Appendix F

Background Review Document
Current Status of *In Vitro* Test Methods for Identifying
Mild/Moderate Ocular Irritants:
The Isolated Chicken Eye (ICE) Test Method

This page intentionally left blank

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

2010

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

This page intentionally left blank

Table of Contents

List of Tables

List of Figures

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
FHSA	U.S. Federal Hazardous Substances Act
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 Database
IRE	Isolated rabbit eye
MeSH	U.S. National Library of Medicine Medical Subject Heading
MMAS	Modified maximum average score
NC	Not Classified (as irritant)
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIH	National Institutes of Health
NL	Not Labeled (as irritant)
OECD	Organisation for Economic Co-operation 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.
Bruce Fowler, Ph.D.
Edward Murray, Ph.D.
Eric Sampson, 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
Harry Salem, Ph.D.
Peter J. Schultheiss, D.V.M., DACLAM

Department of Energy
* Michael Kuperberg, Ph.D.
+ Marvin Stodolsky, Ph.D.

Department of the Interior
* Barnett A. Rattner, Ph.D.
+ Sarah Gerould, Ph.D. (to Feb. 2009)

Department of Transportation
* George Cushmac, Ph.D.
+ Steve Hwang, Ph.D.

Environmental Protection Agency
Office of Pesticide Programs
* John R. "Jack" Fowle III, Ph.D., DABT
+ Vicki Dellarco, Ph.D.
+ Tina Levine, Ph.D.
Deborah McCall
Christine Augustyniak, Ph.D. *(U.S. Coordinator, OECD Test Guidelines Program)*
Office of Pollution Prevention and Toxics
Jerry Smrchek, Ph.D. *(U.S. Coordinator, OECD Test Guidelines Program*, to July 2009*)*
Office of Research and Development
Suzanne McMaster, Ph.D. (to Dec. 2008)
Julian Preston, Ph.D. (to July 2009)
Stephanie Padilla, Ph.D. (to July 2009)
Office of Science Coordination and Policy
Karen Hamernik, Ph.D. (to July 2009)

Food and Drug Administration
Office of the Commissioner
* Suzanne Fitzpatrick, Ph.D., DABT
Center for Biologics Evaluation and Research
Richard McFarland, Ph.D., M.D.
Ying Huang, Ph.D.
Center for Devices and Radiological Health
Melvin E. Stratmeyer, Ph.D.
Vasant G. Malshet, Ph.D., DABT
Center for Drug Evaluation and Research
+ Abigail C. Jacobs, Ph.D.
Paul C. Brown, Ph.D.
Center for Food Safety and Applied 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
Paul Howard, Ph.D.
Donna Mendrick, Ph.D.
William T. Allaben, Ph.D. (to Jan. 2009)
Office of Regulatory Affairs
Lawrence D'Hoostelaere, Ph.D.

National Cancer Institute
* T. Kevin Howcroft, Ph.D.
Chand Khanna, D.V.M., Ph.D.
Alan Poland, M.D. (to Oct. 2008)

National Institute of Environmental Health Sciences
* William S. Stokes, D.V.M., DACLAM
+ Raymond R. Tice, Ph.D.
Rajendra S. Chhabra, Ph.D., DABT
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.

* Principal agency representative
\+ Alternate principal agency representative

Acknowledgements

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

U.S. Consumer Product Safety Commission
Marilyn L. Wind, Ph.D.
Adrienne Layton, Ph.D.

U.S. Department of Defense
Harry Salem, Ph.D.

U.S. Department of Transportation
Steve Hwang, Ph.D.

U.S. Environmental Protection Agency
Office of Pesticide Programs
Meta Bonner, Ph.D.
Jonathan Chen, Ph.D.
John R. "Jack" Fowle III, Ph.D., DABT
Masih Hashim, D.V.M., Ph.D.
Karen Hicks
Marianne Lewis
Debbie McCall
Timothy McMahon, Ph.D.
Mark Perry
John Redden
Jenny Tao, Ph.D.
Office of Research and Development
Andrew Geller, Ph.D.
Office of Science Coordination and Policy
Karen Hamernik, Ph.D.

U.S. Food and Drug Administration
Center for Drug Evaluation and Research
Paul Brown, Ph.D.
Wiley Chambers, M.D.
Abigail (Abby) Jacobs, Ph.D.
Jill Merrill, Ph.D., DABT (OTWG Chair)
Center for Food Safety and Applied Nutrition
Robert Bronaugh, Ph.D.
Donnie Lowther
Office of the Commissioner
Suzanne Fitzpatrick, Ph.D., DABT

National Institute of Environmental Health Sciences
Warren Casey, Ph.D., DABT
Mark F. Cesta, D.V.M, DACVP
Raymond (Buck) Grissom, Ph.D.
William Stokes, D.V.M., DACLAM

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

European Centre for the Validation of Alternative Methods – Liaison
João Barroso, Ph.D.
Thomas Cole, 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., DACLAM
Director; Project Officer

Warren Casey, Ph.D., DABT
Deputy Director

Deborah McCarley
Special Assistant; Assistant Project Officer

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

David Allen, Ph.D.
Jonathan Hamm, Ph.D.
Nelson Johnson
Brett Jones, Ph.D.
Elizabeth Lipscomb, Ph.D.
Linda Litchfield
Steven Morefield, M.D.
Catherine Sprankle
James Truax, M.A.
Linda Wilson

Statistical Consultant for ILS, Inc.

Joseph Haseman, Ph.D.

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

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

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

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

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

Penny Jones
Unilever Research
Sharnbrook, United Kingdom

Menk Prinsen
TNO Nutrition & Food Research Institute
The Netherlands

Horst Spielmann, Dr.med.
ZEBET
Berlin, Germany

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

ICCVAM and NICEATM gratefully acknowledge the following individuals and institutions that submitted data to NICEATM for the evaluation of alternative ocular safety testing methods and approaches

ECVAM
Chantra Eskes, Ph.D.

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

TNO Nutrition & Food Research Institute
Menk Prinsen

This page intentionally left blank

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, and U.S. Occupational Safety and Health 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. Tests now require only one to three rabbits, compared to six rabbits per test 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) test methods for the identification of ocular corrosives or severe (irreversible) ocular irritants. ICCVAM used the EPA (2003a), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (UN 2007), and European Union (EU 2001) regulatory hazard classification systems. In ICCVAM's assessment, the performance of the BCOP and ICE test methods substantiated their use in testing some substances for regulatory hazard classification. The IRE and HET-CAM test methods lacked sufficient performance and/or sufficient data to substantiate their use for regulatory hazard classification.

ICCVAM recommended that the BCOP and ICE test methods 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* (ICCVAM 2006b). 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 purposes.

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 [EPA Category II, III; EU Category R36, GHS Category 2A, 2B]) and substances Not Classified as irritant (GHS NC or Not Labeled, EPA Category IV, FHSA Not Labeled, or EU Not Labeled) according to the GHS (UN 2007), EPA (EPA 2003a), FHSA (FHSA 2005), and EU (EU 2001) classification systems. The Federal Hazardous Substances Act (FHSA) classification system (FHSA 2005) as defined in the "Test for Eye Irritants" (i.e., "Irritant" or Not Labeled [as an irritant]) and published in 16 CFR 1500.42 (CPSC 2003) is also provided in the current background review documents. The FHSA classification system was not used in the previous analyses of test methods used for the identification of severe ocular irritants or corrosives because the FHSA classification is limited to irritants and is not intended to identify corrosive substances or to differentiate between severe and nonsevere 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 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 background review documents (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 and the Japanese Center for the Validation of Alternative Methods will provide input and contribute to the OTWG throughout the evaluation process.

An international independent scientific peer review panel (Panel) met in public session on May 19-21, 2009, to develop conclusions and recommendations on the *in vitro* BCOP, ICE, IRE, and HET-CAM test methods. The Panel included expert scientists nominated by the European Centre for the Validation of Alternative Methods and the Japanese Center for the Validation of Alternative Methods. We anticipate that these organizations can use the subsequent independent Panel report (ICCVAM 2009) to deliberate and develop their own test method recommendations. The Panel considered these BRDs and evaluated the extent to which the available information supports the draft ICCVAM test method recommendations.

ICCVAM provided the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) with the draft BRD and draft Test Method Evaluation Report, the Panel's report, and all public comments. SACATM discussed these at their June 25–26, 2009, meeting, where public stakeholders were given another opportunity to comment. After SACATM's meeting, ICCVAM considered the SACATM comments, the Panel report, and all public comments before finalizing the Background Review Document and test method recommendations. These recommendations 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 who helped prepare this BRD, including the following staff from the NICEATM support contractor, Integrated Laboratory Systems, Inc.: David Allen, Jon Hamm, Nelson Johnson, Brett Jones, Elizabeth Lipscomb, Linda Litchfield, Steven Morefield, 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. (U.S. EPA) and Jill Merrill, Ph.D. (U.S. Food and Drug Administration), 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 Ocular Toxicity Working Group from the European Centre for the Validation of Alternative Methods and the Japanese Center for the Validation of Alternative Methods, 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., DACLAM
Rear Admiral/Assistant Surgeon General, U.S. Public Health Service
Director, NICEATM
Executive Director, ICCVAM

Executive Summary

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 the evaluation of several activities related to reducing, refining, and replacing the use of rabbits in the current *in vivo* Draize rabbit eye test (69 FR 13859 [March 24, 2004]). In response to this nomination, ICCVAM evaluated the validation status of the bovine corneal opacity and permeability (BCOP), hen's egg test–chorioallantoic membrane (HET-CAM), isolated chicken eye (ICE), and isolated rabbit eye (IRE) test methods. To evaluate how well these test methods identify ocular corrosives and severe irritants, ICCVAM used the EPA (2003a), European Union (EU 2001), and United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (UN 2007) classification systems.

ICCVAM considered the performance of two of these *in vitro* test methods, the BCOP and the ICE, to be sufficient to support their use in testing certain types of substances for regulatory hazard classification. The IRE and HET-CAM test methods lacked sufficient performance and/or sufficient data to support their use for regulatory hazard classification. ICCVAM recommended that the BCOP and ICE test methods should be used in a tiered-testing strategy that would classify positive substances as ocular corrosives or severe irritants without animal testing. These recommendations were accepted by U.S. Federal agencies, 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 to identify nonsevere ocular irritants (those that cause reversible ocular damage [EPA Category II and III; EU R36; GHS Category 2A and 2B]) and substances not labeled as irritants (EPA Category IV; EU Not Labeled; GHS Not Classified) according to the EPA (2003a), EU (2001), and GHS (UN 2007) classification systems. The FHSA classification system, which is based on the testing guidelines and associated criteria included in 16 CFR 1500.42 (CPSC 2003), is also included in these evaluations. The FHSA classification system was not used in the original analyses (ability of the test methods to identify ocular corrosives and severe irritants) because the FHSA ocular hazard category that is assigned based on results from the Draize rabbit eye test (Draize et al. 1944) does not distinguish between ocular corrosives and severe irritants and less severe irritants. For this reason, an evaluation to identify ocular corrosives and severe irritants using the FHSA classification system was not possible.

