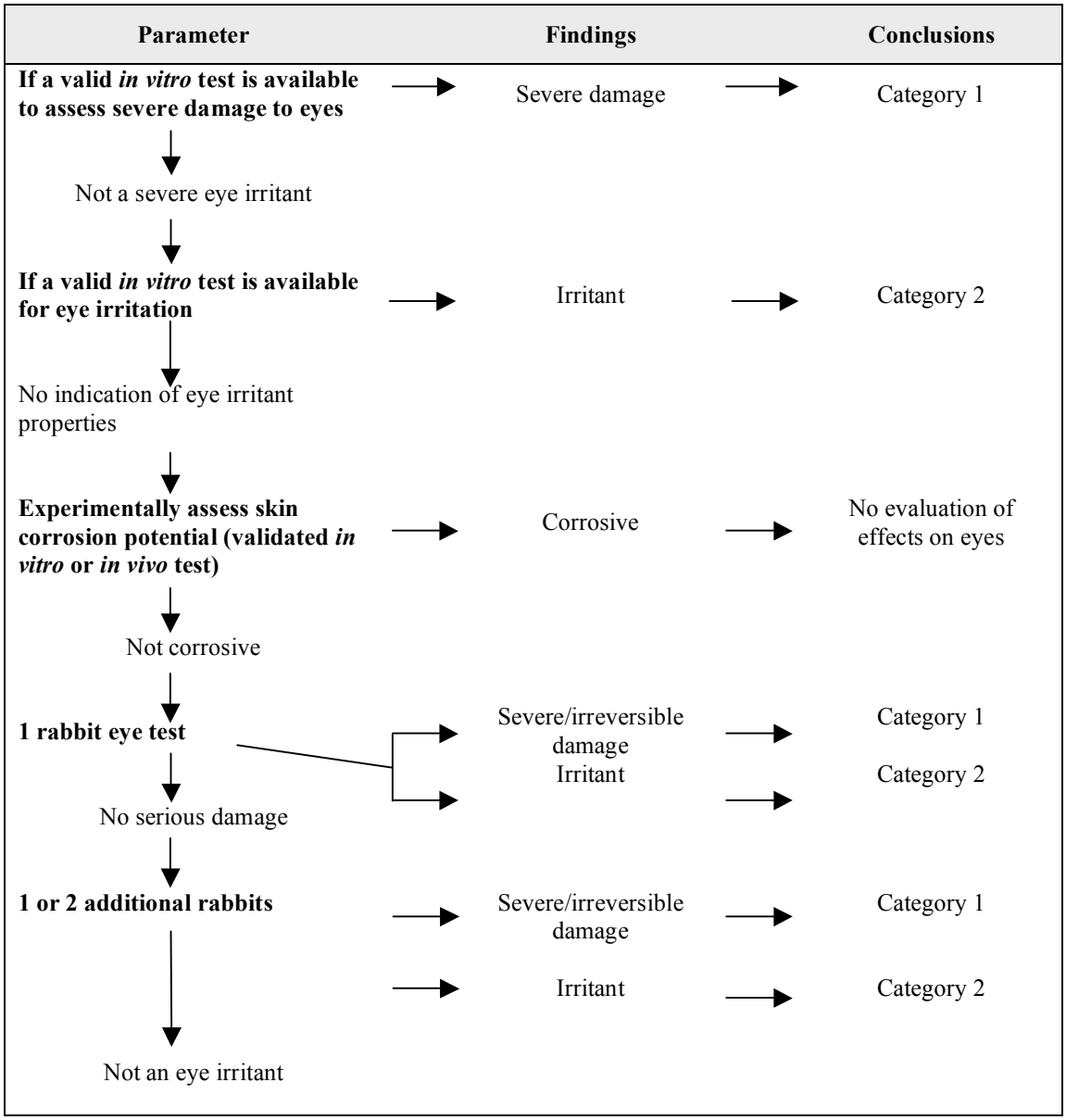
numbers and types of substances tested, and the availability of a standardized protocol. This BRD was prepared for use by ICCVAM expert panel review of ICE as a method to identify all categories of ocular irritants and substances not labeled as irritants. Parallel reviews of the IRE, HET-CAM, and BCOP test methods are being conducted. Results of the Expert Panel Report, combined with the analyses presented in the BRDs, will be 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.

For a more detailed discussion of the background of the ICE test method, including its scientific basis and regulatory rationale and applicability, see the ICCVAM BRD, *Current Status of* In Vitro *Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method* (ICCVAM 2006a).

1.2 Use of the ICE Test Method in Overall Strategy of Hazard or Safety Assessment

As shown in **Figure 1-1**, the GHS allows for use of validated and accepted *in vitro* methods to identify severe ocular irritants/corrosives and ocular irritants without further testing. The ICE test method is currently recommended for use in identifying ocular corrosives and severe irritants in a tiered-testing strategy for regulatory classification and labeling (e.g., GHS, UN 2003). As indicated above, ICCVAM is now conducting an evaluation to further characterize the usefulness and limitations of the ICE test method for identifying nonsevere irritants and substances not labeled as irritants.

752 **Figure 1-1 GHS Testing Strategy for Serious Eye Damage and Eye Irritation[1]**



753 Abbreviation: GHS = United Nations Globally Harmonized System for Classification and Labelling of
754 Chemicals

755 [1]Adapted from UN (2003).

756

757 ## 1.3 Validation of the ICE Test Method

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

Validation is the process by which the reliability and relevance of an assay for a specific purpose are established (ICCVAM 1997). Relevance is defined as the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 1997). For the ICE test method described in this BRD, relevance is restricted to how well the assay identifies substances that are capable of producing effects to the eye across all hazard categories. 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 available information on the ICE test method. Where adequate data are available, the qualitative and quantitative performances of the assay are evaluated.

1.4 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 as detailed in the *Background Review Document, Current Status of* In Vitro *Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method* (ICCVAM 2006a). A subsequent literature search conducted in January 2009 revealed no new articles containing results from an ICE test method. Therefore, the database used in this analysis is to the same as the database used in ICCVAM (2006a).

2.0 ICE Test Method Protocol Components

The ICE test method is an organotypic model that provides short-term maintenance of the chicken eye *in vitro*. In this test method, damage by the test substance is assessed by determination of corneal swelling, opacity, and fluorescein retention. While the latter two parameters involve a qualitative assessment, analysis of corneal swelling provides for a quantitative assessment. 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* ocular irritancy classification. Either of these outcomes can then be used to predict the *in vivo* ocular irritation potential of a test substance.

For a detailed description of how the ICE test method is conducted, see ICCVAM (2006a). Briefly, 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 and eyes dissected within two hours after death. After dissection, the eyes placed in a superfusion apparatus, where isotonic saline is applied to the cornea, at a rate of two to three drops per minute, through a steel tube attached to a peristaltic pump. Substances are applied as a single dose (0.03 mL for liquids, 0.03 g for solids) for 10 seconds. Corneal reactions are measured at regular intervals up to four hours post-treatment, while fluorescein retention is evaluated at 30 minutes post-treatment only. Mean values for each parameter (corneal swelling, corneal opacity, and fluorescein retention) are determined and the maximum mean values of these measurements are used for hazard classification purposes using established decision criteria.

This approach entails calculation of mean corneal opacity, corneal swelling, and fluorescein retention scores at each time point for each test substance and relating the maximum scores for each endpoint to one of four irritancy categories as follows (see **Tables 2-1**, **2-2**, and **2-3**):

Table 2-1 Categorization of Corneal Thickness Measurements

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

Table 2-2 Categorization of Corneal Opacity Scores

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

Table 2-3 Categorization of Fluorescein Retention Scores

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 were then combined into an overall *in vitro* ocular irritancy classification for comparison to the *in vivo* ocular irritancy classification according to the following scheme (**Table 2-4**, INVITTOX 1994).

Table 2-4 *In Vitro* Ocular Irritancy Classification Scheme for the ICE Test Method

Classification	Combinations of the 3 Endpoints
Nonirritant	3 x I
	2 x I, 1 x II
Mild Irritant	3 x II
	2 x II, 1 x I
	2 x II, 1 x III
	2 x I, 1 x IV[1]
	1 x I, 1 x II, 1 x III[1]
Moderate Irritant	3 x III
	2 x III, 1 x II
	2 x III, 1 x IV
	2 x III, 1 x I[1]
	2 x II, 1 x IV[1]
	I x II, 1 x III, 1 x IV[1]
Severe Irritant	3 x IV
	2 x IV, 1 x III
	2 x IV, 1 x II[1]
	2 x IV, 1 x I[1]

[1]Combinations less likely to occur.
For the purposes of this evaluation, Nonirritant = EPA Category IV, EU Not Labeled, or GHS Not Classified; Mild Irritant = EPA Category III, GHS Category 2B; Moderate Irritant = EPA Category II, GHS Category 2A; Severe Irritant = EPA Category I, EU Category R41, GHS Category 1. The Mild Irritant and Moderate Irritant categories were combined to generate EU Category R36.

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, see **Table 2-4**).

3.0 Substances Used for Validation of the ICE Test Method

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 not labeled) also should be assessed to determine limits to the range of responses that can be evaluated by the *in vitro* test method.

