Draft Background Review Document

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

Bovine Corneal Opacity and Permeability Test Method

Interagency Coordinating Committee on the Validation of Alternative Methods

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

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

March 2009
NIH Publication No.

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

[This Page Intentionally Left Blank]

Table of Contents

List of Tables

Page Number

206 # List of Figures

208

List of Acronyms and Abbreviations

BCOP	Bovine Corneal Opacity and Permeability
BRD	Background Review Document
CASRN	*Chemical Abstracts Service Registry Number*
CPSC	(U.S.) Consumer Product Safety Commission
CV	Coefficient of variation
EC	*European Commission*
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EC/HO	European Commission/British Home Office
ECVAM	European Centre for the Validation of Alternative Methods
EEC	European Economic Community
EPA	(U.S.) Environmental Protection Agency
EU	European Union
FDA	(U.S.) Food and Drug Administration
FR	*Federal Register*
g	Gram
GHS	Globally Harmonized System (of Classification and Labeling of Chemicals)
GLP	Good Laboratory Practices
HET-CAM	Hen's Egg Test – Chorioallantoic Membrane
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye
IIVS	Institute for In Vitro Sciences
IRE	Isolated Rabbit Eye
IVIS	*In Vitro* Irritancy Score
μg	Microgram
μL	Microliter
μm	Micrometer
MAS	Maximum average score
MeSH	(National Library of Medicine) Medical subject heading

mL	Milliliter
MMAS	Modified maximum average score
NA	Not applicable
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NS	Not specified
NTP	(U.S.) National Toxicology Program
OD	Optical density
OECD	Organisation for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances
OSHA	Occupational Safety and Health Administration
OTWG	Ocular Toxicity Working Group
P.L.	Public Law
r	rho (correlation coefficient)
SD	Standard deviation
TG	Test Guideline
UN	United Nations

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

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

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

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

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

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

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

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

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

•Principal agency representative
◊Alternate principal agency representative

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

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

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

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

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

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

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

347 ***[This Page Intentionally Left Blank]***

348

Acknowledgements

Interagency Coordinating Committee on the Validation of Alternative Methods Ocular Toxicity Working Group

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

Department of Defense
Harry Salem, Ph.D.

Department of Transportation
Steve Hwang, Ph.D.

U.S. Environmental Protection Agency
Office of Pesticide Programs
Meta Bonner, Ph.D.
Jonathan Chen, Ph.D.
Masih Hashim, D.V.M., Ph.D.
Karen Hicks
Marianne Lewis
Deborah McCall
Timothy McMahon, Ph.D.
Mark Perry, Ph.D.
John Redden, Ph.D.
Amy Rispin, Ph.D.
Jenny Tao, Ph.D.
Office of Research and Development
Andrew Geller, Ph.D.
Office of Science Coordination and Policy
Karen Hamernik, Ph.D. (OTWG Co-Chair)

U.S. Food and Drug Administration
Center for Drug Evaluation and Research
Paul C. Brown, Ph.D.
Abigail Jacobs, Ph.D. (IWG Co-Chair)
Jill Merrill, Ph.D. (Co-Chair)
Center for Food Science and Nutrition
Robert Bronaugh, Ph.D.
Donnie Lowther
Office of Science and Health Coordination
Suzanne Fitzpatrick, Ph.D., D.A.B.T.

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

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

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

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

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

National Institute of Environmental Health Sciences

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

Deborah McCarley
Special Assistant; Assistant Project Officer

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

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

Jonathan Hamm, Ph.D.
Senior Toxicologist

Nelson Johnson
Senior Project Coordinator/Technical Writer

Elizabeth Lipscomb, Ph.D.
Staff Toxicologist

Linda Litchfield
Meeting Coordinator/Admin. Asst.