Because the FHSA classification system (2005) is based on a sequential testing strategy that uses up to 18 animals, only a small percentage of the substances in the ICE database would be classifiable if the FHSA criteria were strictly applied. To maximize the number of substances included in these analyses, "proportionality" criteria were applied for the purpose of assigning an FHSA classification to test results that would require additional testing according to the FHSA sequential testing strategy. These "proportionality" criteria (FHSA-20% and FHSA-67%) are as follows:

- FHSA-20% is based on the proportion of positive animals needed to identify a substance as an irritant using the FHSA sequential testing strategy, where 20% of the animals must demonstrate a positive response for a substance to be identified as an irritant. A substance tested using 3 to 6 animals would not be labeled if ≤1/6 animals were positive based on the FHSA criteria. The substance would be labeled as an irritant if there were ≥1 positive animal in a 3- to 5-animal test or ≥2 positive animals in a 6-animal test.
- FHSA-67% is based on the proportion of positive animals needed to identify a substance as an irritant using the "first test" of the FHSA sequential testing strategy, where 67% of the animals must demonstrate a positive response for a substance to be identified as an irritant. A substance tested using 3 to 6 animals would not be labeled as an irritant if ≤1/6 animals were positive based on the FHSA criteria. The substance would be labeled as an

irritant if there were ≥2/3, 3/4, 4/5, or 4/6 positive animals. If 1/3, 1/4, 2/4, 1/5, 2/5, 3/5, 2/6, or 3/6 animals were positive, further testing would be required.

Together, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the ICCVAM Ocular Toxicity Working Group 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 all available information for the ICE test method and its current validation status, including what is known about its reliability and accuracy, and the scope of the substances tested. Original 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.

ICE Test Method Protocol

The ICE test method is an *in vitro* model that provides short-term maintenance of the chicken eye. Damage caused by a 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 (1) converted into a quantitative score that is used to calculate an overall irritation index or (2) assigned a qualitative category that is used to assign an *in vitro* ocular irritancy classification. Either outcome can then be used to predict the *in vivo* ocular irritation potential of a test substance.

Validation Database

No new ICE data have been obtained since ICCVAM evaluated the ICE test method for identifying ocular corrosives and severe irritants (ICCVAM 2006a). Therefore, the same database was used in the current evaluation. The ICE validation database contains a total of 175 substances. The most commonly tested chemical classes tested are alcohols, carboxylic acids, esters, and heterocyclics. Of the 175 substances, 48% (85/175) could not be assigned a specific chemical class. The most commonly tested product classes are solvents, soaps/surfactants, industrial chemicals, and pesticides/herbicides. Thirteen percent (23/175) could not be assigned a product class.

In order to calculate the appropriate EPA (2003a), EU (2001), FHSA (2005), and GHS (UN 2007) ocular irritancy hazard classifications, detailed *in vivo* data consisting of cornea, iris, and conjunctiva scores for each animal at 24, 48, and 72 hours following test substance administrations and/or assessment of the presence or absence of lesions at 7, 14, and 21 days are needed. Some of the test substances had only limited *in vivo* data and so could not be used to evaluate test method accuracy and reliability. To maximize the number of substances included in the FHSA analyses, "proportionality" criteria (FHSA-20% and FHSA-67%), as outlined above, were applied for the purpose of assigning a FHSA classification to test results that would require additional testing according to the FHSA sequential testing strategy.

ICE Test Method Accuracy

Identification of All Ocular Hazard Categories

ICCVAM evaluated how well the ICE test method identified all categories of ocular irritation potential as defined by the EPA (2003a), GHS (UN 2007), and EU (2001) classification systems. Because the FHSA classification system does not distinguish between ocular corrosives and severe irritants and less severe irritants, an evaluation for all ocular hazard categories using the FHSA classification system was not possible. Analyses were also performed excluding specific chemical classes and/or physical properties that were previously identified as discordant in the ICE test method (alcohols, surfactants, and solids) relative to the *in vivo* hazard classification (ICCVAM 2006a).

As shown in **Table 1**, overall correct classifications ranged from 59% (83/141) to 77% (118/153) when using the entire database, depending on the hazard classification system used. When discordant

classes are excluded, overall correct classifications improved slightly to a range of 64% (49/77) to 80% (66/82), depending on the classification system used.

Distinguishing Substances Not Labeled as Irritants from All Other Hazard Categories

ICCVAM also evaluated how well the ICE test method distinguished substances not labeled as irritants (EPA Category IV, EU Not Labeled, FHSA Not Labeled, GHS Not Classified) from all other ocular hazard categories (EPA Categories I, II, III; EU R41, R36; FHSA Irritant; GHS Categories 1, 2A, 2B) as defined by the EPA (2003a), GHS (UN 2007), EU (2001), and FHSA (2005) classification systems. Analyses were also performed excluding specific chemical classes and/or physical properties that were previously identified as discordant in the ICE test method (alcohols, surfactants, and solids) relative to the *in vivo* hazard classification (ICCVAM 2006a).

As shown in **Table 2**, overall accuracy ranged from 78% (110/141) to 85% (130/153), depending on the hazard classification system used. The lowest false negative rate (6% [4/62]) was noted for the GHS system, followed by 9% (7/76) for the FHSA-67% criteria, 12% (10/84) for the FHSA-20% criteria, 14% (11/81) for the EPA system, and 22% (13/60) for the EU system. Among these false negatives, at least one substance was classified as an ocular corrosive and severe irritant based on Draize rabbit eye test data (n=1 each for the EPA and GHS systems, and n=6 for the EU system). The lowest false positive rate (11% [10/93]) was noted for the EU system, followed by 22% (13/59) for the EPA system, 24% (15/62) for the FHSA-20% and FHSA-67% criteria, and 34% (27/79) for the GHS system. The exclusion of discordant classes had no affect on accuracy (ranged from 75% [58/77] to 85% [70/82] when discordant classes were removed versus 78% [110/141] to 85% [130/153] for overall accuracy, depending on the hazard classification system used).

ICE Test Method Reliability

Interlaboratory Reproducibility

Previous quantitative and qualitative evaluations of the reliability of the ICE test method have been conducted (ICCVAM 2006a). Because the database used for the current evaluation of the ICE test method has not changed, the quantitative evaluation of test method reliability remains unchanged. Additional qualitative analyses of interlaboratory reproducibility were conducted to evaluate how well the ICE hazard classifications agreed among the four participating laboratories from the interlaboratory validation study (Balls et al. 1995). These evaluations were based on the use of the ICE test method (1) to identify all ocular hazard categories according to the EPA, GHS, or EU systems, and (2) to distinguish substances not labeled as irritants (EPA Category IV, GHS Not Classified, EU Not Labeled) from all other ocular hazard categories (EPA Categories I, II, III; GHS Categories 1, 2A, 2B; EU R41, R36). Because the performance of the ICE test method was similar for the EPA and FHSA hazard classification systems, additional reliability analyses were not conducted for the FHSA hazard classification system.

Table 1 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, GHS, and EU Classification Systems[1]

Hazard Classification System	Overall Correct Classification	Severe[2]		Moderate[3]			Mild[4]			Not Labeled[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)
Without Alcohols, Surfactants, and Solids[6]	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 (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)
Without 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)
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)
Without Alcohols, Surfactants, and Solids	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)

Abbreviations: EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = Globally Harmonized System; ICE = isolated chicken eye; NA = not applicable.

[1] EPA classification system (EPA 2003a); GHS classification system (UN 2007); EU classification system (EU 2001). Because the FHSA classification system does not distinguish between ocular corrosives and severe irritants and less severe irritants, an evaluation for all ocular hazard categories using the FHSA classification system was not possible.

[2] Severe = EPA Category I; GHS Category 1, EU R41.

[3] Moderate = EPA Category II; GHS Category 2A; EU R36.

[4] Mild = EPA Category III; GHS Category 2B.

[5] Not Labeled = EPA Category IV; GHS Not Classified; EU Not Labeled.

[6] Alcohols, surfactants, and solids were previously identified as discordant in the ICE test method relative to the *in vivo* hazard classification (ICCVAM 2006a).

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

Hazard Classification System	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall (EPA)[1]	140	83	116/140	86	70/81	78	46/59	22	13/59	14	11/81
Without Alcohols, Surfactants, and Solids[2]	78	82	69/78	85	33/39	79	31/39	21	8/39	15	6/39
Overall (GHS)[3]	141	78	110/141	94	58/62	66	52/79	34	27/79	6	4/62
Without Alcohols, Surfactants, and Solids	77	75	58/77	89	24/27	68	34/50	32	16/50	11	3/27
Overall (EU)[4]	153	85	130/153	78	47/60	89	83/93	11	10/93	22	13/60
Without Alcohols, Surfactants, and Solids	82	85	70/82	81	51/26	88	49/56	12	7/56	19	5/26
Overall (FHSA-20%)[5]	146	83	121/146	88	74/84	76	47/62	24	15/62	12	10/84
Without Alcohols, Surfactants, and Solids	76	82	62/76	86	31/36	78	31/40	23	9/40	14	5/36
Overall (FHSA-67%)[5]	138	84	116/138	91	69/76	76	47/62	24	15/62	9	7/76
Without Alcohols, Surfactants, and Solids	72	82	59/72	88	28/32	78	31/40	23	9/40	13	4/32

Abbreviations: EPA = U.S. Environmental Protection Agency; EU = European Union; FHSA = Federal Hazardous Substances Act; GHS = Globally Harmonized System; ICE = isolated chicken eye; N = number of substances included in this analysis; No. = data used to calculate the percentage.

[1] EPA classification system (EPA 2003a): Category IV vs. Category I/II/III.

[2] Alcohols, surfactants, and solids were previously identified as discordant in the ICE test method relative to the in vivo hazard classification (ICCVAM 2006a).

[3] GHS classification system (UN 2007): Not Classified vs. Category 1/2A/2B.

[4] EU classification system (EU 2001): Not Labeled vs. R41/R36.

[5] FHSA classification system (FHSA 2005): Not Labeled vs. Irritant. To maximize the number of substances included in the FHSA analyses, "proportionality" criteria (FHSA-20% and FHSA-67%) were applied for the purpose of assigning a FHSA classification to test results that would require additional testing according to the FHSA sequential testing strategy.

Using the first approach (identifying all ocular hazard categories), there was 100% agreement among the four laboratories for a majority of the Draize ocular corrosives and severe irritants based on all three classification systems, whether they were correctly identified or underclassified by the ICE test method. For example, for the EPA system, there was 100% agreement for 70% (7/10) of the correctly identified Category I substances. There was also 100% agreement among the four laboratories for at least 50% (3/6 to 3/5) of the correctly identified moderate ocular irritants (EPA Category II, GHS Category 2A, EU R36). For the mild ocular irritants (EPA Category III, GHS Category 2B), there was 100% agreement among the four laboratories for 0% (0/2) to 13% (1/8) of the correctly identified substances. The four laboratories had only 50% agreement for 50% (4/8 or 1/2) of these substances for the EPA and GHS classification systems. A majority of the substances not classified as irritants (EPA Category IV, EU Not Labeled, GHS Not Classified) based on Draize results were overclassified by the ICE test method. The four laboratories had at least 75% agreement for all but two of these substances. For example, there was at least 75% agreement for 85% (11/13) of the GHS Not Labeled substances overclassified by the ICE test method. The four laboratories had at least 75% agreement for 76% (13/17) of the EU Not Labeled substances, whether they were correctly identified or overclassified by the ICE test method. For example, there was at least 75% agreement for 77% (7/9) of the EU Not Labeled substances that were correctly identified and 75% (6/8) of those overclassified by the ICE test method.

Using the second approach (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) study. There was 100% agreement among the four laboratories for 81% (38/47) of the substances correctly identified as irritants according to the EPA system (Category I, II, III). While none of the EPA Category IV substances was correctly identified by the ICE test method, there was 75% agreement among the four laboratories for both of the Category IV substances that were overpredicted by the ICE test method.