As noted in **Section 1.5**, no new ICE data were obtained since the ICCVAM evaluation of ICE for identifying ocular corrosives and severe irritants (ICCVAM 2006a). Therefore, the same database was used in the current evaluation (i.e., Prinsen and Koëter [1993], Balls et al. [1995], Prinsen [1996], Prinsen [2000] and Prinsen [2005]).

Tables 3-1 and **Table 3-2** show the chemical classes and product classes for the test substances included in the database used in this assessment. Information, including substance name, Chemical Abstracts Service Registry Number (CASRN), chemical and/or product class, concentration(s) tested, purity, supplier or source, and literature reference using the test substance are provided in **Appendix A**. 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.

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

Abbreviation: ICE = isolated chicken eye

As shown in **Table 3-1**, the most common chemical classes tested in the ICE test method are alcohols (n = 15), carboxylic acids (n = 12), esters (n = 10) and heterocyclics (n = 9). Of the 175 substances included in **Appendix A**, 85 (including formulations 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 solvents (n = 37), soaps/surfactants (n = 34), industrial chemicals (n = 20), and pesticides/herbicides (n = 15). Of the 175 substances included in **Appendix A**, 23 could not be assigned a product class.

881 **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

882

883

4.0 *In Vivo* Reference Data Used for an Assessment of ICE Test Method Accuracy

A detailed description of the test method protocol used to generate the *in vivo* reference data (i.e., the Draize rabbit eye test) is provided in ICCVAM (2006a). There also are a number of national and international test guidelines that describe this procedure (EPA 1998, OECD 2002, CPSC 2003, EU 2004). The subjective scoring system used for assigning an ocular hazard classification is based on a discrete scale for grading the severity of ocular lesions on the cornea, iris, and conjunctiva.

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. These data were used by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to assign an ocular hazard classification according to the EPA (1996), the EU (2007), and the GHS (UN 2007) ocular irritancy classification systems (**Appendix C**). Exceptions included the following:

- For Prinsen (2000), 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 Prinsen (1996), 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 2007) classification systems for most test substances. However, for some test substances, adequate information was provided such that they could be included in the evaluation.
- For Prinsen and Koëter (1993), 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 2007) 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.

4.1 *In Vivo* Classification Criteria Used for BRD Analysis

As described in ICCVAM (2006a), 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, 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. Scoring continues 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. The rules used for classification according to the EPA, EU, or GHS classification systems are detailed in ICCVAM (2006a).

4.2 *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 Good Laboratory Practice (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 reports. Based on the available information, all of the reports included *in vivo* data obtained according to GLP guidelines.

5.0 ICE Test Method Data and Results

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 for an accuracy analysis of the ICE test method for the identification of all categories of ocular irritation. As detailed in **Section 6.0**, these data were evaluated collectively (i.e., data from all studies combined), and on a per study basis[1].

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

Original study records, containing data for the substances screened with the ICE test method in Prinsen (1996), Prinsen (2000), and Prinsen (2005), were kindly provided by Mr. Menk Prinsen of TNO (TNO Nutrition and Food Research, Toxicology and Applied Pharmacology, Zeist, The Netherlands). Summaries of ICE results (i.e., total scores) but not original data were obtained for the 60 substances evaluated by Balls et al. (1995). No other ICE test method data have been obtained by NICEATM.

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

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., modified maximum average score [MMAS]). However, because the current evaluation is focused on 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, an *in vitro* classification system was used to assign an ocular irritation classification for each test substance (see **Section 2.0**).

5.3 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

[1] Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

test method endpoint values (maximum mean), *in vitro* classification, and reference. 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 MeSH system (available at http://www.nlm.nih.gov/mesh). **Appendix A** provides information on the names, synonyms, CASRN, and chemical/product class, where available, for each substance while **Appendix B** contains the *in vitro* ICE test method data sorted by reference and alphabetically by substance name.

5.4 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 in ICCVAM 2006a).

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 identify all categories of ocular irritation potential, as defined by the GHS, EPA, and EU classification systems (EPA 1996; EU 2001; UN 2007), was evaluated. This same analysis was also performed with specific chemical classes and/or physical properties excluded based an them previously being identified as discordant in ICE (ICCVAM 2006a).

These evaluations were conducted on the overall data set comprised by combining results from the reports indicated in **Section 5.0** then assigning an overall ocular irritancy

classification for each substance (**Appendix B** and **C**). When the same substance was evaluated in multiple laboratories, an overall ICE classification was based on the majority classification among all of the studies. When there was an equal number of differing irritancy classifications for substances (e.g., two tests classified a substance as a not labeled and two tests classified a substance as a mild irritant), the more severe irritancy classification was used for the overall classification for the substance (mild irritant, in this case).

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 174 substances, 141 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the GHS classification system (UN 2007) (see **Appendix C**). Based on results from *in vivo* rabbit eye experiments, 20% (29/141)[2] were classified as Category 1, 16% (22/141[3]) were classified as Category 2A, 8% (11/141) were classified as Category 2B, and 56% (79/141) were classified as Not Labeled. The remaining 19% (33/174) of the substances could not be classified according to the GHS classification system due to the lack of adequate animal data and are so noted in **Appendix C**.

6.1.1.1 Identification of Category 1 Substances (Ocular Corrosives/Severe Irritants)

The ICE test method correctly identified 52% (15/29) of the Category 1 substances (**Table 6-1**). Among the remaining 48% (14/29) Category 1 substances that were underpredicted by ICE, 10% (3/29) were classified as Category 2A, 35% (10/29) were classified as Category 2B, and 3% (1/29) was classified as Not Classified.

[2] 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. 1% sodium hydroxide was duplicated in the database. Sodium hydroxide (Prinsen and Koëter, 1993) was removed because the *in vivo* classification corresponded to a 10% solution.
[3] Triton X-100 (10%) and dibenzyl phosphate were excluded because they were classified *in vitro* as 2A/2B.

Table 6-1 Evaluation of the Performance of the ICE Test Method in Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System[1], by Study and Overall

Data Source	Overall Correct Classification	Category 1		Category 2A			Category 2B			Not Classified	
		actual	under	over	actual	under	over	actual	under	over	actual
Prinsen and Koëter (1993)	63% (5/8)	100% (2/2)	0% (0/2)	100% (1/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	100% (1/1)	25% (1/4)	75% (3/4)
Balls et al. (1995)	38% (19/50)	55% (11/20)	45% (9/20)	46% (6/13)	38% (5/13)	16% (2/13)	50% (2/4)	50% (2/4)	0% (0/4)	92% (12/13)	8% (1/13)
Prinsen (1996)	81% (29/36)	50% (1/2)	50% (1/2)	0% (0/3)	33% (1/3)	67% (2/3)	0% (0/2)	100% (2/2)	0% (0/2)	14% (4/29)	86% (25/29)
Prinsen (2005)	63% (29/46)	0% (0/4)	100% (4/4)	20% (1/5)	40% (2/5)	40% (2/5)	0% (0/4)	100% (4/4)	0% (0/4)	30% (10/33)	70% (23/33)
Overall[2]	59% (83/141)	52% (15/29)	48% (14/29)	36% (8/22)	36% (8/22)	28% (6/22)	18% (2/11)	73% (8/11)	9% (1/11)	34% (27/79)	66% (52/79)

Abbreviations: GHS = Globally Harmonized System; ICE = Isolated Chicken Eye;
[1]GHS classification system (UN 2007)
[2]Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

6.1.1.2 Identification of Category 2A Substances (Moderate Ocular Irritants)

For the 22 substances that could be evaluated, the ICE test method correctly identified 36% (8/22) as moderate irritants while 36% (8/22) were overpredicted and 28% (6/22) were underpredicted (**Table 6-1**).

6.1.1.3 Identification of Category 2B Substances (Mild Ocular Irritants)

For the 11 substances that could be evaluated, the ICE test method correctly identified 73% (8/11) as mild irritants while 18% (2/22) were overpredicted and 9% (1/11) were underpredicted (**Table 6-1**).

6.1.1.4 Identification of Not Classified Substances

For the 79 substances that could be evaluated, the ICE test method correctly identified 66% (52/79) as substances not labeled as irritants while 34% (27/79) were overpredicted (**Table 6-1**).

6.1.1.5 Ability to Distinguish Not Classified Substances from All Other Classes

In addition to evaluating the ability of the ICE test method to identify each individual ocular hazard category according to the GHS classification system, ICCVAM also evaluated the ability of the ICE test method to distinguish Not Classified substances from all irritant classes[4]. Using this approach for the 141 substances, the ICE test method has an overall accuracy of 78% (110/141), a sensitivity of 94% (58/62), a specificity of 66% (52/79), a false positive rate of 34% (27/79), and a false negative rate of 6% (4/62) (**Table 6-2**). One (25%) of the 4 false positive substances (4-carboxybenzaldehyde) was from one of the discordant classes (solids).

As detailed below, the results from each individual study were also evaluated separately.

Prinsen and Koëter (1993): Based upon the *in vivo* rabbit data, eight substances could be assigned a GHS classification. Based on these eight substances, the ICE test method has an accuracy of 75% (6/8), sensitivity of 75% (3/4), specificity of 75% (3/4), false positive rate of 25% (1/4), and a false negative rate of 25% (1/4) (**Table 6-2**).

[4] ICCVAM (2006) provides an evaluation of the ICE test method for distinguishing ocular corrosives and severe irritants from all other classes. Since the database of ICE test method results has not changed, this analysis has not been repeated here.