Greg Moyer, M.B.A.
Project Manager

Catherine Sprankle
Senior Communications Specialist

James Truax
Senior Project Coordinator/Technical Writer

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

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

Bob Guest
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 BCOP Background Review Document

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

Access Business Group
Luann Potts
Tom Truszkowski

Cosmetics, Toiletry, and Fragrance Association
Carol Eisenmann, Ph.D.

ECVAM
Chantra Eskes, Ph.D.

ExxonMobil Biomedical Sciences, Inc.
James Freeman, Ph.D.

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

JohnsonDiversey, Inc.
John Hamilton, Ph.D.
Sarah Willems, B.S.

Johnson & Johnson Pharmaceutical Research and Development – A Division of Janssen Pharmaceutica N.V.
Freddy Van Goethem, Ph.D.
Philippe Vanparys, Ph.D.

L'OREAL
Christine Van den Berghe, Ph.D.

Merck
Joseph Sina, Ph.D.

S.C. Johnson & Son, Inc./JohnsonDiversy, Inc.
Nicole Cuellar, M.S.
Judith Swanson, B.S./B.A.

491 ***[This Page Intentionally Left Blank]***

492

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 Health and Safety Administration have testing requirements and guidelines for assessing the ocular irritation potential of substances such as pesticides, household products, pharmaceuticals, cosmetics, and agricultural and industrial chemicals.

Although ocular safety assessment has clearly helped to protect consumers and workers, concerns have been raised about the humane aspects of the Draize rabbit eye test (Draize et al. 1944). Regulatory authorities have adopted various modifications that reduce the number of animals used and the potential pain and distress associated with the procedure. Significant progress has been made during the last decade. Now only one to three rabbits are required per test, compared to six rabbits in the original protocol. Provisions have been added that allow for animals with severe lesions or discomfort to be humanely euthanized.

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

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

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

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

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

We gratefully acknowledge the organizations and scientists who provided data and information for this document. We also acknowledge the efforts of those individuals contributing to the preparation of this summary review document, including the following staff from the NICEATM Support Contractor, Integrated Laboratory Systems, Inc.: David Allen, Jon Hamm, Nelson Johnson, Elizabeth Lipscomb, Linda Litchfield, Gregory Moyer, Catherine Sprankle, and Jim Truax. We also thank the members of the ICCVAM Ocular Toxicity Working Group, chaired by Karen Hamernik, Ph.D. (EPA) and Jill Merrill, Ph.D. (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., D.A.C.L.A.M.
Rear Admiral, U.S. Public Heath Service
Director, NICEATM
Executive Director, ICCVAM
March 2009

Executive Summary

Background

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

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

BCOP Test Method Protocol

The BCOP assay is an *in vitro* eye irritation test method using isolated bovine eyes procured from cattle slaughtered for meat and/or other purposes. In the BCOP assay, opacity is determined by the amount of light transmitted through the cornea, and permeability is determined by the amount of sodium fluorescein dye that passes through all corneal cell layers. Both measurements are used to calculate an *In Vitro* Irritancy Score, which is used to assign an *in vitro* irritancy classification for prediction of the *in vivo* ocular irritation potential of a test substance. The BCOP test method is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the bovine cornea *in vitro*. In this test method, damage caused by the test substance is assessed by quantitative measurements of changes in corneal opacity and permeability with an opacitometer and a visible light spectrophotometer, respectively.

Validation Database

An online literature search conducted in support of the evaluation of the validation status of the BCOP test method for its ability to identify ocular corrosives and severe irritants identified four publications containing BCOP test method results. However none of these publications included raw data or reference *in vivo* data, or they included data cited from earlier studies that were already included in the validation database; as such these were not added to the database. The results from BCOP tests for 66 antimicrobial cleaning products (AMCPs) were obtained from a submission to ICCVAM that describes a non-animal approach for evaluating eye irritation potential and labeling requirements for AMCPs. Therefore, the previous validation database for the BCOP test method (ICCVAM, 2006a) was updated to include BCOP test results for the 66 AMCPs. The updated BCOP validation database contains a total of 211 substances, including 135 commercial products or formulations. A variety of chemical and product classes have been tested in the BCOP assay. The chemical classes with the greatest amount of *in vitro* BCOP data are alcohols, carboxylic acids, esters, formulations, heterocyclic compounds, hydrocarbons, ketones, and onium compounds. The formulations tested include hair shampoos, personal care cleansers, detergents, bleaches, insect repellents, petroleum products and fabric softener. Other chemical classes tested include amines, ethers/polyethers, inorganic and organic salts and