The four laboratories had 100% agreement for 87% (33/38) of the substances correctly identified as irritants according to the GHS system (Category 1, 2A, 2B). While only one of the GHS substances not labeled as irritants was correctly identified by the ICE test method (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 GHS substances not labeled as irritants that were overpredicted by the ICE test method. There was 100% agreement among the four laboratories for 85% (22/26) of the substances correctly identified as irritants according to the EU system (R36 or R41). The laboratories had at least 75% agreement for 77% (7/9) of the substances correctly identified as Not Labeled.

1.0 Introduction

1.1 Background

The current Draize rabbit eye test method identifies both irreversible (i.e., corrosive) and reversible ocular effects. It also provides quantitative scoring with which to categorize the severity of reversible effects such as mild, moderate, or severe irritation. The current U.S. Environmental Protection Agency health effects test guideline for acute eye irritation (EPA 1998) and United Nations Globally Harmonized System (GHS) of Classification and Labelling of Chemicals (UN ocular testing strategy) indicate that if serious ocular damage is anticipated (e.g., a lesion considered to be irreversible or persisting for 21 days), then a test on a single animal may be considered. If serious damage is observed, no further animal testing is necessary (EPA 1998; UN 2007). If no serious damage is observed, additional test animals (1 or 2 rabbits) may be evaluated sequentially until concordant irritant or nonirritant responses are observed based on the GHS (UN 2007) or until unequivocal results are obtained in a minimum of three animals according to the EPA test guideline (EPA 1998). In the FHSA classification system (FHSA 2005), which is based on the testing guidelines and associated criteria included in 16 CFR 1500.42 (CPSC 2003), corrosive substances are identified by other test methods (e.g., Draize skin test or human accidental exposure data) and excluded from further irritant testing.

In 2006, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) completed an evaluation of the isolated chicken eye (ICE) test method for its ability to identify ocular corrosives and severe irritants (ICCVAM 2006a). 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, GHS Category 1, European Union [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 test method can be classified as ocular corrosives or severe irritants without further testing in rabbits, while a substance that tests negative would need additional testing in rabbits using a sequential testing strategy as outlined in Organisation for Economic Co-operation and Development Test Guideline 405 (OECD 2002).

ICCVAM is now evaluating the usefulness and limitations of the ICE test method for identifying nonsevere irritants (i.e., those that induce reversible ocular damage [EPA Category II and III; EU R36; GHS Category 2A and 2B]) and substances not labeled as irritants (i.e., EPA Category IV; EU Not Labeled; FHSA Not Labeled; GHS Not Classified) according to the EPA, EU, FHSA, and GHS classification systems (EPA 2003a; EU 2001; FHSA 2005; UN 2007). However because the FHSA classification system (2005) is based on a sequential testing strategy, which uses up to 18 animals, only a small percentage of the substances in the ICE database would be classifiable if the FHSA criteria were strictly applied. In order to maximize the number of substances included in these analyses, “proportionality” criteria (i.e., FHSA-20% and FHSA-67%) were applied for the purpose of assigning an FHSA classification for test results that would require additional testing according to the FHSA sequential testing strategy (see **Section 4.1**).

As part of the 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. An ICCVAM expert panel used this BRD when reviewing the ICE test method to identify all categories of ocular irritants and substances not labeled as irritants.

Parallel review of the ICE, isolated rabbit eye (IRE), hen’s egg test–chorioallantoic membrane (HET-CAM), and bovine corneal opacity and permeability (BCOP) test methods were conducted. The expert panel report and the analyses presented in the BRDs were used to support ICCVAM

recommendations on the proposed standardized test method protocols, proposed list of recommended reference substances, and additional optimization and/or validation studies that may be necessary to further develop and characterize the usefulness and limitations of these methods.

For a more detailed discussion on the background of the ICE test method, including its scientific basis and regulatory rationale and applicability, see 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).

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 ocular corrosives/severe irritants without further testing. The GHS currently recommends the ICE test method for use in identifying ocular corrosives and severe irritants in a tiered-testing strategy for regulatory classification and labeling (UN 2007).

1.3 Validation of the ICE Test Method

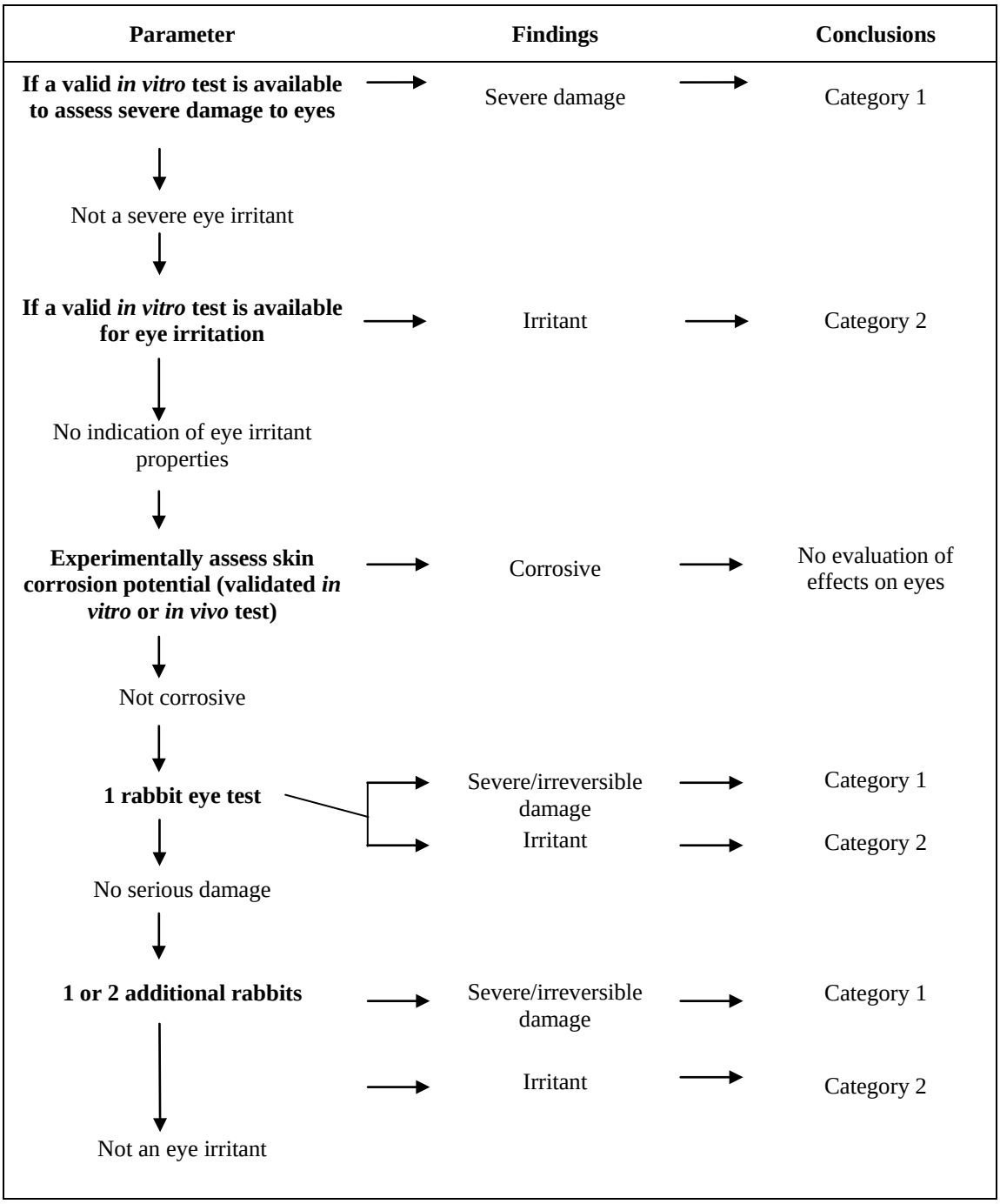
The ICCVAM Authorization Act of 2000 (Sec. 4([c]) mandates that "each Federal Agency … shall ensure that any new or revised … test method … is determined to be valid for its proposed use prior to requiring, recommending, or encouraging [its use]" (Public Law 106-545).

Validation is the process that establishes the reliability and relevance of a test method for a specific purpose (ICCVAM 1997). *Relevance* is defined as the extent to which a test method 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 test method identifies (1) substances that are capable of producing nonsevere ocular irritation or (2) substances not labeled as irritants.

Reliability is defined as the reproducibility of a test method within and among laboratories. Reliability should be based on performance with a diverse set of substances that (1) represent the types of chemical and product classes likely to be tested and (2) cover the range of responses that need to be identified. The validation process will provide data and information to 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 substances' eye irritation potential.

The first stage in this evaluation is the preparation of a BRD that presents and evaluates the relevant data and information about the test method, 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 performance of the test method is evaluated.

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



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

[1] Adapted from UN (2007).

1.4 Search Strategies and Selection of Citations for the ICE BRD

The ICE test method data summarized in this BRD are derived from peer-reviewed scientific literature detail 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 the same as the database previously used (ICCVAM 2006a).

2.0 Isolated Chicken Eye Test Method Protocol Components

The ICE test method is an *in vitro* model that provides short-term maintenance of the chicken eye. Damage caused 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 (1) converted into a quantitative score that is used to calculate an overall irritation index or (2) assigned a qualitative categorization that is used to assign an *in vitro* ocular irritancy classification. Either outcome 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 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). 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 are dissected within 2 hours after death. After dissection, the eyes are placed in a superfusion apparatus, where isotonic saline is applied to the cornea at a rate of 2 to 3 drops per minute through a steel tube attached to a peristaltic pump. Substances are applied as a single dose (30 µL for liquids, 30 mg for solids) for 10 seconds.

Corneal swelling and opacity are measured at regular intervals for up to 4 hours after treatment. Fluorescein retention is evaluated 30 minutes after treatment only. Mean values for each parameter (corneal swelling, corneal opacity, and fluorescein retention) are determined. The maximum mean value for each parameter is classified in one of four irritancy categories as shown in **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 30 Minutes After 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 endpoint are 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

Overall *In Vitro* Classification	Combinations of the Three Endpoints
Nonirritant	3 x I
	2 x 1, 1 x II
Mild Irritant	3 x II
	2 x II, 1 x I
	2 x 11, 1 x III
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, GHS Not Classified, EU Not Labeled, FHSA Not Labeled; Mild Irritant = EPA Category III, GHS Category 2B; Moderate Irritant = EPA Category II, GHS Category 2A; Severe Irritant = EPA Category I, GHS Category 1, EU Category R41. The Mild and Moderate Irritant categories were combined to generate EU Category R36. The Mild, Moderate, and Severe Irritant categories were combined to generate FHSA Irritant.

To date, this scheme has been published only as an application to the EU classification system (EU 2001). However, using this same scheme, 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 corresponding *in vivo* classification for each of the EPA, GHS, and EU classification systems (EPA 2003a; EU 2001; UN 2007). For the FHSA classification system, the *in vivo* classification was compared to the *in vitro*

classification based on the EPA classification system. *In vitro* classifications of Mild, Moderate, and Severe Irritant were classified as FHSA Irritant; and Nonirritant was classified as FHSA Not Labeled.

3.0 Substances Used for Validation of the ICE Test Method

Validation studies for *in vitro* ocular test methods 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.

No new ICE test method data have been obtained since ICCVAM originally evaluated the ICE test method for identification of ocular corrosives and severe irritants (ICCVAM 2006a). Therefore, the same database (n=175 substances) (derived from Balls et al. 1995; Prinsen 1996, 2000, 2005; Prinsen and Koëter 1993) was used in the current evaluation.