Table 6-2 Accuracy of the ICE Test Method for Distinguishing Not Classified from All Other Irritant Classes as Defined by the GHS Classification System[1], by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	8	75	6/8	75	3/4	75	3/4	25	1/4	25	1/4
Balls et al. (1995)	50	72	36/50	95	35/37	8	1/13	92	12/13	5	2/37
Prinsen (1996)	36	89	32/36	100	7/7	86	25/29	14	4/29	0	0/7
Prinsen (2005)	46	76	35/46	92	12/13	70	23/33	30	10/33	8	1/13
Overall[4]	141	78	110/141	94	58/62	66	52/79	34	27/79	6	4/62

[1]GHS = United Nations Globally Harmonized System for Classification and Labelling of Chemicals (UN 2007). NC vs. Cat 1/2A/2B.
[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]Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

Balls et al. (1995): Based upon the *in vivo* rabbit data, 50 substances could be assigned a GHS classification. Based on these 50 substances, the ICE test method has an accuracy of 72% (36/50), sensitivity of 95% (35/37), specificity of 8% (1/13), false positive rate of 92% (12/13), and a false negative rate of 5% (2/37) (**Table 6-2**). One of the two false negative substances (4-carboxybenzaldehyde) was from one of the discordant classes (solids).

Prinsen (1996): Based upon the *in vivo* rabbit data, 36 substances could be assigned a GHS classification. Based on these 36 substances, the ICE test method has an accuracy of 89% (32/36), sensitivity of 100% (7/7), specificity of 86% (25/29), false positive rate of 14% (4/29), and a false negative rate of 0% (0/7) (**Table 6-2**).

Prinsen (2005): Based upon the *in vivo* rabbit data, 46 substances could be assigned a GHS classification. Based on these 46 substances, the ICE test method has an accuracy of 76% (35/46), sensitivity of 92% (12/13), specificity of 70% (22/33), false positive rate of 30% (10/33), and a false negative rate of 8% (1/13) (**Table 6-2**).

6.1.1.6 Performance of the ICE Test Method with Discordant Classes Excluded

The previously identified limitations for the ICE test method are based upon the false positive rate for alcohols and the false negative rates for solids and surfactants when the ICE is used to identify ocular corrosives and severe irritants (ICCVAM 2006a). For this reason, the performance of the ICE test method for identifying all ocular irritant classes was evaluated with these substances excluded from the database. The overall performance statistics were slightly improved (e.g., overall correct classification increased from 59% to 64%) when these substances were excluded (**Table 6-3**).

When the ability of the ICE test method to distinguish Not Classified substances from all irritant classes was evaluated with the discordant classes removed, overall accuracy of the ICE method was actually slightly reduced from 78% (110/141) to 75% (58/77), false negative rates increased from 6% (4/62) to 11% (3/27), and false positive rates decreased from 34% (27/79) to 32% (16/50) (**Table 6-4**).

Table 6-3 Evaluation of the Performance of the ICE Test Method in Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

ICE	Overall Correct Classification	Category 1		Category 2A			Category 2B			Not Classified	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Overall	59% (83/141)	52% (15/29)	48% (14/29)	36% (8/22)	36% (8/22)	28% (6/22)	18% (2/11)	73% (8/11)	9% (1/11)	34% (27/79)	66% (52/79)
w/o Alcohols	62% (80/130)	52% (14/27)	48% (13/27)	19% (3/16)	44% (7/16)	38% (6/16)	10% (1/10)	80% (8/10)	10% (1/10)	34% (26/77)	66% (51/77)
w/o Surfactants	61% (74/121)	52% (11/21)	48% (10/21)	40% (8/20)	35% (7/20)	25% (5/20)	20% (2/10)	70% (7/10)	10% (1/10)	30% (21/70)	70% (49/70)
w/o Solids	57% (57/107)	59% (10/17)	41% (7/17)	38% (8/21)	38% (8/21)	24% 5/21)	25% (2/8)	63% (5/8)	12% (1/8)	38% (23/61)	62% (38/61)
w/o Alcohols and Surfactants	64% (70/110)	53% (10/19)	47% (9/19)	21% (3/14)	43% (6/14)	36% (5/14)	11% (1/9)	78% (7/9)	11% (1/9)	29% (20/68)	71% (48/68)
w/o Alcohols, Surfactants, and Solids	64% (49/77)	63% (5/8)	37% (3/8)	23% (3/13)	46% (6/13)	31% (4/13)	17% (1/6)	67% (4/6)	17% (1/6)	32% (16/50)	68% (34/50)

Abbreviations: GHS = United Nations Globally Harmonized System for Classification and Labelling of Chemicals; ICE = Isolated Chicken Eye

[1]GHS classification system (UN 2007).

Table 6-4 Accuracy of the ICE Test Method for Distinguishing Not Classified from All Other Irritant Classes as Defined by the GHS Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

ICE	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall	141	78	110/141	94	58/62	66	52/79	34	27/79	6	4/62
w/o Alcohols	129	78	100/129	92	49/53	67	51/76	33	25/76	8	4/53
w/o Surfactants	122	79	96/122	92	47/51	69	49/71	31	22/71	8	4/51
w/o Solids	107	76	81/107	93	43/46	62	38/61	38	23/61	7	3/46
w/o Alcohols and Surfactants	109	78	85/109	90	37/41	71	48/68	29	20/68	10	4/41
w/o Alcohols, Surfactants and Solids	77	75	58/77	89	24/27	68	34/50	32	16/50	11	3/27

[1]GHS = United Nations Globally Harmonized System for Classification and Labelling of Chemicals (UN 2007). NC vs. Cat 1/2A/2B.
[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.

Further analysis of substances according to chemical class for which hazard classification was underpredicted by ICE indicated that carboxylic acids had the highest proportion of underpredicted substances (19% [4/21]). Regarding the physical form of underpredicted substances, 12 were liquids and 8 were solids. Six surfactants were underpredicted by ICE (**Table 6-5**).

According to the GHS classification system, the most overpredicted substances (false positives) were alcohols, which accounted for 24% (9/37) of the overpredicted substances. Regarding the properties of interest for these substances among the overpredicted substances, 73% (27/37) were liquids, 4 were solids, and six were surfactants (**Table 6-5**).

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 174 substances, 140 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the EPA classification system (EPA 1996) (see **Appendix C**). Based on results from *in vivo* rabbit eye experiments, 19% (27/140)[5] were classified as Category 1, 11% (16/140)[6] were classified as Category II, 27% (38/140) were classified as Category III, and 42% (59/140) were classified as Category IV. The remaining 20% (34/174) of substances could not be classified according to the EPA classification system due to the lack of adequate animal data and are so noted in **Appendix C**.

6.1.2.1 Identification of Category I Substances (Ocular Corrosives/Severe Irritants)

The ICE test method correctly identified 52% (13/27) of the Category I substances (**Table 6-6**). Among the remaining 52% (14/27) Category I substances that were underpredicted by ICE, 11% (3/27) were classified as Category II, 37% (10/27) were classified as Category III, and 4% (1/27) was classified as Category IV.

[5] 1% sodium hydroxide was duplicated in the database. Sodium hydroxide (Prinsen and Koëter, 1993) was removed because the in vivo classification corresponded to a 10% solution.
[6] Triton X-100 (10%) and dibenzyl phosphate were removed because they were classified as II/III.

Table 6-5 Evaluation of Under- and Overprediction of the ICE Test Method Using the GHS[1] Classification System In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method by Chemical Class or Physical Property

Category	N	Underprediction (*In Vivo/In Vitro*)						Overprediction (*In Vivo/In Vitro*)					
		Category 1			Category 2A		Cat2B	Cat 2A	Category 2B		Not Classified		
		NI	2B	2A	NI	2B	NI	1	2A	1	2B	2A	1
Overall	141	3% (1/29)	34% (10/29)	10% (3/29)	9% (2/22)	18% (4/22)	9% (1/11)	36% (8/22)	18% (2/11)	0% (0/11)	27% (21/79)	8% (6/79)	0% (0/79)
Chemical Class[2]													
Alcohol	12	0% (0/2)	50% (1/2)	0% (0/2)	0% (0/6)	0% (0/6)	-	83% (5/6)	100% (1/1)	-	67% (2/3)	33% (1/3)	0% (0/3)
Carboxylic Acid	10	0% (0/7)	43% (3/7)	0% (0/7)	100% (1/1)	-	-	-	-	-	50% (1/2)	0% (0/2)	-0% (0/2)
Ester	9	0% (0/1)	0% (0/1)	0% (0/1)	33% (1/3)	0% (0/3)	0% (0/1)	33% (1/3)	0% (0/1)	0% (0/1)	50% (2/4)	50% (2/4)	0% (0/4)
Heterocyclic	9	0% (0/6)	11% (1/6)	11% (1/6)	0% (0/1)	0% (0/1)	-	0% (0/1)	-	-	50% (1/2)	0% (0/2)	0% (0/2)
Onium Compound	8	0% (0/6)	0% (0/6)	33% (2/6)	-	-	0% (0/1)	-	0% (0/1)	0% (0/1)	100% (1/1)	-	-
Properties of Interest													
Liquids	100	6% (1/18)	17% (3/18)	11% (2/18)	5% (1/19)	21% (4/19)	13% (1/8)	37% (7/19)	-	-	27% (15/55)	9% (5/55)	0% (0/55)
Solids	35	0% (0/12)	58% (7/12)	0% (0/12)	50% (1/2)	0% (0/2)	0% (0/3)	0% (0/2)	0% (0/3)	0% (0/3)	22% (4/18)	0% (0/18)	0% (0/18)
Pesticide	10	0% (0/4)	50% (2/4)	0% (0/4)	0% (0/1)	100% (1/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	50% (2/4)	0% (0/4)	0% (0/4)
Surfactant-Total	21	0% (0/9)	22% (2/9)	22% (2/9)	-	100% (2/2)	0% (0/1)	-	0% (0/1)	0% (0/1)	67% (6/9)	0% (0/9)	0% (0/9)
-nonionic	4	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	100% (2/2)	-	-
Anionic	2	-	100% (1/1)	-	-	-	-	-	-	-	100% (1/1)	-	-