organic sulfur compounds. The most common product classes tested in the BCOP assay are chemical/synthetic intermediates, cleaners, drugs/pharmaceuticals/therapeutic agents, petroleum products, solvents, shampoos and surfactants. Other product classes tested include pesticides, plasticizers, reagents, and bactericides.

Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal at 24, 48, and 72 hours and/or assessment of the presence or absence of lesions at 7, 14, and 21 days was necessary to calculate the appropriate EPA (1996), EU (2001) and GHS (UN 2003) ocular irritancy hazard classification. Thus, some of the test substances for which there was only limited *in vivo* data could not be used for evaluating test method accuracy and reliability.

BCOP Test Method Accuracy

Identification of All Ocular Hazard Categories

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

As indicated in **Table 1**, overall correct classifications ranged from 49% (91/187) to 54% (101/186), depending on the hazard classification system evaluated when using the entire database; and 47% (31/66) to 54% (35/65) depending on the hazard classification system evaluated when discordant classes are removed. Using alternative decision criteria for the identification of corrosive/severe ocular irritants (i.e., IVIS $\geq$ 75 [used in the AMCP submission protocol] instead of IVIS $\geq$55.1 [as per the ICCVAM recommended BCOP protocol]) does not improve test method performance.

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

Severe using ≥55.1											
	Overall Correct Classification	**Severe[2]**		**Moderate[3]**			**Mild[4]**			**Not Labeled[5]**	
		actual	**under**	**over**	**actual**	**under**	**over**	**actual**	**under**	**over**	**actual**
GHS	49% (91/187)	85% (55/65)	15% (10/65)	62% (16/26)	27% (7/26)	11% (3/26)	67% (4/6)	33% (2/6)	0% (0/6)	70% (63/90)	30% (27/90)
EPA	55% (101/187)	84% (53/63)	16% (10/63)	50% (11/22)	32% (7/22)	18% (4/22)	50% (28/57)	36% (21/57)	14% (8/57)	53% (24/45)	47% (21/45)
EU	50% (59/118)	79% (26/33)	21% (7/33)	48% (10/21)	52% (11/21)	0% (0/21)	NA	NA	NA	66% (42/64)	34% (22/64)
Severe using ≥75											
		Severe		**Moderate**			**Mild**			**Not Labeled**	
		actual	**under**	**over**	**actual**	**under**	**over**	**actual**	**under**	**over**	**actual**
GHS	50% (94/187)	78% (51/65)	22% (14/65)	31% (8/26)	54% (14/26)	15% (4/26)	67% (4/6)	33% (2/6)	0% (0/6)	70% (63/90)	30% (27/90)
EPA	49% (92/187)	78% (49/63)	22% (14/63)	36% (8/22)	45% (10/22)	19% (4/22)	47% (27/57)	39% (22/57)	14% (8/57)	53% (24/45)	47% (21/45)
EU	51% (60/118)	73% (24/33)	27% (9/33)	29% (6/21)	67% (14/21)	4% (1/21)	NA	NA	NA	66% (42/64)	34% (22/64)

Abbreviations: GHS = Globally Harmonized System; EPA = U.S. Environmental Protection Agency; EU = European Union; BCOP= Bovine Corneal Opacity and Permeability; NA = Not Applicable

[1]GHS classification system (UN 2003), EPA classification system (EPA 1996); EU classification system (EU 2001)

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

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

[4]Mild = GHS Category 2B; EPA Category III; EU R36.