Tables 3-1 and **3-2** show the chemical and product classes of the test substances 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 for the test substances are provided in **Annex I**. If not assigned in the study report, the product class was sought from other sources, including the National Library of Medicine's ChemIDplus® 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), which ensures consistency in classifying substances among all *in vitro* ocular test methods under consideration. A substance could be classified 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 chemical classes tested most often in the ICE test method are alcohols, carboxylic acids, esters, and heterocyclics. Of the 175 substances included in the database used for this assessment (see **Annex I)**, 85 (including formulations of unidentified composition) could not be assigned a specific chemical class.

As shown in **Table 3-2**, the product classes tested most in the ICE test method are solvents, soaps/surfactants, industrial chemicals, and pesticides/herbicides. Of the 175 substances (see **Annex I)**, 23 could not be assigned a product class.

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 byproduct	1	Paint	4
Cleaner	8	Pesticide/Herbicide	15
Copolymer	3	Pharmaceutical compound	5
Cosmetic ingredient	1	Preservative	6
Detergent	8	Raw material	9
Developer	1	Reagent	4
Disinfectant	5	Resin	2
Dyes and stains	10	Silicone resin	1
Elastomer	2	Soap	9
Enzyme inhibitor	1	Solvent	37
Enzyme solution	3	Surfactant	25

4.0 *In Vivo* Reference Data Used to Assess Isolated Chicken Eye 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 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 number of national and international test guidelines also describe this procedure (EPA 1998; OECD 2002; CPSC 2003; EU 2004). The subjective scoring system used to assign 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 NICEATM to assign an ocular hazard classification according to the EPA (2003a), EU (2001), FHSA (2005), and the GHS (UN 2007) ocular irritancy classification systems (**Annex III**). Exceptions include 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 (2003a), GHS (UN 2007), and FHSA (2005) 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 were 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) includes toxicological summaries that provide a recommended EU classification for each of the chemicals. In three cases, there were adequate summary *in vivo* data with which to also generate irritancy classifications for the EPA (2003a) 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 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), the *in vivo* rabbit eye test database that was used to analyze the accuracy of the ICE test method includes studies conducted using from one to six rabbits. However, some of the *in vivo* classification systems considered for the accuracy analyses are designed for application to studies using no more than three rabbits. Thus, to maximize the amount of data used to evaluate 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. The criteria used for classification according to the EPA (2003a), EU (2001), or GHS (UN 2007), classification systems were detailed in the 2006 ICCVAM BRD (ICCVAM 2006a). Each of these classification systems requires that the Draize scoring system be used. For these classification systems, 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 the 2006 ICCVAM BRD, the following four criteria must have been met.

- 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 fewer than three rabbits.
- A volume of 0.1 mL or 100 mg was tested in each rabbit. A study in which a lower quantity was applied to the eye could be 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 were made at least 24, 48, and 72 hours after test substance application if no severe effect was observed.
- Observations of the eye were 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.

For the FHSA classification system (FHSA 2005), the testing guidelines and associated criteria are included in 16 CFR 1500.42 (CPSC 2003). The FHSA classification system is based on using up to three sequential tests for each test substance with six animals used per test (**Table 4-1**). Decisions on further sequential testing are based on the number of positive responses in each test. The severity of effects for each endpoint (i.e., corneal ulceration and opacity, conjunctival redness and/or swelling, and iritis) is measured at 24, 48, and 72 hours after test substance administration. Positive responses include corneal ulceration (other than a fine stippling), corneal opacity or iritis ≥1, and conjunctival swelling and/or redness ≥2. In the first test, six animals are tested. If ≥4 animals are positive, the test is positive. If ≤1 animal tests positive, the test is negative. If 2/6 or 3/6 animals are positive, then a second test is performed with six additional animals. A third test is needed if 1/6 or 2/6 animals are positive with the second test.

Table 4-1 FHSA Classification System (16 CFR 1500.42)[1,2]

Positive Response for a Single Rabbit[3] ≥1 of the following at 24, 48, and/or 72 hours	***In Vivo* Effect**
Corneal ulceration (other than a fine stippling) Corneal opacity (CO) ≥1 Iritis (IR) ≥1 Conjuctival redness (CR) and/or chemosis (CC) ≥2	First Test – If ≥4/6 animals are positive, the test is positive. If ≤1 animal is positive, the test is negative. If 2/6 or 3/6 animals are positive, the test is repeated using a different group of six animals. Second Test – If ≥3/6 animals are positive, the test is positive. If 0/6 animals are positive, the test is negative. If 1/6 or 2/6 animals are positive, the test is repeated using a different group of six animals. Third Test – Should a third test be needed, the test is positive if ≥1/6 animals are positive. If 0/6 animals are positive, the test is negative.

Abbreviations: CC = conjunctival chemosis; CFR = Code of Federal Regulations; CO = corneal opacity; CR = conjunctival redness; FHSA = Federal Hazardous Substances Act; IR = iritis.

[1] For the FHSA Classification System (2005), the testing guidelines and associated criteria are included in 16 CFR 1500.42 (CPSC 2003).

[2] At least three animals per test (one animal screen for corrosive/severe irritants permitted). Maximum score in any animal used for classification.

[3] The following scores are considered positive: CO or IR ≥1 or CR or CC ≥2. Therefore, CO and IR scores of 0 or CR and CC scores ≤1 are considered negative.

The FHSA classification system (FHSA 2005) is a binary system, which classifies substances that test positive (according to the criteria provided in **Table 4-1)** as irritants and substances that test negative

as not requiring labeling (i.e. FHSA Not Labeled). Based on the FHSA sequential testing strategy, a substance can be classified as an eye irritant hazard with as few as 22% of the animals having a positive response (i.e., 2/6 [first test] +1/6 [second test] +1/6 [third test] = 4/18 or 22%).

Because the FHSA classification system is based on a sequential testing strategy, which uses up to 18 animals, only a small percentage of the substances in ICE database would be classifiable if the FHSA criteria were strictly applied. In order to maximize the number of substances include in these analyses, "proportionality" criteria were developed by NICEATM for the purpose of assigning an FHSA classification for test results that would require additional testing according to the FHSA sequential testing strategy (**Table 4-2**).

Table 4-2 Proposed FHSA "Proportionality" Criteria

No. of Animals in Test	FHSA-20%[1]		FHSA-67%[1]		
	NL	Irritant	NL	Irritant	Further Testing Required[2]
3	0/3	≥1 (≥33%)	0/3	≥2 (≥67%)	1/3
4	0/4	≥1 (≥25%)	0/4	≥3 (≥75%)	1/4, 2/4
5	0/5	≥1 (≥20%)	0/5	≥4 (≥80%)	1/5, 2/5, 3/5
6	0/6, 1/6	≥2 (≥33%)	0/6, 1/6	≥4 (≥67%)	2/6, 3/6

Abbreviations: CPSC = U.S. Consumer Product Safety Commission; FHSA = Federal Hazardous Substances Act; NL = Not Labeled (as irritant); No. = number.

[1] FHSA-20% and FHSA-67% analysis methods are based on the proportionality of positive animals needed to identify a substance as an irritant.

[2] For FHSA-67%, Further Testing Required refers to substances that do not meet adequate positive or negative criteria to be classified.

These "proportionality" criteria (i.e., FHSA-20% and FHSA-67%) are as follows:

- (FHSA-20%) – FHSA-20% is based on the proportion of positive animals needed to identify a substance as an irritant using the FHSA sequential testing strategy, where 20% of the animals need to demonstrate a positive response for a substance to be identified as an irritant. A substance tested using 3 to 6 animals would not be labeled if ≤ 1/6 animals were positive based on the FHSA criteria. The substance would be labeled as an irritant if there were ≥1 positive animal in a 3 to 5 animal test or ≥2 positive animals in a 6 animal test.
- (FHSA-67%) – FHSA-67% is based on the proportion of positive animals needed to identify a substance as an irritant using the "first test" of the FHSA sequential testing strategy, where 67% of the animals need to demonstrate a positive response for a substance to be identified as an irritant. A substance tested using 3 to 6 animals would not be labeled if ≤ 1/6 animals were positive based on the FHSA criteria. The substance would be labeled as an irritant if there were ≥2/3, 3/4, 4/5, or 4/6 positive animals. If 1/3, 1/4, 2/4, 1/5, 2/5, 3/5, 2/6, or 3/6 animals were positive, further testing would be required.

4.2 *In Vivo* Data Quality

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with Good Laboratory Practice (GLP) guidelines. GLP guidelines are nationally and

internationally recognized rules designed to produce high-quality laboratory records (OECD 1998; EPA 2003b, 2003c; 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, thereby ensuring the integrity, reliability, and accountability of a study.

The extent to which the *in vivo* rabbit eye studies that were used to provide the comparative data in the published ICE validation studies complied 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 Isolated Chicken Eye Test Method Data and Results

A total of five reports, three published (Balls et al. 1995; Prinsen 1996; Prinsen and Koëter 1993) and two unpublished (Prinsen 2000, 2005), included sufficient data for an accuracy analysis of the ICE test method for the identification of all categories of ocular irritation. **Section 6.0** details how 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 Nutrition and Food Research. Summaries of ICE results (i.e., total scores) but no 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 focuses on the regulatory applicability of the ICE test method, and MMAS scores are not used for regulatory classification, this BRD did not use this approach. 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

The information extracted for the database used in this assessment includes, when provided, the following specifics:

- Name
- CASRN (if available)
- Chemical class and/or product class
- Concentration(s) tested
- Purity
- Form tested
- ICE test method endpoint values (maximum mean)
- *In vitro* classification
- Supplier or source
- Literature reference

If not provided, the CASRN was obtained from various sources, including the National Library of Medicine's ChemIDplus® 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 to each test substance based on the MeSH® classification system (available at http://www.nlm.nih.gov/mesh). **Annex I** provides information on the names, synonyms, CASRNs, and chemical/product classes, where available, for

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

each substance. **Annex II** provides 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 (EPA 2003b, 2003c; FDA 2003; OECD 1998). The data quality was evaluated by reviewing 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 Isolated Chicken Eye Test Method Accuracy

A critical component of an ICCVAM evaluation of a test method's validation status is an assessment of the proposed test method's accuracy when compared to that of the current reference test method (ICCVAM 2003). This aspect of test method 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

ICCVAM evaluated the ability of the ICE test method to identify all categories of ocular irritation potential as defined by the EPA (EPA 2003a), GHS (UN 2007), and EU (EU 2001) classification systems. Given that the FHSA classification system is used to identify eye irritants based on incidence and does not differentiate between irreversible (i.e., corrosive or severe) and reversible (i.e., nonsevere) ocular effects based on Draize rabbit eye test results, an evaluation for all ocular hazard categories using the FHSA classification system was not possible.

Analyses were also performed with specific chemical classes and/or physical properties excluded based on their previous identification as discordant in the ICE test method (ICCVAM 2006a). These evaluations were conducted on the overall data set created by combining results from the reports discussed in **Section 5.0** (Balls et al. 1995; Prinsen 1996, 2005; Prinsen and Koëter 1993) then assigning an overall ocular irritancy classification for each substance. (See **Annexes II** and **III**). 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 were an equal number of different irritancy classifications for substances (e.g., two tests classified a substance as 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 (i.e., mild irritant, in this case).

6.1 GHS Classification System: ICE Test Method Accuracy

The four studies (Balls et al. 1995; Prinsen 1996, 2005; Prinsen and Koëter 1993) 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 **Annex III**). 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%

---

[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. Another chemical (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 Category 2A/2B.