Category	N	Underprediction (*In Vivo/In Vitro*)						Overprediction (*In Vivo/In Vitro*)					
		Category 1			Category 2A		Cat2B	Cat 2A	Category 2B		Not Classified		
		NI	2B	2A	NI	2B	NI	1	2A	1	2B	2A	1
Overall	141	3% (1/29)	34% (10/29)	10% (3/29)	9% (2/22)	18% (4/22)	9% (1/11)	36% (8/22)	18% (2/11)	0% (0/11)	27% (21/79)	8% (6/79)	0% (0/79)
Cationic	7	0% (0/6)	0% (0/6)	33% (2/6)	-	-	-	-	-	-	100% (1/1)	-	-
pH-Total	22	0% (0/20)	30% (6/20)	10% (2/20)	-	-	-	-	-	-	100% (2/2)	-	-
-acidic (pH < 7.0)	14	0% (0/20	25% (3/12)	8% (1/12)	-	-	-	-	-	-	100% (2/2)	-	-
-basic (pH > 7.0)	8	0% (0/20	38% (3/8)	13% (1/8)	-	-	-	-	-	-	-	-	-

Abbreviations: GHS = United Nations Globally Harmonized System for Classification and Labelling of Chemicals; ICE = Isolated Chicken Eye

[1]GHS classification system (UN 2007)

[2]Chemical classes included in this table are represented by at least five substances tested in the ICE test method and assignments are based upon MeSH categories (www.nlm.nih.gov/mesh) as defined in Appendix A.

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

Data Source	Overall Correct Classification	Category I		Category II			Category III			Category IV	
		actual	under	over	actual	under	over	actual	under	over	actual
Prinsen and Koëter (1993)	75% (6/8)	100% (2/2)	0% (0/2)	100% (1/1)	0% (0/1)	0% (0/1)	0% (0/2)	50% (1/2)	50% (1/2)	0%	100%
Balls et al. (1995)	53% (10/19)	47% (9/19)	30% (3/10)	50% (5/10)	20% (2/10)	50% (10/20)	40% (8/20)	10% (2/20)	100% (1/1)	0% (0/1)	53% (10/19)
Prinsen (1996)	81% (29/36)	50% (1/2)	50% (1/2)	0% (0/3)	67% (2/3)	33% (1/3)	0% (0/6)	67% (4/6)	33% (2/6)	12% (3/25)	88% (22/25)
Prinsen (2005)	63% (29/46)	0% (0/4)	100% (4/4)	50% (1/2)	50% (1/2)	0% (0/2)	10% (1/10)	70% (7/10)	20% (2/10)	30% (9/30)	70% (21/30)
Overall[2]	62% (87/140)	48% (13/27)	52% (14/27)	31% (5/16)	50% (8/16)	19% (3/16)	29% (11/38)	53% (20/38)	18% (7/38)	22% (13/59)	78% (46/59)

Abbreviations: EPA = U.S. Environmental Protection Agency; ICE = Isolated Chicken Eye;
[1]EPA classification system (EPA 1996)
[2] Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

6.1.2.2 Identification of Category II Substances (Moderate Ocular Irritants)

For the 16 substances that could be evaluated, the ICE test method correctly identified 50% (8/16) as moderate irritants while 31% (5/16) were overpredicted and 19% (3/16) were underpredicted (**Table 6-6**).

6.1.2.3 Identification of Category III (Mild Ocular Irritants)

For the 38 substances that could be evaluated, the ICE test method correctly identified 53% (20/38) as mild irritants while 29% (11/38) were overpredicted and 18% (7/38) were underpredicted (**Table 6-6**).

6.1.2.4 Identification of Category IV Substances

For the 59 substances that could be evaluated, the ICE test method correctly identified 78% (46/59) as substances not labeled as irritants while 22% (46/59) were overpredicted (**Table 6-6**).

6.1.2.5 Ability to Identify Category IV Substances from All Other Classes

Using this approach for the 140 substances, the ICE test method has an overall accuracy of, of 83% (116/140), a sensitivity of 86% (70/81), a specificity of 78% (46/59), a false positive rate of 22% (13/59), and a false negative rate of 14% (11/81) (**Table 6-7**).

As detailed below, the results from each individual study were also evaluated separately.

Prinsen and Koëter (1993): Based upon the *in vivo* rabbit data eight substances could be assigned a GHS classification. Based on these eight substances the ICE test method has an accuracy of 88% (7/8), sensitivity of 80% (4/5), specificity of 100% (3/3), false positive rate of 0% (0/3), and a false negative rate of 20% (1/5) (**Table 6-7**).

Balls et al. (1995): Based upon the *in vivo* rabbit data 50 substances could be assigned a GHS classification. Based on these 50 substances the ICE test method has an accuracy of 90% (45/50), sensitivity of 92% (45/49), specificity of 0% (0/1), false positive rate of 100% (1/1), and a false negative rate of 8% (4/49) (**Table 6-7**). Two (4-carboxybenzaldehyde and Maneb) of the four false negative substances were from the discordant classes (both solids).

Table 6-7 Accuracy of the ICE Test Method for Distinguishing Category IV Substances from All Other Irritant Classes as Defined by the EPA Classification System[1], by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	8	88	7/8	80	4/5	100	3/3	0	0/3	20	1/5
Balls et al. (1995)	50	90	45/50	92	45/49	0	0/1	100	1/1	8	4/49
Prinsen (1996)	36	83	30/36	73	8/11	88	22/25	12	3/25	27	3/11
Prinsen (2005)	46	74	34/46	81	13/16	70	21/30	30	9/30	19	3/16
Overall	140	83	116/140	86	70/81	78	46/59	22	13/59	14	11/81

Abbreviations: EPA = U.S. Environmental Protection Agency; ICE – isolated chicken eye.

[1] EPA classification system (EPA 1996). Cat IV vs. Cat I/II/III.
[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]Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

Prinsen (1996): Based upon the *in vivo* rabbit data 36 substances could be assigned a GHS classification. Based on these 36 substances the ICE test method has an accuracy of 83% (30/36), sensitivity of 73% (8/11), specificity of 88% (22/25), false positive rate of 12% (3/25), and a false negative rate of 27% (3/11) (**Table 6-7**).

Prinsen (2005): Based upon the *in vivo* rabbit data 46 substances could be assigned a GHS classification. Based on these 46 substances the ICE test method has an accuracy of 74% (34/46), sensitivity of 81% (13/16), specificity of 70% (21/30), a false positive rate of 30% (13/59), and a false negative rate of 14% (11/81) (**Table 6-7**).

6.1.2.6 Performance of the ICE Test Method with Discordant Classes Excluded

The previously identified limitations for the ICE test method are based upon the false positive rate for alcohols and the false negative rates for solids and surfactants when the ICE is used to identify ocular corrosives and severe irritants (ICCVAM 2006a). When the ability of the ICE test method to distinguish Category IV substances from all irritant classes was evaluated with the discordant classes removed, the overall performance statistics were generally unchanged (e.g., overall correct classification increased from 82% to 83%) when these substances were excluded. False negative rates changed from 14% (11/81) to 15% (6/39) and false positive rates changed from 22% (13/59) to 21% (8/39) when the discordant classes were removed (**Table 6-9**).

Table 6-8 Evaluation of the Performance of the ICE Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

ICE	Overall Correct Classification	Category I		Category II			Category III			Category IV	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Overall	62% (87/140)	48% (13/27)	52% (14/27)	31% (5/16)	50% (8/16)	19% (3/16)	29% (11/38)	53% (20/38)	18% (7/38)	22% (13/59)	78% (46/59)
w/o Alcohols	64% (82/128)	48% (12/25)	52% (13/25)	18% (2/11)	55% (6/11)	27% (3/11)	26% (9/35)	54% (19/35)	20% (7/35)	21% (12/57)	79% (45/57)
w/o Surfactants	62% (76/122)	50% (10/20)	50% (10/20)	31% (5/16)	50% (8/16)	19% (3/16)	31% (10/32)	47% (15/32)	22% (7/32)	19% (10/53)	81% (43/53)
w/o Solids	64% (68/107)	59% (10/17)	41% (7/17)	33% (5/15)	53% (8/15)	13% (2/15)	38% (11/29)	52% (15/29)	10% (3/29)	24% (11/46)	76% (35/46)
w/o Alcohols and Surfactants	65% (71/110)	50% (9/18)	50% (9/18)	18% (2/11)	55% (6/11)	27% (3/11)	28% (8/29)	48% (14/29)	24% (7/29)	19% (10/52)	81% (42/52)
w/o Alcohols, Surfactants, and Solids	67% (52/78)	67% (6/9)	33% (3/9)	20% (2/10)	60% (6/10)	20% (2/10)	17% (1/6)	67% (4/6)	17% (1/6)	21% (8/39)	79% (31/39)

Abbreviations: EPA = U.S. Environmental Protection Agency; ICE = Isolated Chicken Eye

[1]EPA classification system (EPA 1996).