[5]Not Labeled = Not Classified.

675 ***Distinguishing Substances Not Labeled as Irritants from All Other Hazard Categories***

676 The ability of the BCOP test method to distinguish substances not labeled as irritants (i.e.,
677 EPA Category IV, EU Not Labeled, GHS Not Classified) from all other ocular hazard
678 categories (i.e., EPA Category I, II, III; EU R41, R36; GHS Category 1, 2A, 2B), as defined
679 by the GHS, EPA and EU classification systems (EPA 1996; EU 2001; UN 2003) was also
680 evaluated. Again, this analysis was performed with specific chemical classes and/or physical
681 properties excluded based on them previously being identified as discordant in BCOP
682 (ICCVAM 2006a).

683 As indicated in **Table 2**, overall accuracy for the identification of substances not labeled as
684 irritants (i.e., EPA Category IV, EU Not Labeled, GHS Not Classified) from all other
685 categories ranged from 64% (76/118) to 85% (103/121), depending on the hazard
686 classification system used. While false positive rates were high (53% [24/45] to 70% [63/90]
687 depending on the hazard classification system used), the false negative rates were low (6%
688 [8/141] and 0% [0/54 or 0/97] for the EU and GHS systems, respectively). Among the eight
689 false negatives identified for the EPA system, 100% (8/8) were EPA Category III substances
690 based on Draize data. For 38% (3/8) of these substances, the categorization was based on at
691 least one rabbit with a corneal opacity score of one that was not resolved until day three of
692 the study. An additional substance was categorized based on all six rabbits with a
693 conjunctival redness score of three that was not resolved until day seven of the study.
694 Considering the severity and number of ocular lesions noted *in vivo*, these false negative
695 results cannot be minimized as they present a significant risk to a user potentially exposed to
696 these types of materials.

Table 2 Accuracy of the BCOP Test Method for Distinguishing Not Classified Substances from All Other Irritant Classes

Hazard Classification System	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
EPA	121	85	103/121	93	87/94	59	16/27	41	11/27	7	7/94
GHS	118	64	76/118	100	54/54	34	22/64	66	42/64	0	0/54
GHS	122	68	83/122	100	61/61	36	22/61	64	39/61	0	0/61

[1]GHS = Globally Harmonized System (UN 2003). NC vs. Cat 1/2A/2B.
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]No. = Data used to calculate the percentage

The accuracy analysis also indicated that hydrocarbons are often overpredicted (56% to 73% [6/11 to 8/11], depending on the classification system used) in the BCOP test method. Alcohols (19% to 53% [3/16 to 9/17], ketones (33% to 56% [3/9 to 5/9]), carboxylic acids (31% to 43% [4/13 to 6/14]), esters (40% to 50% [4/10 to 5/10]), ethers (17% to 50% [1/6 to 3/6]) and heterocyclic compounds (8% to 31%[1/2 to 4/13]) also had high rates of over prediction. Although there were a small number of underpredicted substances (7 to 11), alcohols were generally underpredicted by all hazard classification systems using the BCOP test method. Furthermore, carboxylic acids (2), esters (2) and heterocyclic compounds were underpredicted in one hazard classification system employed.

BCOP Test Method Reliability

Interlaboratory Reproducibility

Quantitative and qualitative evaluations of BCOP test method reliability have been conducted previously (ICCVAM, 2006a). However, additional qualitative analyses of interlaboratory reproducibility were conducted to evaluate the extent of agreement of BCOP hazard classifications among the participating laboratories from the three different interlaboratory validation studies (Balls et al. 1995, Gautheron et al. 1994, and Southee 1998. As was done for the accuracy evaluation, these qualitative evaluations of reproducibility were based on 1) the use of the BCOP test method for identifying all ocular hazard categories according to the EPA, EU or GHS systems, and 2) the use of the BCOP test method to distinguish substances not labeled as irritants (i.e., EPA Category IV, EU Not Labeled, GHS Not Classified) from all other ocular hazard categories (i.e., EPA Category I, II, III; EU R41, R36; GHS Category 1, 2A, 2B).