(79/141) were classified as Not Labeled as Irritant. The remaining 33 substances could not be classified according to the GHS classification system due to the lack of adequate animal data and are so noted in **Annex III**.

6.1.1 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) were classified as Not Classified as Irritant.

6.1.2 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**).

6.1.3 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**).

6.1.4 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 classified as irritants, while 34% (27/79) were overpredicted (**Table 6-1**).

6.1.5 Ability to Distinguish Substances Not Classified as Irritants 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 substances not classified as irritants 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 negative 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. Among 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] The 2006 ICCVAM BRD provides an evaluation of the ICE test method for distinguishing ocular corrosives and severe irritants from all other classes. Because the database of ICE test method results has not changed, this analysis has not been repeated here.

Table 6-1 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	Severe (Category 1)		Moderate (Category 2A)			Mild (Category 2B)			Not Classified as Irritant	
		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 and were not evaluated separately.

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

Data Source	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	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[2]	141	78	110/141	94	58/62	66	52/79	34	27/79	6	4/62

Abbreviations: GHS = Globally Harmonized System; ICE = isolated chicken eye; N = number of substances included in this analysis; No. = data used to calculate the percentage.

[1] GHS classification system (UN 2007): Not Classified as Irritant vs. Category 1/2A/2B.

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

Balls et al. (1995): Based upon the *in vivo* rabbit data, 50 substances could be assigned a GHS classification. Among 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. Among 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. Among 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**).

6.1.6 Performance of the ICE Test Method with Discordant Classes Excluded

The previous ICCVAM BRD identified limitations of the ICE test method 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 in identifying all ocular irritant classes was evaluated with these substances excluded from the database. The overall performance statistics improved slightly (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 substances not labeled as irritants 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**). Following the removal of substances belonging to discordant classes (i.e., alcohols, surfactants and solids; see also ICCVAM 2006a), there were three GHS ocular irritants classified as Not Classified as Irritant using the ICE test method (i.e., false negatives; see **Table 6-5**). Among the three false negatives for the GHS system, 33% (1/3) were GHS Category 2B substances, 33% (1/3) were GHS Category 2A substances, and 33% (1/3) were GHS Category 1 substances.

Table 6-3 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 Discordant Chemical and Physical Classes Excluded

ICE	Overall Correct Classification	Severe (Category 1)		Moderate (Category 2A)			Mild (Category 2B)			Not Classified as Irritant	
		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)
Without 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)
Without 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)
Without 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)
Without 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)
Without 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 = Globally Harmonized System; ICE = isolated chicken eye.

[1] GHS classification system (UN 2007).

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

ICE	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	141	78	110/141	94	58/62	66	52/79	34	27/79	6	4/62
Without Alcohols	129	78	100/129	92	49/53	67	51/76	33	25/76	8	4/53
Without Surfactants	122	79	96/122	92	47/51	69	49/71	31	22/71	8	4/51
Without Solids	107	76	81/107	93	43/46	62	38/61	38	23/61	7	3/46
Without Alcohols and Surfactants	109	78	85/109	90	37/41	71	48/68	29	20/68	10	4/41
Without Alcohols, Surfactants, and Solids	77	75	58/77	89	24/27	68	34/50	32	16/50	11	3/27

Abbreviations: GHS = Globally Harmonized System; ICE = isolated chicken eye; N = number of substances included in this analysis/the total number of substances in the study; NC = Not Classified (as an irritant); No. = data used to calculate the percentage.

[1] GHS classification system (UN 2007): NC vs. Category 1/2A/2B.

Table 6-5 ICE False Negative Substances[1]

Substance	*In Vivo* Classification					*In Vivo* Scores		
	EPA	GHS	EU	FHSA-20%	FHSA-67%	N	Corneal Opacity: Score (Day Cleared)	Conjunctival Redness: Score (Day Cleared)
TNO-94[2]	I	1	R41	Irr	Irr	3	N=1 2(7)	N=2 3(14)
TNO-28[3] (toilet bowl cleaner-1)	I	1	R41	Irr	Irr	3	None	N=1 2(7) N=1 3(28)
Methyl cyanoacetate	II	2A	R36	Irr	Irr	3	N=1 1(2) N=1 1(7)	N=1 3(7) N=2 3(14)
TNO-9 (paint)	II	NC	NL	Irr	Irr	3	N=1 2(14)	N=1 2(2) N=1 3(3)
DMSO	III	2B	NL	Irr	FTR	3	None	N=1 2(3) N=1 2(4)
Methyl cyclopentane	III	NC	NL	NL	NL	6	None	N=1 2(2)
TNO-3 (pesticide)	III	NC	NL	Irr	Irr	3	None	N=1 2(2) N=1 2(3)
TNO-29 (toilet bowl cleaner-2)	III	2A	R36	Irr	Irr	3	N=1 1(2) N=1 1(3)	N=1 3(7) N=1 2(14) N=1 3(14)
TNO-52	III	2A	R36	Irr	Irr	3	N=3 1(7)	N=3 3(14)

Abbreviations: DMSO = dimethyl sulfoxide; EPA = U.S. Environmental Protection Agency; EU = European Union; FHSA = U.S. Federal Hazardous Substances Act; FTR = further testing required; GHS = Globally Harmonized System; ICE = isolated chicken eye; Irr = irritant; N = number of animals; NC = Not Classified (as irritant); NL = Not Labeled (as irritant); TNO = TNO Nutrition and Food Research Institute, Netherlands.

For the purposes of this evaluation, *clearing* is defined in the EPA hazard classification system as corneal opacity or iritis scores = 0 or redness or chemosis scores = 1; in the GHS and EU hazard classification systems as corneal opacity, iritis, redness, or chemosis scores = 0.

[1] False negative compounds (shaded here) are those that test as nonirritants *in vitro* but are mild, moderate, or severe ocular irritants/corrosive *in vivo*, i.e., EPA Categories I, II, and III; GHS Categories 1, 2A, and 2B; and EU R41 and R36.

[2] One animal with ischemic necrosis of conjunctiva; study terminated.

[3] One animal with ischemic necrosis of conjunctiva.

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

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

6.2 EPA Classification System: ICE Test Method Accuracy

The four studies (Balls et al. 1995; Prinsen 1996, 2005; Prinsen and Koëter 1993) 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 2003a) (see **Annex III**). 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 34 substances could not be classified according to the EPA classification system due to the lack of adequate animal data and are so noted in **Annex III**.

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

The ICE test method correctly identified 48% (13/27) of the Category I substances (**Table 6-7**). Among the remaining 52% (14/27) of the Category I substances underpredicted by the ICE test method, 11% (3/27) were classified as Category II, 37% (10/27) were classified as Category III, and 4% (1/27) were classified as Category IV.

[5] One substance (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 Category II/III.

Table 6-6 Under- and Overprediction of the ICE Test Method Using the GHS Classification System[1] 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*)					
		Severe (Category 1)			Moderate (Category 2A)		Mild (Cat 2B)	Moderate (Cat 2A)	Mild (Category 2B)		Not Classified (NC)		
		NC	2B	2A	NC	2B	NC	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[3]	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)
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)
Solids[3]	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)

continued

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

Category	N	Underprediction (*In Vivo/In Vitro*)						Overprediction (*In Vivo/In Vitro*)					
		Severe (Category 1)			Moderate (Category 2A)		Mild (Cat 2B)	Moderate (Cat 2A)	Mild (Category 2B)		Not Classified (NC)		
		NC	2B	2A	NC	2B	NC	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)
Properties of Interest (continued)													
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)	-	-
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 = Globally Harmonized System; ICE = isolated chicken eye; N = number of substances; NC = Not Classified/not labeled as irritant.

[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 National Library of Medicine medical subject heading (MeSH) categories (www.nlm.nih.gov/mesh) as defined in Annex I.

[3] Physical form (i.e., solid and liquid) not known for some substances; therefore, the overall number does not equal the sum of the solid and liquid substances.

Table 6-7 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	Severe (Category I)		Moderate (Category II)			Mild (Category III)			Not Labeled (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% (0/3)	100% (3/3)
Balls et al. (1995)	46% (23/50)	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)
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 2003a).

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

6.2.2 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 Category II irritants, while 31% (5/16) were overpredicted and 19% (3/16) were underpredicted (**Table 6-7**).

6.2.3 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 6-7**).

6.2.4 Identification of Category IV Substances (Not Labeled)

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 6-7**).

6.2.5 Ability to Distinguish Category IV Substances from All Other Classes

Using this approach for the 140 substances, the ICE test method had an overall accuracy 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 6-8**).

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 an EPA classification. Among these eight substances, the ICE test method had 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 6-8**).

Balls et al. (1995): Based upon the *in vivo* rabbit data, 50 substances could be assigned an EPA classification. Among 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 6-8**). Two (4-carboxybenzaldehyde and maneb) of the four false negative substances were from the discordant classes (both solids).

Prinsen (1996): Based upon the *in vivo* rabbit data, 36 substances could be assigned an EPA classification. Among these 36 substances, the ICE test method had 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 6-8**).

Prinsen (2005): Based upon the *in vivo* rabbit data, 46 substances could be assigned an EPA classification. Among these 46 substances, the ICE test method had an accuracy of 74% (34/46), sensitivity of 81% (13/16), specificity of 70% (21/30), a false positive rate of 30% (9/30), and a false negative rate of 19% (3/16) (**Table 6-8**).

Table 6-8 Accuracy of the ICE Test Method in Distinguishing Category IV Substances from All Other Irritant Classes as Defined by the EPA Classification System,[1] by Study and Overall

Data Source	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	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[2]	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; N = number of substances included in this analysis/the total number of substances in the study; No. = data used to calculate the percentage.

[1] EPA classification system (EPA 2003a): Category IV vs. Category I/II/III.

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

6.2.6 Performance of the ICE Test Method with Discordant Classes Excluded

The ICE test method limitations identified in the 2006 ICCVAM BRD were based upon the false positive rate for alcohols and the false negative rates for solids and surfactants when the ICE test method is used to identify ocular corrosives and severe irritants (ICCVAM 2006a). For this reason, the performance of the ICE test method in identifying all ocular irritant classes was evaluated with these substances excluded from the database. The overall performance statistics improved slightly (e.g., overall correct classification increased from 59% to 64%) when these substances were excluded (**Table 6-9**).

When the ability of the ICE test method to distinguish Category IV substances from all other irritant classes was evaluated with the discordant classes removed, the overall accuracy was generally unchanged (e.g., overall accuracy decreased from 83% to 82%) 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 6-10**).

Following the removal of substances belonging to discordant classes (i.e. alcohols, surfactants and solids, see also ICCVAM [2006a]), there were six EPA ocular irritants classified as Category IV using the ICE test method (i.e. were false negatives, see **Table 6-5**). Among the six false negatives for the EPA system, 50% (3/6) were EPA Category III substances, 33% (2/6) were EPA Category II substances, and 17% (1/6) were EPA Category I substances.

Further analysis of substances for which hazard classification was underpredicted by the ICE test method 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 the ICE test method (**Table 6-11**).

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 6-11**).

Table 6-9 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 Discordant Chemical and Physical Classes Excluded

ICE	Overall Correct Classification	Severe (Category I)		Moderate (Category II)			Mild (Category III)			Not Labeled (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)
Without 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)
Without 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)
Without 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)
Without 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)
Without 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 2003a).