Table 6-9 Accuracy of the ICE Test Method for Distinguishing Category IV Substances from All Other Irritant Classes as Defined by the EPA Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

ICE	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall	140	83	116/140	86	70/81	78	46/59	22	13/59	14	11/81
w/o Alcohols	128	82	105/128	85	60/71	79	45/57	21	12/57	15	11/71
w/o Surfactants	122	82	100/122	84	57/68	80	43/54	20	11/54	16	11/68
w/o Solids	107	84	90/107	90	55/61	76	35/46	24	11/46	10	6/61
w/o Alcohols and Surfactants	110	814	89/110	81	47/58	81	42/52	19	10/52	19	11/58
w/o Alcohols, Surfactants and Solids	78	82	69/78	85	33/39	79	31/39	21	8/39	15	6/39

Abbreviations: EPA = U.S. Environmental Protection Agency; ICE – Isolated Chicken Eye

[1] EPA classification system (EPA 1996). Cat IV vs. Cat I/II/III.
[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.

Further analysis of substances for which hazard classification was underpredicted by ICE according to chemical class indicated that carboxylic acids had the highest proportion of underpredicted substances (17% [4/24]). Of the underpredicted substances, 11 were liquids and 12 were solids. Two surfactants were underpredicted by ICE (**Table 6-10**).

According to the EPA classification system, the most overpredicted substances (false positives) were alcohols, which accounted for 21% (6/29) of the overpredicted substances. Of the overpredicted substances, 79% (23/29) were liquids, 2 were solids, and 1 was a surfactant (**Table 6-10**).

Table 6-10 Evaluation of Under and Overprediction of the ICE Test Method Using the EPA[1] Classification System In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method by Chemical Class or Physical Property

Category	N	Underprediction (*In Vivo/In Vitro*)						Overprediction (*In Vivo/In Vitro*)					
		Category I			Category II		Cat III	Cat II	Category III		Category IV		
		IV	III	II	IV	III	IV	I	II	I	III	II	I
Overall	140	4% (1/27)	37% (10/27)	11% (3/27)	19% (3/16)	0% (0/16)	18% (7/38)	31% (5/16)	21% (8/38)	8% (3/38)	22% (13/59)	0% (0/59)	0% (0/50)
Chemical Class[2]													
Alcohol	12	0% (0/2)	50% (1/2)	0% (0/2)	0% (0/5)	0% (0/5)	-0% (0/3)	60% (3/5)	0% (0/3)	67% (2/3)	50% (1/2)	0% (0/2)	0% (0/2)
Carboxylic Acid	10	0% (0/7)	43% (3/7)	0% (0/7)	100% (1/1)	-	0% (0/2)	-	50% (1/2)	0% (0/2)	-	-	-
Ester	9	-	-	-	25% (1/4)	0% (0/4)	0% (0/5)	25% (1/4)	40% (2/5)	0% (0/5)	-	-	-
Heterocyclic	8	0% (0/5)	0% (0/5)	20% (1/5)	0% (0/1)	0% (0/1)	0% (0/2)	0% (0/1)	0% (0/2)	0% (0/2)	-	-	-
Onium Compound	7	0% (0/5)	0% (0/5)	40% (2/5)	-	-	0% (0/2)	-	0% (0/2)	0% (0/2)	-	-	-
Properties of Interest													
Liquids	101	6% (1/17)	18% (3/17)	12% (2/17)	13% (2/15)	0% (0/15)	11% (3/28)	27% (4/15)	25% (7/28)	11% (3/28)	22% (9/41)	0% (0/41)	0% (0/41)
Solids	34	0% (0/10)	70% (7/10)	0% (0/10)	50% (1/2)	0% (0/2)	44% (4/9)	0% (0/2)	0% (0/9)	0% (0/9)	15% (2/13)	0% (0/13)	0% (0/13)
Pesticide	10	0% (0/4)	75% (3/4)	0% (0/4)	0% (0/1)	0% (0/1)	50% (2/5)	0% (0/1)	0% (0/5)	0% (0/5)	50% (1/2)	0% (0/2)	0% (0/2)
Surfactant-Total	20	0% (0/7)	29% (2/7)	0% (0/7)	-	0% (0/1)	0% (0/6)	-	17% (1/6)	0% (0/6)	0% (0/6)	0% (0/6)	0% (0/6)
-nonionic	4	-	-	-	-	0% (0/1)	-	-	100% (1/1)	-	-	-	-
Anionic	2	-	100%	-	-	-	-	-	-	-	-	-	-

Category	N	Underprediction (*In Vivo/In Vitro*)						Overprediction (*In Vivo/In Vitro*)					
		Category I			Category II		Cat III	Cat II	Category III		Category IV		
		IV	III	II	IV	III	IV	I	II	I	III	II	I
Overall	140	4% (1/27)	37% (10/27)	11% (3/27)	19% (3/16)	0% (0/16)	18% (7/38)	31% (5/16)	21% (8/38)	8% (3/38)	22% (13/59)	0% (0/59)	0% (0/50)
			(1/1)										
Cationic	6	0% (0/5)	0% (0/5)	40% (2/5)	-	-	-	-	-	-	-	-	-
pH-Total	19	0% (0/16)	25% (4/16)	6% (1/16)	0% (0/1)	0% (0/1)	0% (0/2)	0% (0/1)	0% (0/2)	0% (0/2)	-	-	-
-acidic (pH < 7.0)	12	0% (0/10)	30% (3/10)	10% (1/10)	-	-	0% (0/2)	-	0% (0/2)	0% (0/2)	-	-	-
-basic (pH > 7.0)	7	0% (0/6)	17% (1/6)	0% (0/6)	0% (0/1)	0% (0/1)	-	0% (0/1)	-	-	-	-	-

Abbreviations: EPA = U.S. Environmental Protection Agency; ICE = Isolated Chicken Eye

[1] EPA classification system (EPA 1996)

[2]Chemical classes included in this table are represented by at least five substances tested in the ICE test method and assignments are based upon MeSH categories (www.nlm.nih.gov/mesh) as defined in Appendix A.

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 174 substances, 153 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification (duplicates removed) according to the EU classification system (EU 2001) (see **Appendix C**). Based on results from *in vivo* rabbit eye experiments, 21% (32/153)[7] were classified as severe irritants (i.e., R41), 18% (28/153) were classified as moderate irritants (i.e., R36), and 61% (93/153) were classified as Not Labeled. The remaining 12% (21/174) substances that could not be classified according to the EU classification system due to the lack of adequate animal data and are so noted in **Appendix C**.

6.1.3.1 Identification of R41 Substances (Ocular Corrosives/Severe Irritants)

The ICE test method correctly identified 59% (19/32) of the R41 substances (**Table 6-11**). Among the remaining 41% (13/32) Category 1 substances that were underpredicted by ICE, 22% (7/32) were classified as R36, 31% (10/32), and 19% (6/32) were classified as Not Labeled.

6.1.3.2 Identification of R36 Substances (Moderate Ocular Irritants)

For the 28 substances that could be evaluated, the ICE test method correctly identified 57% (16/28) as moderate irritants while 18% (5/28) were overpredicted and 25% (7/28) were underpredicted (**Table 6-11**).

6.1.3.3 Identification of Not Labeled Substances

For the 93 substances that could be evaluated, the ICE test method correctly identified 89% (83/93) as substances not labeled as irritants while 11% (10/93) were overpredicted (**Table 6-11**).

[7] 1% sodium hydroxide was duplicated in the database. Sodium hydroxide (Prinsen and Koëter, 1993) was removed because the in vivo classification corresponded to a 10% solution.

Table 6-11 Evaluation of the Performance of the ICE Test Method in Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System[1], by Study and Overall

Data Source	Overall Correct Classification	R41		R36			Not Labeled	
		actual	under	over	actual	under	over	actual
Prinsen and Koëter (1993)	100% (19/19)	100% (7/7)	0% (0/7)	0% (0/3)	100% (3/3)	0% (0/3)	0% (0/9)	100% (9/9)
Balls et al. (1995)	52% (25/48)	56% (10/18)	44% (8/18)	29% (4/14)	50% (7/14)	31% (3/14)	50% (8/16)	50% (8/16)
Prinsen (1996)	94% (34/36)	50% (1/2)	50% (1/2)	0% (0/3)	67% (2/3)	33% (1/3)	8% (3/36)	92% (33/36)
Prinsen (2005)	80% (37/46)	0% (0/4)	100% (4/4)	17% (1/6)	50% (3/6)	33% (2/6)	6% (2/36)	94% (34/36)
Overall[2]	77% (118/153)	59% (19/32)	41% (13/32)	18% (5/28)	57% (16/28)	25% (7/28)	11% (10/93)	89% (83/93)

Abbreviations: EU = European Union; ICE = Isolated Chicken Eye ; NA = Not Applicable

[1]EU classification system (EU 2001)

[2]Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

6.1.3.4 Ability to Identify Not Labeled Substances from All Other Classes

In addition to evaluating the ability of the ICE test method to identify each individual ocular hazard category according to the EU classification system, ICCVAM also evaluated the ability of the ICE test method to distinguish Not Labeled substances from all irritant classes[8]. Using this approach of identifying substances not labeled as irritants from all other classes for the 153 substances considered, the ICE test method has an overall accuracy of 85% (130/153), a sensitivity of 78% (46/60), a specificity of 89% (83/93), a false positive rate of 11% (10/93), and a false negative rate of 22% (13/60) (**Table 6-12**).