Table 6-10 Accuracy of the ICE Test Method in Distinguishing Category IV Substances from All Other Irritant Classes as Defined by the EPA Classification System,[1] with Discordant Chemical and Physical Classes Excluded

ICE	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	140	83	116/140	86	70/81	78	46/59	22	13/59	14	11/81
Without Alcohols	128	82	105/128	85	60/71	79	45/57	21	12/57	15	11/71
Without Surfactants	122	82	100/122	84	57/68	80	43/54	20	11/54	16	11/68
Without Solids	107	84	90/107	90	55/61	76	35/46	24	11/46	10	6/61
Without Alcohols and Surfactants	110	81	89/110	81	47/58	81	42/52	19	10/52	19	11/58
Without 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; N = number of substances included in this analysis/the total number of substances in the study; No. = data used to calculate the percentage.

[1] EPA classification system (EPA 2003a): Category IV vs. Category I/II/III.

Table 6-11 Under- and Overprediction of the ICE Test Method Using the EPA Classification System[1] 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*)					
		Severe (Category I)			Moderate (Category II)		Mild (Cat III)	Moderate (Cat II)	Mild (Cat III)		Not Labeled (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[3]	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[3]	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)

continued

Table 6-11 Under- and Overprediction of the ICE Test Method Using the EPA Classification System[1] in Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method by Chemical Class or Physical Property (continued)

Category	N	Underprediction (*In Vivo/In Vitro*)						Overprediction (*In Vivo/In Vitro*)					
		Severe (Category I)			Moderate (Category II)		Mild (Cat III)	Moderate (Cat II)	Mild (Cat III)		Not Labeled (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)
Properties of Interest (continued)													
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% (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; N = number of substances included in this analysis/the total number of substances in the study.

[1] EPA classification system (EPA 2003a).

[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 National Library of Medicine medical subject heading (MeSH) categories (www.nlm.nih.gov/mesh) as defined in Annex I.

[3] Physical form (i.e., solid and liquid) not known for some substances, and therefore the overall number does not equal the sum of the solid and liquid substances.

6.3 EU Classification System: ICE Test Method Accuracy

The four studies (Balls et al. 1995; Prinsen 1996, 2005; Prinsen and Koëter 1993) contained ICE test method data on 174 substances, 153 of which had sufficient *in vivo* data to be assigned an EU ocular irritancy classification, duplicates removed (EU 2001) (see **Annex III**). Based on results from *in vivo* rabbit eye experiments, 21% (32/153[7]) were classified as severe irritants (R41), 18% (28/153) were classified as moderate irritants (R36), and 61% (93/153) were classified as Not Labeled. The remaining 21 substances could not be classified according to the EU classification system due to the lack of adequate animal data and are so noted in **Annex III**.

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

The ICE test method correctly identified 59% (19/32) of the R41 substances (**Table 6-12**). Among the remaining 41% (13/32) R41 substances that were underpredicted by the ICE test method, 22% (7/32) were classified as R36, and 19% (6/32) were classified as Not Labeled.

6.3.2 Identification of R36 Substances (Moderate Ocular Irritants)

Of 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 6-12**).

6.3.3 Identification of Not Labeled Substances

Of 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 6-12**).

Table 6-12 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	Severe (R41)		Moderate (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.

[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 and were not evaluated separately.

[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.

6.3.4 Ability to Distinguish 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 evaluated the ability of the ICE test method to distinguish substances not labeled as irritants from all irritant classes.[8] Using this approach for the 153 substances considered, the ICE test method has an overall accuracy of 85% (130/153), a sensitivity of 78% (47/60), a specificity of 89% (83/93), a false positive rate of 11% (10/93), and a false negative rate of 22% (13/60) (**Table 6-13**).

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 an EU classification. Among 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 6-13**).

Balls et al. (1995): Based upon the *in vivo* rabbit data, 48 substances could be assigned an EU classification. Among 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 6-13**). Six of the 7 substances identified as false negatives were from the discordant classes (alcohol, solids, surfactants).

Prinsen (1996): Based upon the *in vivo* rabbit data, 36 substances could be assigned an EU classification. Among 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 6-13**).

Prinsen (2005): Based upon the *in vivo* rabbit data 46 substances could be assigned an EU classification. Among 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% (3/10) (**Table 6-13**).

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

Table 6-13 Accuracy of the ICE Test Method in Distinguishing Not Labeled Substances from All Other Irritant Classes as Defined by the EU Classification System,[1] by Study and Overall

Data Source	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	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[2]	153	85	130/153	78	47/60	89	83/93	11	10/93	22	13/60

Abbreviations: EU = European Union; ICE = isolated chicken eye; N = number of substances included in this analysis/the total number of substances in the study; No. = data used to calculate the percentage.

[1] EU classification system (EU 2001): Not Labeled vs. R41/R36.

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

6.3.5 Performance of the ICE Test Method with Discordant Classes Excluded

The ICE test method limitations identified in the 2006 ICCVAM BRD were based upon the false positive rate for alcohols and the false negative rates for solids and surfactants when the ICE test method 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 slightly improved (77% versus 80%) when these substances were excluded relative to the performance with the entire database (**Table 6-14**).

When the evaluation was broadened to the ability of the ICE test method to distinguish Not Labeled substances from all other 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 6-15**).

Following the removal of substances belonging to discordant classes (i.e. alcohols, surfactants, and solids, see also ICCVAM [2006a]), there were five EU ocular irritants classified as Not Labeled using the ICE test method (i.e., they were false negatives, see **Table 6-5**). Among the five false negatives for the EU system, 60% (3/5) were EU Category R36 substances, and 40% (2/5) were EU Category R41 substances.

Further analysis of underpredicted (false negative) results by chemical class indicated that onium compounds were the most underpredicted, with 3 of the 20 substances underpredicted. Six *in vivo* severe substances (carboxylic acid, heterocyclic, and an inorganic) were underclassified 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 6-16**).

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 6-16**).

Table 6-14 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 Discordant Chemical and Physical Classes Excluded

ICE	Overall Correct Classification	Severe (R41)		Moderate (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)
Without 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)
Without 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)
Without 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)
Without 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)
Without 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.

[1] EU classification system (EU 2001).

Table 6-15 Accuracy of the ICE Test Method in Distinguishing Not Labeled Substances from All Other Irritant Classes as Defined by the EU Classification System,[1] with Discordant Chemical and Physical Classes Excluded

ICE	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	153	85	130/153	78	47/60	89	83/93	11	10/93	22	13/60
Without Alcohols	139	85	118/139	75	39/52	91	79/87	9	8/87	25	13/52
Without Surfactants	132	85	112/132	78	38/49	89	74/83	11	9/83	22	11/49
Without Solids	116	85	99/116	84	37/44	86	62/72	14	10/72	16	7/44
Without Alcohols and Surfactants	118	85	100/118	73	30/41	91	70/77	9	7/77	27	11/41
Without 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; N = number of substances included in this analysis/the total number of substances in the study; No. = data used to calculate the percentage.

[1] EU classification system (EU 2001): Not Labeled vs. R41/R36.

Table 6-16 Under- and Overprediction of the ICE Test Method Using the EU Classification System[1] 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*)		
		Severe (R41)		Mild/Mod (R36)	Mild/Mod (R36)	Not Labeled (NL)	
		NL	R36	NL	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[3]	112	8% (2/24)	21% (5/24)	23% (5/22)	18% (4/22)	14% 9/66	2% (1/66)
Solids[3]	39	27% (4/15)	13% (2/15)	66% (2/3	0% (0/3)	0% (0/21)	0% (0/21)

continued

Table 6-16 Under- and Overprediction of the ICE Test Method Using the EU Classification System[1] in Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method by Chemical Class or Physical Property (continued)

Category	N	Underprediction (*In Vivo/In Vitro*)			Overprediction (*In Vivo/In Vitro*)		
		Severe (R41)		Mild/Mod (R36)	Mild/Mod (R36)	Not Labeled (NL)	
		NL	R36	NL	R41	R36	R41
Overall	153	18% (6/32)	22% (7/32)	25% (7/28)	18% (5/28)	10% (9/93)	1% (1/93)
Properties of Interest (continued)							
Pesticide	11	20% (1/5)	20% (1/5)	1% (1/1)	-	0% (0/5)	0% (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; N = number of substances included in this analysis/the total number of substances in the study; NL = Not Labeled (as irritant).

[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 National Library of Medicine medical subject heading (MeSH) categories (www.nlm.nih.gov/mesh) as defined in Annex I.

[3] Physical form (i.e., solid and liquid) not known for some substances; therefore, the overall number does not equal the sum of the solid and liquid substances.

6.4 FHSA Classification System: ICE Test Method Accuracy

The four studies (Prinsen and Köeter 1993; Balls et al. 1995; Prinsen 1996; Prinsen 2005) contained ICE test method data on 174 substances, 146 and 138 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the FHSA-20% and FHSA-67% criteria, respectively. Based on results from *in vivo* rabbit eye experiments using the FHSA-20% criteria, 58% (84/146) were classified as irritants and 42% (62/146) were classified as Not Labeled. The remaining 28 substances could not be classified according to the FHSA-20% criteria due to lack of adequate data and are so noted in **Annex III**.

Based on results from *in vivo* rabbit eye experiments using the FHSA-67% criteria, 55% (76/138) were classified as irritants and 45% (62/138) were classified as Not Labeled. The remaining 36 substances could not be classified according to the FHSA-67% criteria due to lack of adequate data and are so noted in **Annex III**.

6.4.1 Ability to Distinguish Not Labeled Substances from Irritants

ICCVAM evaluated the ability of the ICE test method to distinguish substances not labeled as irritants from irritants. Using this approach for the 146 substances classified according to the FHSA-20% criteria, the ICE test method has an overall accuracy of 83% (121/146), a sensitivity of 88% (74/84), a specificity of 76% (47/62), a false positive rate of 24% (15/62), and a false negative rate of 12% (10/84) (**Table 6-17**).

Using this approach for the 138 substances classified according to the FHSA-67% criteria, the ICE test method has an overall accuracy of 84% (116/138), a sensitivity of 91% (69/76), a specificity of 76% (47/62), a false positive rate of 24% (15/62), and a false negative rate of 9% (7/76) (**Table 6-18**).

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

Prinsen and Köeter (1993): Based upon *in vivo* rabbit data using the FHSA-20% criteria (**Table 6-17**), ten substances could be assigned a classification. Among these ten substances, the ICE test method has an accuracy of 80% (8/10), sensitivity of 83% (5/6), specificity of 75% (3/4), a false positive rate of 25% (1/4), and a false negative rate of 17% (1/6).

Based upon *in vivo* rabbit data using the FHSA-67% analysis method (**Table 6-18**), nine substances could be assigned a classification. Among these nine substances, the ICE test method has an accuracy of 89% (8/9), sensitivity of 100% (5/5), specificity of 75% (3/4), a false positive rate of 25% (1/4), and a false negative rate of 0% (0/5).

Balls et al. (1995): Based upon *in vivo* rabbit data using the FHSA-20% criteria (**Table 6-17**), 53 substances could be assigned a classification. Among these 53 substances, the ICE test method has an accuracy of 91% (48/53), sensitivity of 94% (47/50), specificity of 33% (1/3), a false positive rate of 67% (2/3), and a false negative rate of 6% (3/50).

Based upon *in vivo* rabbit data using the FHSA-67% criteria (**Table 6-18**), 48 substances could be assigned a classification. Among these 48 substances, the ICE test method has an accuracy of 90% (43/48), sensitivity of 93% (42/45), specificity of 33% (1/3), a false positive rate of 67% (2/3), and a false negative rate of 7% (3/45).