As detailed below, the results from each individual study were also evaluated separately.

Prinsen and Koëter (1993): Based upon the *in vivo* rabbit data, 19 substances could be assigned a GHS classification. Based on these 19 substances, the ICE test method has an accuracy of 100% (19/19), sensitivity of 100% (10/10), specificity of 100% (9/9), false positive rate of 0% (0/9), and a false negative rate of 0% (0/10) (**Table 6-12**).

Balls et al. (1995): Based upon the *in vivo* rabbit data, 48 substances could be assigned a GHS classification. Based on these 48 substances, the ICE test method has an accuracy of 69% (33/48), sensitivity of 78% (25/32), specificity of 50% (8/16), false positive rate of 50% (8/16), and a false negative rate of 32% (7/32) (**Table 6-12**). Six of the 7 substances identified as false negatives were from the discordant classes (alcohol, solids, surfactants).

[8] ICCVAM (2006) provides an evaluation of the ICE test method for distinguishing ocular corrosives and severe irritants from all other classes. Since the database of ICE test method results has not changed, this analysis has not been repeated here.

Table 6-12 Accuracy of the ICE Test Method for Distinguishing Not Labeled Substances from All Other Irritant Classes as Defined by the EU Classification System[1], by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)[4]	19	100	19/19	100	10/10	100	9/9	0	0/9	0%	0/10
Balls et al. (1995)	48	69	33/48	78	25/32	50	8/16	50	8/16	32	7/32
Prinsen (1996)	36	94	34/36	60	3/5	100	31/31	0	0/31	40	2/5
Prinsen (2005)	46	89	41/46	70	7/10	94	34/36	6	2/36	30	3/10
Overall[4]	153	85	130/153	78	47/60	89	83/93	11	10/93	22	13/60

Abbreviations: EU = European Union; ICE = Isolated Chicken Eye

[1]EU classification system (EU 2001). Not Labeled vs. R41/R36.
[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]Because Prinsen (2000) includes only four test substances, data from this study were included only in the overall analysis, but were not evaluated separately.

Prinsen (1996): Based upon the *in vivo* rabbit data 36 substances could be assigned a GHS classification. Based on these 36 substances the ICE test method has an accuracy of 94% (34/36), sensitivity of 60% (3/5), specificity of 100% (31/31), false positive rate of 0% (0/31), and a false negative rate of 40% (2/5) (**Table 6-12**).

Prinsen (2005): Based upon the *in vivo* rabbit data 46 substances could be assigned a GHS classification. Based on these 46 substances the ICE test method has an accuracy of 89% (41/46), sensitivity of 70% (7/10), specificity of 94% (34/36), a false positive rate of 6% (2/36), and a false negative rate of 30% (13/60) (**Table 6-12**).

6.1.3.5 Performance of the ICE Test Method with Discordant Classes Excluded

The previously identified limitations for the ICE test method are based upon the false positive rate for alcohols and the false negative rates for solids and surfactants when the ICE is used to identify ocular corrosives and severe irritants (ICCVAM 2006a). For this reason, the performance of the ICE test method for identifying all ocular irritant classes was evaluated with these substances excluded from the database. However, the performance statistics were moderately improved when these substances were excluded relative to the performance with the entire database (**Table 6-13**).

Table 6-13 Evaluation of the Performance of the ICE Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

ICE	Overall Correct Classification	R41		R36			Not Labeled	
		Actual	Under	Over	Actual	Under	Over	Actual
Overall	77% (118/153)	59% (19/32)	41% (13/32)	18% (5/28)	57% (16/28)	25% (7/28)	11% (10/93)	89% (83/93)
w/o Alcohols	78% (109/139)	59% (17/29)	41% (12/29)	13% (3/23)	57% (13/23)	30% (7/23)	9% (8/87)	91% (79/87)
w/o Surfactants	79% (104/132)	63% (15/24)	37% (9/24)	20% (5/25)	60% (15/25)	20% (5/25)	11% (9/83)	89% (74/83)
w/o Solids	77% (89/116)	63% (12/19)	37% (7/19)	20% (5/25)	60% (15/25)	20% (5/25)	14% (10/72)	86% (62/72)
w/o Alcohols and Surfactants	81% (95/118)	62% (13/21)	38% (8/21)	15% (3/20)	60% (12/20)	25% (5/20)	9% (7/77)	91% (70/77)
w/o Alcohols, Surfactants, and Solids	80% (66/82)	67% (6/9)	33% (3/9)	18% (3/17)	65% (11/17)	18% (3/17)	13% (7/56)	87% (49/56)

Abbreviations: EU = European Union; ICE = Isolated Chicken Eye; NA = Not applicable

[1]EU classification system (EU 2001).

When the evaluation was broadened to the ability of the ICE test method to distinguish Not Labeled substances from all irritant classes and the discordant classes were removed, overall accuracy of the ICE method was unchanged at 85% (130/153) and (70/82), false positive and false negative rates also were generally comparable when the discordant classes were removed. False negative rates changed from 22% (13/60) to 19% (5/26) and false positive rates changed from 11% (10/93) to 12% (7/56) when the discordant classes were removed (**Table 6-14**).

Further analysis of underprediction (false negative) results by chemical class indicated that onium compounds were the most underpredicted with 3 of the 20 substances that were underpredicted. Six *in vivo* severe substances (carboxylic acid, heterocyclic, and an inorganic) were under classified as Not Labeled. One of these substances had a pH < 7 while 3 had a pH >7. Regarding the physical form of underpredicted substances, 12 were liquids, 8 were solids, and 6 were surfactants (**Table 6-15**).

According to the EU classification system, the most overpredicted substances (false positives) were alcohols, which accounted for 4 of the 15 substances overpredicted overall. Regarding the physical form of overpredicted substances, 14 were liquids, and 2 were surfactants (**Table 6-15**).

Table 6-14 Accuracy of the ICE Test Method for Distinguishing Not Labeled Substances from All Other Irritant Classes as Defined by the EU Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

ICE	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall	153	85	130/153	78	47/60	89	83/93	11	10/93	22	13/60
w/o Alcohols	139	85	118/139	75	39/52	91	79/87	9	8/87	25	13/52
w/o Surfactants	132	85	112/132	78	38/49	89	74/83	11	9/83	22	11/49
w/o Solids	116	85	99/116	84	37/44	86	62/72	14	10/72	16	7/44
w/o Alcohols and Surfactants	118	85	100/118	73	30/41	91	70/77	9	7/77	27	11/41
w/o Alcohols, Surfactants and Solids	82	85	70/82	81	51/26	88	49/56	12	7/56	19	5/26

Abbreviations: EU = European Union; ICE = Isolated Chicken Eye

[1] EU classification system (EU 2001). NV vs. R41/R36.
[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.

Table 6-15 Evaluation of Under- and Overprediction of the ICE Test Method Using the EU[1] Classification System In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method by Chemical Class or Physical Property

Category	N	Underprediction (*In Vivo/In Vitro*)			Overprediction (*In Vivo/In Vitro*)		
		R41		R36	R36	Not Labeled	
		NI	R36	NI	R41	R36	R41
Overall	153	18% (6/32)	22% (7/32)	25% (7/28)	18% (5/28)	10% (9/93)	1% (1/93)
Chemical Class[2]							
Alcohol	14	0% (0/3)	33% (1/3)	0% (0/5)	40% (2/5)	17% (1/6)	17% (1/6)
Carboxylic Acid	10	17% (1/6	0% (0/6)	50% (1/2)	0% (0/2)	0% (0/2)	0% (0/2)
Ester	9	0% (0/1)	0% (0/1)	33% (1/3)	33% (1/3)	40% (2/5)	0% (0/5)
Heterocyclic	9	17% (1/6)	17% (1/6)	0% (0/1)	0% (0/1)	0% (0/2)	0% (0/2)
Inorganic	5	50% (1/2)	0% (0/2)	0% (0/1)	0% (0/1)	0% (0/2)	0% (0/2)
Onium Compound	8	0% (0/6)	33% (2/6)	100% (1/1)	-	0% (0/1)	0% (0/1)
Polyether	5	-	100% (1/1)	100% (1/1)	-	0% (0/3)	0% (0/3)
Properties of Interest							
Liquids	112	8% (2/24)	21% (5/24)	23% (5/22)	18% (4/22)	14% 9/66	2% (1/66)
Solids	39	27% (4/15)	13% (2/15)	66% (2/3	0% (0/3)	0% (0/21)	0% (0/21)
Pesticide	11	20%	20%	1%	-	0%	0%

Category	N	Underprediction (*In Vivo/In Vitro*)			Overprediction (*In Vivo/In Vitro*)		
		R41		R36	R36	Not Labeled	
		NI	R36	NI	R41	R36	R41
Overall	153	18% (6/32)	22% (7/32)	25% (7/28)	18% (5/28)	10% (9/93)	1% (1/93)
		(1/5)	(1/5)	(1/1)		(0/5)	(0/5)
Surfactant-Total	24	0% (0/9)	44% (4/9)	67% (2/3)	0% (0/3)	17% (2/12)	0% (0/12)
-nonionic	5	-	100% (1/1)	100% (1/1)	-	67% (2/3)	0% (0/3)
Anionic	3	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	-0% (0/1)	-0% (0/1)
Cationic	7	0% (0/6)	33% (2/6)	-	-	0% (0/1)	0% (0/1)
pH-Total	20	22% (4/18)	17% (3/18)	-	-	0% (0/2)	0% (0/2)
-acidic (pH < 7.0)	13	9% (1/11)	18% (2/11)	-	-	0% (0/2)	0% (0/2)
-basic (pH > 7.0)	7	43% (3/7)	14% (1/7)	-	-	-	-

Abbreviations: EU = European Union; ICE = Isolated Chicken Eye

[1] EU classification system (EU 2001)

[2]Chemical classes included in this table are represented by at least five substances tested in the ICE test method and assignments are based upon MeSH categories (www.nlm.nih.gov/mesh) as defined in Appendix A.

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). Quantitative and qualitative evaluations of ICE test method reliability have been conducted previously (ICCVAM 2006a). Since the database used for the current evaluation of the ICE test method has not changed, the quantitative evaluation of test method reliability remains unchanged. However, additional qualitative analyses of test method reproducibility were conducted to evaluate the extent of agreement of ICE hazard classifications among the laboratories.

7.1 Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System

Of 14 substances classified by the GHS as Not Labeled 1/14 (7%) were correctly identified while 2/4 (50%) GHS Category 2B substances were correctly identified, 6/14 (43%) substances classified as GHS Category 2A were correctly identified, and 11/22 (50%) GHS Category 1 substances were correctly identified.

- The four participating laboratories were in 100%, 75% and 50% agreement to the ocular irritancy classification when assessing Not Classified substances from all other classes of 44/59 (75%), 8/59 (14%), and 7/59 (12%), respectively (**Table 7-1**)[9].
- All four participating laboratories agreed on the classification of 7/11 (64%) substances that were correctly identified as GHS Category 1[10], 3/6 (50%) substances correctly classified as GHS Category 2A, 0/2 (0%) substances correctly classified as GHS Category 2B and 0/1 (0%) substance correctly classified as GHS Not Labeled (**Table 7-2**).

[9] Because the database of ICE test method results has not changed, the qualitative evaluation of reproducibility presented in ICCVAM (2006) is not repeated here.

[10] 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 7-1 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Not Classified or Category 1/2A/2B 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 (%)
+/+	38	4[2]	33 (87)	3 (8)	2 (5)
+/-	2	4	0	0	2 (100)
-/+	13	4	7 (54)	4 (31)	2 (15)
-/-	1	4	0	1 (100)	0
?/-	1	4	0	0	1 (100)
?/+	4	4	4 (100)	0	0
TOTAL	59	4[2]	44 (75)	8 (14)	7 (12)

Abbreviation: GHS = United Nations Globally Harmonized System of Classification and Labelling of Chemicals

[1]A "+" indicates that the substance was assigned an overall classification of Mild, Moderate or Corrosive/Severe irritant (I, 2A, 2B); a "-" indicates that the substance was assigned a classification of nonirritant (NI); 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.

Table 7-2 Evaluation of the Interlaboratory Variability of Balls et al. (1995) In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the GHS Classification System

In vivo Classification (No.)[1]	Classification *(in vitro)*	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 (%)
NC (14)	Actual	1 (7)	4	0	1 (100)	0
	Over	13 (93)	4	7 (54)	4 (31)	2 (15)
2B (4)	Under	0	4	0	0	0
	Actual	2 (50)	4	0	1 (50)	1 (50)
	Over	2 (50)	4	0	2 (100)	0
2A (14)	Under	2 (14)	4	0	0	2 (100)
	Actual	6 (43)	4	3 (50)	0	3 (50)
	Over	6 (43)	4	1 (17)	0	5 (83)
1 (22)	Under	11 (50)	4	9 (82)	2 (18)	0
	Actual	11 (50)	4[2]	7 (64)	3 (27)	1(9)

Abbreviation: GHS = United Nations Globally Harmonized System of Classification and Labelling of Chemicals; NC = Not Classified

[1]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 for 5 substances. 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.

- Three of the four laboratories were in agreement for the 2/11 (18%) of the GHS Category 1 substances correctly identified, 0 GHS Category 2A substances, 1/2 (50%) GHS Category 2B substances, and 1/1 (100%) of the Not Labeled substances correctly.

- Two of the four laboratories were in agreement for 1/11 (9%) of the GHS Category 1 substances identified correctly, 3/6 (50%) of GHS Category 2A substances, 1/2 (50%) of GHS Category 2B substances, and 0/1 (0%) of the GHS Not Classified substances (**Table 7-2**). The labs with discordant data were not consistent within or across the irritant classes.

Of 14 substances classified by the GHS as Not Classified, 13/14 (93%) were incorrectly identified while 2/4 (50%) GHS Category 2B substances were incorrectly identified, 8/14 (57%) Category 2A substances were incorrectly identified, and 11/22 (50%) GHS Category 1 substances were incorrectly identified (**Table 7-2**).

- The four participating laboratories were in 100% agreement in incorrectly classifying 9 (82%) of the GHS Category 1 substances, 1 (17%) of the GHS Category 2A substances, 0 of the GHS Category 2B substances, and 7 (54%) of the GHS Not Classified substances (**Table 7-2**).

- Three of the four laboratories were in agreement of incorrectly classifying 2 (18%) of the GHS Category 1 substances, 0 of the GHS Category 2A substances, 2 (100%) of Category 2B substances, and 4 (31%) of the GHS Not Classified substances (**Table 7-2**).

- Two of the four laboratories were in agreement of incorrectly classifying 0 of the GHS Category 1 substances, 7 (50%) of the GHS Category 2A substances, 0 of the GHS Category 2B substances, and 2 (15%) of the GHS Not Classified substances (**Table 7-2**).

7.2.2 Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System

Of 2 substances classified by the EPA as Category IV, 0/2 (0%) were correctly identified while 8/20 (40%) EPA Category III substances were correctly identified, 5/12 (42%) of the substances EPA Category II substances were correctly identified, and 10/19 (53%) of the EPA Category I substances were correctly identified.

- The four participating laboratories were in 100%, 75% and 50% agreement in regard to the ocular irritancy classification when assessing Category IV substances from all other classes of 44/59 (75%), 8/59 (14%), and 7/59 (12%), respectively (**Table 7-3**)[11].
- All four participating laboratories agreed on the classification of 7/10 (70%) substances that were correctly identified as EPA Category I[12], 3/5 (60%) substances correctly classified as EPA Category II, 1/8 (13%) correctly classified as EPA Category III, and 0 substances classified as Category IV (**Table 7-4**).
- Three of the four laboratories were in agreement for the 2/10 (20%) of the EPA Category Is correctly identified, 1/5 (20%) of the EPA Category IIs, 3/8 (38%) of the EPA Category IIIs, and 0 of the substances classified as Category IV (**Table 7-4).** The discordant laboratory was not consistent among these substances.
- Two of the four laboratories were in agreement for 1/10 (10%) of the EPA Category I substances identified correctly, 1/5 (20%) of the EPA Category II substances, 4/8 (50%) of the EPA Category III substances correctly identified, and 0 of the substances classified as Category IV (**Table 7-4**).

[11] Because the database of ICE test method results has not changed, the qualitative evaluation of reproducibility presented in ICCVAM (2006) is not repeated here.

[12] 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.

Of two substances classified by the EPA as Category IV, 2/2 (100%) were incorrectly identified while 12/20 (60%) substances classified as EPA Category III were incorrectly identified, 5/12 (42%) EPA Category II substances were incorrectly identified, and 9/19 (47%) EPA Category I substances were incorrectly identified (**Table 7-4**).

- The four participating laboratories were in 100% agreement in incorrectly classifying 7/9 (78%) of the EPA Category I substances, 1/5 (20%) of the EPA Category II substances, and 6/12 (50%) of the EPA Category III substances, and 0 of the EPA Category IV substances (**Table 7-4**).
- Three of the four laboratories were in agreement of incorrectly classifying 2/9 (22%) of the EPA Category I substances, 1/5 (20%) of the EPA Category II substances, 4/12 (33%) of the Category III substances, and 2/2 (100%) of the EPA Category IV substances, (**Table 7-4).** The lab with the discordant results was not consistent within and across the irritant classes.
- Two of the four laboratories were in agreement of incorrectly classifying 0/9 (0%) of the EPA Category I substances, 3/5 (60%) of the EPA Category II substances, 2/12 (17%) of the EPA Category III substances, and 0 of the EPA Category IV substances (**Table 7-4**).