Prinsen (1996): Based upon *in vivo* rabbit data using the FHSA-20% criteria (**Table 6-17**), 38 substances could be assigned a classification. Among these 38 substances, the ICE test method has an accuracy of 84% (32/38), sensitivity of 77% (10/13), specificity of 88% (22/25), a false positive rate of 12% (3/25), and a false negative rate of 23% (3/13).

Based upon *in vivo* rabbit data using the FHSA-67% criteria (**Table 6-18**), 37 substances could be assigned a classification. Among these 37 substances, the ICE test method has an accuracy of 86%

(32/37), sensitivity of 83% (10/12), specificity of 88% (22/25), a false positive rate of 12% (3/25), and a false negative rate of 17% (2/12).

Prinsen (2005): Based upon *in vivo* rabbit data using the FHSA-20% criteria (**Table 6-17**), 44 substances could be assigned a classification. Among these 44 substances, the ICE test method has an accuracy of 73% (32/44), sensitivity of 79% (11/14), specificity of 70% (21/30), a false positive rate of 30% (9/30), and a false negative rate of 21% (3/14).

Based upon *in vivo* rabbit data using the FHSA-67% criteria (**Table 6-18**), 43 substances could be assigned a classification. Among these 43 substances, the ICE test method has an accuracy of 74% (32/43), sensitivity of 85% (11/13), specificity of 70% (21/30), a false positive rate of 30% (9/30), and a false negative rate of 15% (2/13).

Table 6-17 Accuracy of the ICE Test Method in Distinguishing Not Labeled Substances from Irritants as Defined by the FHSA-20% Criteria,[1] by Study and Overall

Data Source	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	10	80	8/10	83	5/6	75	3/4	25	1/4	17	1/6
Balls et al. (1995)	53	91	48/53	94	47/50	33	1/3	67	2/3	6	3/50
Prinsen (1996)	38	84	32/38	77	10/13	88	22/25	12	3/25	23	3/13
Prinsen (2005)	44	73	32/44	79	11/14	70	21/30	30	9/30	21	3/14
Overall[2]	146	83	121/146	88	74/84	76	47/62	24	15/62	12	10/84

Abbreviations: FHSA = Federal Hazardous Substances Act; ICE = isolated chicken eye; N = number of substances; No. = data used to calculate the percentage.

[1] For the FHSA classification system (FHSA 2005), "proportionality" criteria (i.e., FHSA-20%) were applied for the purpose of assigning an FHSA classification for test results that would require additional testing according to the FHSA sequential testing strategy in order to maximize the number of substances included in these analyses.

[2] Because Prinsen (2000) includes only one test substance that could be classified by FHSA-20%, data from this study were included only in the overall analysis and were not evaluated separately.

Table 6-18 Accuracy of the ICE Test Method in Distinguishing Not Labeled Substances from Irritants as Defined by the FHSA-67% Criteria,[1] by Study and Overall

Data Source	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993)	9	89	8/9	100	5/5	75	3/4	25	1/4	0	0/5
Balls et al. (1995)	48	90	43/48	93	42/45	33	1/3	67	2/3	7	3/45
Prinsen (1996)	37	86	32/37	83	10/12	88	22/25	12	3/25	17	2/12
Prinsen (2005)	43	74	32/43	85	11/13	70	21/30	30	9/30	15	2/13
Overall[2]	138	84	116/138	91	69/76	76	47/62	24	15/62	9	7/76

Abbreviations: FHSA = Federal Hazardous Substances Act; ICE = isolated chicken eye; N = number of substances; No. = data used to calculate the percentage.

[1] For the FHSA classification system (FHSA (2005), "proportionality" criteria (i.e., FHSA-67%) were applied for the purpose of assigning an FHSA classification for test results that would require additional testing according to the FHSA sequential testing strategy in order to maximize the number of substances included in these analyses.

[2] Because Prinsen (2000) includes only one test substance that could be classified by FHSA-67%, data from this study were included only in the overall analysis and were not evaluated separately.

6.4.2 Performance of the ICE Test Method with Discordant Classes Excluded

The previous ICCVAM BRD identified limitations of the ICE test method based upon the false positive rate for alcohols and the false negative rates for solids and surfactants when the ICE test method is used to identify ocular corrosives and severe irritants (ICCVAM 2006a). For this reason, the performance of the ICE test method in identifying FHSA irritants using the FHSA-20% and FHSA-67% criteria was evaluated with these substances excluded from the database. The overall performance statistics using the FHSA-20% criteria (**Table 6-19**) or the FHSA-67% criteria (**Table 6-20**) were not affected by the exclusion of substances belonging to any of the three discordant classes or by any combinations of them**.**

The ability of the ICE test method to distinguish substances not labeled as irritants from irritants as defined by the FHSA-20% criteria was evaluated with the discordant classes removed separately and in combination (**Table 6-19**). The overall accuracy of the ICE database was 83% (121/146) compared to 82% (62/76) with all previously discordant alcohols, surfactants, and solids removed. The overall false negative rate of 12% (10/84) ranged from a low of 8% (5/60) with solids removed to a high of 17% (10/59) with alcohols and surfactants removed. However, the overall false positive rate increased from 24% (47/62) to 27% (13/49) when solids were removed and decreased marginally to 21% (11/53) when alcohols and surfactants were removed.

The ability of the ICE test method to distinguish substances not labeled as irritants from irritants as defined by the FHSA-67% criteria was evaluated with the discordant classes removed separately and in combination (**Table 6-20**). The overall accuracy of the ICE database was 84% (116/138) compared to 82% (59/72) with all previously discordant alcohols, surfactants, and solids removed. The overall false negative rate of 9% (7/76) ranged from a low of 7% (4/54) with solids removed to a high of 13% with alcohols and surfactants removed (10/59) or alcohols, surfactants, and solids (9/40) removed. However, the overall false positive rate increased marginally from 24% (15/62) to 27% (13/49) when solids were removed and decreased slightly to 21% (11/53) when alcohols and surfactants were removed.

Following the removal of substances belonging to the discordant classes (i.e., alcohols, surfactants and solids; see ICCVAM 2006a), there were five FHSA-20% criteria ocular irritants and four FHSA-67% criteria ocular irritants classified as Not Labeled as Irritant by the ICE test method (i.e., false negatives; see **Table 6-5**).

Table 6-19 Accuracy of the ICE Test Method in Distinguishing Not Labeled Substances from Irritants as Defined by the FHSA-20% Criteria,[1] with Discordant Chemical and Physical Classes Excluded

ICE	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	146	83	121/146	88	74/84	76	47/62	24	15/62	12	10/84
Without Alcohols	132	83	109/132	78	64/74	78	45/58	22	13/58	14	10/74
Without Surfactants	124	82	102/124	86	59/69	78	43/55	22	12/55	14	10/69
Without Solids	109	83	91/109	92	55/60	73	36/49	27	13/49	8	5/60
Without Alcohols and Surfactants	112	81	91/112	83	49/59	79	42/53	21	11/53	17	10/59
Without Alcohols, Surfactants, and Solids	76	82	62/76	86	31/36	78	31/40	23	9/40	14	5/36

Abbreviations: FHSA = U.S. Federal Hazardous Substances Act; ICE = isolated chicken eye; N = number of substances included in this analysis/the total number of substances in the study; No. = data used to calculate the percentage.

[1] For the FHSA classification system (FHSA 2005), "proportionality" criteria (i.e., FHSA-20%) were applied for the purpose of assigning an FHSA classification for test results that would require additional testing according to the FHSA sequential testing strategy in order to maximize the number of substances included in these analyses.

Table 6-20 Accuracy of the ICE Test Method in Distinguishing Not Labeled Substances from Irritants as Defined by the FHSA-67% Criteria,[1] with Discordant Chemical and Physical Classes Excluded

ICE	N	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	138	84	116/138	91	69/76	76	47/62	24	15/62	9	7/76
Without Alcohols	124	84	104/124	89	59/66	78	45/58	22	13/58	11	7/66
Without Surfactants	116	84	99/118	89	56/63	78	43/55	22	12/55	11	7/63
Without Solids	103	83	86/103	93	50/54	73	36/49	27	13/49	7	4/54
Without Alcohols and Surfactants	106	83	88/106	87	46/53	79	42/53	21	11/53	13	7/53
Without Alcohols, Surfactants, and Solids	72	82	59/72	88	28/32	78	31/40	23	9/40	13	4/32

Abbreviations: FHSA = U.S. Federal Hazardous Substances Act; ICE = isolated chicken eye; N = number of substances included in this analysis/the total number of substances in the study; No. = data used to calculate the percentage.

[1] For the FHSA classification system (FHSA (2005), "proportionality" criteria (i.e., FHSA-67%) were applied for the purpose of assigning an FHSA classification for test results that would require additional testing according to the FHSA sequential testing strategy in order to maximize the number of substances included in these analyses.

7.0 Isolated Chicken Egg Test Method Reliability

Assessment of test method reliability (intralaboratory repeatability and intra- and interlaboratory reproducibility) is essential to 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). Because 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, ICCVAM conducted additional qualitative analyses of interlaboratory reproducibility to evaluate the extent to which the four laboratories participating in the interlaboratory validation study (Balls et al. 1995) agreed on ICE hazard classifications. As was done for the accuracy evaluation, these qualitative evaluations of reproducibility were conducted based on (1) the use of the ICE test method to identify all ocular hazard categories according to the EPA, GHS, and EU 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 as Irritant) from all other ocular hazard categories (i.e., EPA Categories I, II, and III; EU R41 and R36; GHS Categories 1, 2A, and 2B). Given that the performance of the ICE test method was similar for the EPA and FHSA classification systems, additional reliability analyses were not conducted for the FHSA classification system.

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

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

The four participating laboratories were in 100%, 75%, and 50% agreement on the ocular irritancy classification when distinguishing Not Classified substances from all other classes of 75% (44/59), 14% (8/59), and 12% (7/59), respectively (**Table 7-2**).[9]

All four participating laboratories agreed on the classification of 64% (7/11) of substances that were correctly identified as GHS Category 1,[10] 50% (3/6) of substances correctly classified as GHS Category 2A, 0% (0/2) of substances correctly classified as GHS Category 2B, and 0% (0/1) of substances correctly classified as GHS Not Classified (**Table 7-1**).

Three of the four laboratories were in agreement for 27% (3/11) of the correctly identified GHS Category 1 substances, 0% (0/6) of GHS Category 2A substances, 50% (1/2) of GHS Category 2B substances, and 100% (1/1) of the Not Classified substances **(Table 7-1)**.

---

[9] Because the database of ICE test method results has not changed, the qualitative evaluation of reproducibility presented in ICCVAM (2006a) 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) in Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the GHS Classification System[1]

***In Vivo* Classification (No.)[2]**	**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[3]	7 (64)	3 (27)	1(9)

Abbreviation: GHS = Globally Harmonized System; NC = Not Classified; No. = number of substances included in this analysis/the total number of substances in the study.

[1] GHS classification system (UN 2007); Mild, Moderate, or Corrosive/Severe irritant (2B, 2A, or 1, respectively).

[2] 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*.

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

Table 7-2 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Not Classified or Category 1/2A/2B Using the GHS Classification System[1]

Classification (*In Vivo/In Vitro*)[2]	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[3]	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[3]	44 (75)	8 (14)	7 (12)

Abbreviation: GHS = Globally Harmonized System.

[1] GHS classification system (UN 2007).