Table 7-3 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Category IV or Category I/ II/III 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 (%)
+/+	47	4[2]	38 (81)	5 (11)	4 (9)
+/-	4	4	0	1 (25)	3 (75)
-/+	2	4	0	2 (100)	0
-/-	0	4	0	0	0
?/-	0	4	0	0	0
?/+	6	4	6 (100)	0	0
TOTAL	59	4[2]	44 (75)	8 (14)	7 (12)

Abbreviation: EPA = U.S. Environmental Protection Agency

[1]A "+" indicates that the substance was assigned an overall classification of Severe, Moderate or Mild irritant (I, II, III); a "-" indicates that the substance was assigned a classification of 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), a 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.

Table 7-4 Evaluation of the Interlaboratory Variability of Balls et al. (1995) In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the EPA Classification System

In vivo Classification (No.)[1]	Classification *(in vitro)*	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 (%)
IV (2)	Actual	0	4	0	0	0
	Over	2 (100)	4	0	2 (100)	0
III (20)	Under	2 (10)	4	0	1 (50)	1 (50)
	Actual	8 (40)	4	1 (13)	3 (38)	4 (50)
	Over	10 (50)	4	6 (60)	3 (30)	1 (10)
II (12)	Under	2 (17)	4	0	1 (50)	1 (50)
	Actual	5 (42)	4	3 (60)	1 (20)	1 (20)
	Over	3 (25)	4	1 (33)	0	2 (67)
1 (19)	Under	9 (47)	4	7 (78)	2 (22)	0
	Actual	10 (53)	4[2]	7 (70)	2 (20)	1 (10)

Abbreviation: EPA = U.S. Environmental Protection Agency

[1]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 for 6 substances. 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.

7.2.3 Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System

Of 17 substances classified by the EU as Not Labeled, 9/17 (53%) were correctly identified while 7/14 (50%) substances classified as EU moderate irritants were correctly identified, and 10/19 (53%) substances classified by the EU as corrosive/severe irritants were correctly identified.

- The four participating laboratories were in 100%, 75% and 50% agreement in regard to the ocular irritancy classification when assessing Not Labeled substances from all other classes of 36/59 (61%), 15/59 (25%), and 8/59 (14%), respectively (**Table 7-5**)[13].
- All four participating laboratories agreed on the classification of 7/10 (70%) substances that were correctly identified as EU R41, 4/7 (57%) substances correctly classified as EU R36, and 3/9 (33%) correctly classified as EU Not Labeled (**Table 7-6**).
- Three of the four laboratories were in agreement for the 2/10 (20%) of the EU R41 correctly identified, 2/7 (29%) of the EU R36s, 4/9 (44%) of the EU substances classified as Not Labeled (**Table 7-6).** The discordant laboratory was not consistent among these substances.
- Two of the four laboratories were in agreement for 1/10 (10%) of the EU R41 substances identified correctly, 1/7 (14%) of the EU R36 substances, and 2/9 (22%) of the EU substances classified as Not Labeled (**Table 7-6**).

Of 17 substances classified by the EU as Not Labeled 8/17 (47%) were incorrectly identified while 7/14 (50%) substances classified as EU R36 substances were incorrectly identified, and 9/19 (47%) substances classified as EU R41 were incorrectly identified (**Table 7-6**).

- The four participating laboratories were in 100% agreement in incorrectly classifying 7/9 (78%) of the EU R41 substances, 1/7 (14%) of the EU R36 substances, and 5/8 (63%) of the EU Not Labeled substances (**Table 7-6**).

[13] Because the database of ICE test method results has not changed, the qualitative evaluation of reproducibility presented in ICCVAM (2006) is not repeated here.

- Three of the four laboratories were in agreement of incorrectly classifying 2/9 (22%) of the EU R41 substances, 2/7 (29%) of the EU R36 substances, and 1/8 (13%) of the EU Not Labeled substances (**Table 7-6**).
- Two of the four laboratories were in agreement of incorrectly classifying 0/9 (0%) of the EU R41 substances, 4/7 (57%) of the EU R36 substances, 2/8 (25%) of the EU Not Labeled substances (**Table 7-6**).

Table 7-5 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Not Labeled or R41/R36 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 (%)
+/+	26	4[2]	22 (85)	3 (12)	1 (4)
+/-	7	4	2 (29)	3 (42)	2 (29)
-/+	8	4	5 (63)	1 (13)	2 (25)
-/-	9	4	3 (33)	4 (44)	2 (22)
?/-	1	4	0	1 (100)	0
?/+	8	4	4 (50)	3 (38)	1 (13)
TOTAL	59	4[2]	36 (61)	15 (25)	8 (14)

Abbreviation: EU = European Union

[1]A "+" indicates that the substance was assigned an overall classification of Severe or Non-Severe irritant (Category R41 or R36); a "-" indicates that the substance was assigned a classification of nonirritant (Category NI); 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 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.

Table 7-6 Evaluation of the Interlaboratory Variability of Balls et al. (1995) In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the EU Classification System

In vivo Classification[1] (No.)	Classification *(in vitro)*	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 (%)
NL (17)	Actual	9 (53)	4	3 (33)	4 (44)	2 (22)
	Over	8 (47)	4	5 (63)	1 (13)	2 (25)
R36 (14)	Under	3 (21)	4	0	2 (67)	1 (33)
	Actual	7 (50)	4	4 (57)	2 (29)	1 (14)
	Over	4 (29)	4	1 (25)	0	3 (75)
R41 (19)	Under	9 (47)	4	7 (78)	2 (22)	0
	Actual	10 (53)	4[2]	7 (70)	2 (20)	1 (10)

Abbreviation: EU = European Union; NL = Not Labeled

[1]Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), a EU classification could not be made for 9 substances. 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.

8.0 Test Method Data Quality

The database used in this assessment did not change from that used in the previous assessment of the ability of the ICE method to identify ocular corrosives and severe irritants. The evaluation of ICE test method data quality is detailed in ICCVAM (2006a).

9.0 Other Scientific Reports and Reviews

No new data and published or unpublished studies have been located since the previous evaluation of the ICE method in identifying ocular corrosives and severe irritants (ICCVAM 2006a).

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. Substances that are identified as corrosive or severe irritants *in vitro* are excluded from *in vivo* testing. Furthermore, the ability to identify mild and moderate ocular irritants would eliminate the need for *in vivo* testing thus sparing rabbits from the pain associated with these types of substances.

The ICE test method can also reduce animal use because the test method was adapted from the IRE test method that used animal species routinely raised as a food source in large numbers to replace the need for laboratory animals. Additionally, with the ability to identify ocular corrosives and severe ocular irritants as well as mild and moderate ocular irritants 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).

11.0 Practical Considerations

Practical considerations for the ICE method are detailed in ICCVAM (2006a).

12.0 References

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.

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.

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

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.

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. Toxic Substances Control Act. 40 CFR 792.

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.

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. 2006a. Background Review Document: Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method. NIH Publication No: 06-4513. Research Triangle Park:National Toxicology Program.

ICCVAM. 2006b. ICCVAM Test Method Evaluation Report: *In Vitro* Ocular Toxicity Methods for Identifying Severe Irritants and Corrosives. NIH Publication No: 07-4517. 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].

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

Occupational Safety and Health Act of 1970 (OSHA). 1970. Public Law 91-596

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:

1637 OECD Guidelines for Testing of Chemicals. OECD, Paris.

1638 Prinsen MK. 1991a. Acute eye irritation/corrosion study with 1-butanol in albino rabbits.
1639 TNO-report V91.187/210069, November 1991. Dept. of Biological Toxicology, TNO
1640 Toxicology and Nutrition Institute, Zeist, The Netherlands. (Available on request)

1641 Prinsen MK. 1991b. Acute eye irritation/corrosion study with 2-methoxyethanol in albino
1642 rabbits. TNO-report V91.188/210069, November 1991. Dept. of Biological Toxicology,
1643 TNO Toxicology and Nutrition Institute, Zeist, The Netherlands. (Available on request)

1644 Prinsen MK. 1996. The chicken enucleated eye test (CEET): A practical (pre)screen for the
1645 assessment of eye irritation/corrosion potential of test materials. Food Chem Toxicol 34:291-
1646 296.

1647 Prinsen MK. 2000. Chicken encleated eye test with reference surfactants and siloxane
1648 polymers in Phase II of the reference standard validation project; an *ex vivo* alternative to the
1649 Draize eye irritation test with rabbits. TNO Report V99.521b. Unpublished report provided
1650 directly to NICEATM by M Prinsen, TNO Nutrition and Food Research Institute.

1651 Prinsen MK. 2005. *In vitro* and *in vivo* data for 94 substances tested in the isolated chicken
1652 eye test. Unpublished data provided directly to NICEATM by MK Prinsen, TNO Nutrition
1653 and Food Research Institute.

1654 Prinsen MK, Koëter BWM. 1993. Justification of the enucleated eye test with eyes of
1655 slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. Food
1656 Chem Toxicol 31:69-76.

1657 Russell WMS, Burch RL. 1992. The Principles of Humane Experimental Technique. 14th ed.
1658 South Mimms, Potters Bar, Herts:Universities Federation for Animal Welfare.

1659 UN. 2003. Globally Harmonised System of Classification and Labelling of Chemicals
1660 (GHS). New York & Geneva:United Nations Publications.

1661

13.0 Glossary[14]

Accuracy[15]: (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.

---

[14] The definitions in this Glossary are restricted to their uses with respect to the Draize rabbit eye test method and the ICE test method.

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

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.

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

1894 are used as the basis for a conclusion that may not be evident from the individual data.