[2] A "+" indicates that the substance was assigned an overall classification of Mild, Moderate, or Corrosive/Severe irritant (2B, 2A, or 1, respectively). A "-" indicates that the substance was assigned a classification of Not Classified. 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.*

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

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

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

All four participating laboratories (100%) incorrectly classified 82% (9/11) of the GHS Category 1 substances, 13% (1/8) of the GHS Category 2A substances, 0% (0/2) of the GHS Category 2B substances, and 54% (7/13) of the GHS Not Classified substances (**Table 7-1**).

Three of the four laboratories (75%) were in agreement in incorrectly classifying 18% (2/11) of the GHS Category 1 substances, 0% (0/8) of the GHS Category 2A substances, 100% (2/2) of Category 2B substances, and 31% (4/13) of the GHS Not Classified substances (**Table 7-1**).

Two of the four laboratories (50%) were in agreement in incorrectly classifying 0% (0/11) of the GHS Category 1 substances, 88% (7/8) of the GHS Category 2A substances, 0% (0/2) of the GHS Category 2B substances, and 15% (2/13) of the GHS Not Classified substances (**Table 7-1**).

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

Of two substances classified by the EPA as Category IV, 0% (0/2) were correctly identified, while 40% (8/20) EPA Category III substances were correctly identified, 50% (5/10) of the EPA Category

II substances were correctly identified, and 53% (10/19) of the EPA Category I substances were correctly identified (**Table 7-3).**

The four participating laboratories were in 100%, 75%, and 50% agreement in regard to the ocular irritancy classification when distinguishing Category IV substances from all other classes of 75% (44/59), 14% (8/59), and 12% (7/59), respectively (**Table 7-4**).[11]

All four participating laboratories (100%) agreed on the classification of 70% (7/10) of substances that were correctly identified as EPA Category I,[12] 60% (3/5) of substances correctly classified as EPA Category II, 13% (1/8) of substances correctly classified as EPA Category III, and 0 substances classified as Category IV (**Table 7-3**).

Three of the four laboratories (75%) were in agreement for 20% (2/10) of the correctly identified EPA Category I substances, 20% (1/5) of the EPA Category II substances, 38% (3/8) of the EPA Category III substances, and 0 of the substances classified as Category IV (**Table 7-3).** The discordant laboratory was not consistent among these substances.

Two of the four laboratories (50%) were in agreement for 10% (1/10) of the EPA Category I substances identified correctly, 20% (1/5) of the EPA Category II substances, 50% (4/8) of the EPA Category III substances correctly identified, and 0 of the substances classified as Category IV (**Table 7-3**).

Of two substances classified by the EPA as Category IV, 100% (2/2) were incorrectly identified, while 60% (12/20) of substances classified as EPA Category III were incorrectly identified, 50% (5/10) of EPA Category II substances were incorrectly identified, and 47% (9/19) of EPA Category I substances were incorrectly identified (**Table 7-3**).

The four participating laboratories (100%) were in 100% agreement in incorrectly classifying 78% (7/9) of the EPA Category I substances, 20% (1/5) of the EPA Category II substances, 50% (6/12) of the EPA Category III substances, and 0% (0/2) of the EPA Category IV substances (**Table 7-3**).

Three of the four laboratories (75%) were in agreement in incorrectly classifying 22% (2/9) of the EPA Category I substances, 20% (1/5) of the EPA Category II substances, 33% (4/12) of the Category III substances, and 100% (2/2) of the EPA Category IV substances (**Table 7-3).** 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% (0/9) of the EPA Category I substances, 60% (3/5) of the EPA Category II substances, 17% (2/12) of the EPA Category III substances, and 0% (0/2) of the EPA Category IV substances (**Table 7-3**).

---

[11] Because the database of ICE test method results has not changed, the qualitative evaluation of reproducibility presented in ICCVAM (2006a) 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 by one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

Table 7-3 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[1]

In vivo Classification (No.)[2]	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 (10)	Under	2 (20)	4	0	1 (50)	1 (50)
	Actual	5 (50)	4	3 (60)	1 (20)	1 (20)
	Over	3 (30)	4	1 (33)	0	2 (67)
1 (19)	Under	9 (47)	4	7 (78)	2 (22)	0
	Actual	10 (53)	4[3]	7 (70)	2 (20)	1 (10)

Abbreviation: EPA = U.S. Environmental Protection Agency; No. = number of substances included in this analysis/the total number of substances in the study

[1] EPA classification system (EPA 2003a).

[2] 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.*

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

Table 7-4 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Category IV or Category I/ II/III Using the EPA Classification System[1]

Classification (*In vivo/In vitro*)[2]	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[3]	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[3]	44 (75)	8 (14)	7 (12)

Abbreviation: EPA = U.S. Environmental Protection Agency

[1] EPA classification system (2003a).

[2] A "+" indicates that the substance was assigned an overall classification of Severe, Moderate, or Mild irritant (I, II, or III, respectively). A "-" indicates that the substance was assigned a classification of not classified as an irritant (Category IV). A "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), an EPA classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro.*

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

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

Of 17 substances classified by the EU as Not Labeled, 53% (9/17) were correctly identified, while 50% (7/14) of substances classified as EU moderate irritants (R36) were correctly identified, and 53% (10/19) substances classified by the EU as corrosive/severe irritants (R41) were correctly identified (**Table 7-5**).

The four participating laboratories were in 100%, 75%, and 50% agreement in regard to the ocular irritancy classification when distinguishing Not Labeled substances from all other classes of 61% (36/59), 25% (15/59), and 14% (8/59), respectively (**Table 7-6**).[13]

All four participating laboratories (100%) agreed on the classification of 70% (7/10) of the substances that were correctly identified as R41, 57% (4/7) of substances correctly classified as EU R36, and 33% (3/9) of those correctly classified as EU Not Labeled (**Table 7-5**).

Three of the four laboratories (75%) were in agreement on 20% (2/10) of the correctly identified R41 substances, 29% (2/7) of the R36 substances, and 44% (4/9) of the substances classified as EU Not Labeled (**Table 7-5**). The discordant laboratory was not consistent among these substances.

Two of the four laboratories (50%) were in agreement for 10% (1/10) of the R41 substances correctly identified, 14% (1/7) of the R36 substances, and 22% (2/9) of the substances classified as EU Not Labeled (**Table 7-5**).

[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.

Of 17 substances classified by the EU as Not Labeled, 47% (8/17) were incorrectly identified, while 50% (7/14) of substances classified as R36 substances were incorrectly identified, and 47% (9/19) of substances classified as R41 were incorrectly identified (**Table 7-5**).

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

Three of the four laboratories (75%) were in agreement in incorrectly classifying 22% (2/9) of the R41 substances, 29% (2/7) of the R36 substances, and 13% (1/8) of the EU Not Labeled substances (**Table 7-5**).

Two of the four laboratories (50%) were in agreement in incorrectly classifying 0% (0/9) of the R41 substances, 57% (4/7) of the R36 substances, and 25% (2/8) of the EU Not Labeled substances (**Table 7-5**).

Table 7-5 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[1]

In vivo Classification (No.)[2]	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[3]	7 (70)	2 (20)	1 (10)

Abbreviation: EU = European Union; NL = Not Labeled (as an irritant); No. = number of substances included in this analysis

[1] EU classification system (2001).

[2] 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.*

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

Table 7-6 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Not Labeled or R36/R41 Using the EU Classification System[1]

Classification (*In vivo/In vitro*)[2]	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[3]	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[3]	36 (61)	15 (25)	8 (14)

Abbreviation: EU = European Union.

[1] EU classification system (2001).

[2] A "+" indicates that the substance was assigned an overall classification of Severe or Nonsevere irritant (Category R41 or R36). A "-" indicates that the substance was assigned a classification of Not Labeled (as an irritant) (Category NL). 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.*

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

8.0 Isolated Chicken Egg 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 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).

9.0 Other Scientific Reports and Reviews

No new data, nor published or unpublished studies, have been located since the previous evaluation of the ICE test method for identification of 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 non-animal procedures (e.g., *in vitro* technologies), where possible (Russell and Burch 1992)

The ICE test method refines animal use. Because these animals are being humanely killed for nonlaboratory purposes, the testing procedure inflicts no additional pain or distress on animals. 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, which replaces laboratory animals with animal species routinely raised in large numbers as a food source. Additionally, with the ability to identify ocular corrosives and severe 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 *in vitro* test method, only chickens humanely killed for food or other nonlaboratory purposes are used as eye donors (i.e., no live animals are used in this test method).

11.0 Practical Considerations

Practical considerations for the ICE test method are 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).

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

EPA. 2003a. Label Review Manual: 3rd ed. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.

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

EPA. 2003c. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR 160.

EU. 1992. Commission Directive 92/69/EEC of 31 July 1992 adapting to technical progress for the seventeenth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L383:113-115.

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 twenty-ninth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Union L152:1-311.

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

Federal Hazardous Substances Act (FHSA). 2005. Public Law 86-613.

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. Research Triangle Park, NC:National Institute of Environmental Health Sciences.

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

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, NC:National Institute of Environmental Health Sciences.

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, NC:National Institute of Environmental Health Sciences.

ICCVAM. 2009. Independent Scientific Peer Review Panel Report: Evaluation of the Validation Status of Alternative Ocular Safety Testing Methods and Approaches. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

ICCVAM Authorization Act of 2000. 2000. Public Law 106-545. 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. Paris:OECD.

OECD. 2002. Test guideline 405. Acute eye irritation/corrosion, adopted April 24, 2002. In: OECD Guidelines for Testing of Chemicals. Paris:OECD.

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

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

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

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

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

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

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

United Nations. 2007. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). 2d rev ed. New York:UN. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev02/02files_e.html.

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:[15] The experimental system used. Often used interchangeably with *test* and *test method.*

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.

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

Blepharitis: Inflammation of the eyelids.

Bulbar conjunctiva: The portion of the conjunctiva that covers the outer surface of the eye.

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

Chemosis: A form of eye irritation in which the membranes that line the eyelids and surface of the eye (*conjunctiva*) become swollen.

Classification system: An arrangement of quantified results or data into groups or categories according to previously established criteria.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

Coefficient of variation: A statistical representation of the precision of a test. It is expressed as a percentage and is calculated as follows:

$$\left(\frac{standard\ \ deviation}{mean}\right) \times 100\%$$

[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 2003 Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods (NIH Publication No. 03-4508).

Concordance:[15] 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:[15] 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:[15] A substance incorrectly identified as negative by a test method.

False negative rate:[15] 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:[15] A substance incorrectly identified as positive by a test method.

False positive rate:[15] 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):[15] 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:[15] 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:[15] 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:[15] 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:[15] 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 nonreactive material, such as water. This sample is processed with test substance-treated samples and other control samples to determine whether the solvent interacts with the test system.

Negative predictivity:[15] 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; EU R36.

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:[15] 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:[15] 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:[15] The proportion of positives in the population of substances tested (see *two-by-two* table).

Protocol:[15] 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:[15] 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:[15] A new or modified test method that reduces the number of animals required.

Reference test method:[15] 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:[15] A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal wellbeing.

Relevance:[15] 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:[15] 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 interlaboratory reproducibility and intralaboratory repeatability.

Replacement alternative:[15] 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:[15] 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 interlaboratory 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:[15] 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) A substance 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:[15] 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:[15] The experimental system used; used interchangeably with *test method* and *assay*.

Test method:[15] 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:[15] The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two table:[15] 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:[15] 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:[15] The process by which the reliability and relevance of a procedure are established for a specific purpose.

Vascular tunic: The middle of three membranes of the eye, comprising the iris, ciliary body, and choroid. Also referred to as the *uvea*.

Weight of evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.