Using the first approach (i.e., identifying all ocular hazard categories) among the three interlaboratory studies for the Balls et al. (1995) study, there was 100% agreement among the five laboratories for most of the Draize ocular corrosives/severe irritants based on all three classification systems, whether they were correctly identified or underclassified by the BCOP test method (e.g., for the GHS system, there was 100% agreement for 76% [13/17] of the correctly identified Category I substances). There was also 100% agreement among the five laboratories for all of the overpredicted Not Labeled substances and for at least 50% (2/4) of the correctly identified substances.

For the Gautheron et al. (1994) study, there was 100% agreement among the eleven laboratories for most of the Draize ocular corrosives/severe irritants based on all three classification systems, whether they were correctly identified or underclassified by the BCOP test method (e.g., for the GHS system, there was 100% agreement for 67% [4/6] of the correctly identified Category I substances). There was also 100% agreement among the eleven laboratories for most of the overpredicted Not Labeled substances (e.g., 100% agreement for 54% [7/13] of the correctly identified Not Labeled substances) and for most of the incorrectly identified Not Labeled substances (e.g., 100% agreement for 81% [17/21] of the correctly identified substances.

For the Southee (1998) study, there was 100% agreement among the three laboratories for all of the corrosive/severe irritant substances based on all three classification systems, whether they were correctly identified or underclassified by the BCOP test method (e.g., 100% agreement for 100% [4/4] of the Draize ocular corrosives/severe irritants). There was also 100% agreement among the two correctly identified Not Labeled substances.

Using the second approach (i.e., distinguishing Not Labeled substances from all other ocular hazard categories) for the Balls et al. (1995) study, there was 100% agreement among the multiple laboratories for most of the correctly identified ocular irritant classes (e.g., 100% agreement for 97% [37/38] of the correctly identified GHS Category 1 substances), and for half of the correctly identified Not Labeled substances (50% [2/4]).

For the Gautheron et al. (1994) study, there was 100% agreement among the multiple laboratories for most of the correctly identified ocular irritant classes (e.g., 100% agreement for 92% [11/12] of the correctly identified GHS Category 1 substances), and approximately half of the correctly identified Not Labeled substances (e.g., 100% agreement for 54% [7/12] of the GHS Not Labeled substances).

For the Southee (1998) study, there was 100% agreement among the multiple laboratories for all of the correctly identified ocular irritant classes (e.g., 100% agreement for 100% [10/10] of the correctly identified GHS Category 1 substances), and all of the correctly identified Not Labeled substances (e.g., 100% agreement for 100% [2/2] of the GHS Not Labeled substances).

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

766 ***[This Page Intentionally Left Blank]***
767

1.0 Introduction

1.1 Background

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

In 2006, ICCVAM completed an evaluation of the Bovine Corneal Opacity and Permeability (BCOP) test method for its ability to identify ocular corrosives and severe irritants (ICCVAM, 2006a). Following this review, ICCVAM concluded that the BCOP test method could be used, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) (ICCVAM, 2006b). While it was not considered valid as a complete replacement for the *in vivo* rabbit eye test, the BCOP 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 testing positive in this assay can be classified as ocular corrosives or severe irritants without further testing in rabbits, while a substance that tests negative would need additional testing in rabbits using a sequential testing strategy, as outlined in OECD Test Guideline 405 (OECD 2002).

ICCVAM is now conducting an evaluation to further characterize the usefulness and limitations of the BCOP test method for identifying non-severe irritants (i.e., 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, GHS Not Classified). As part of the evaluation process, this Background Review Document (BRD) has been prepared to describe the current

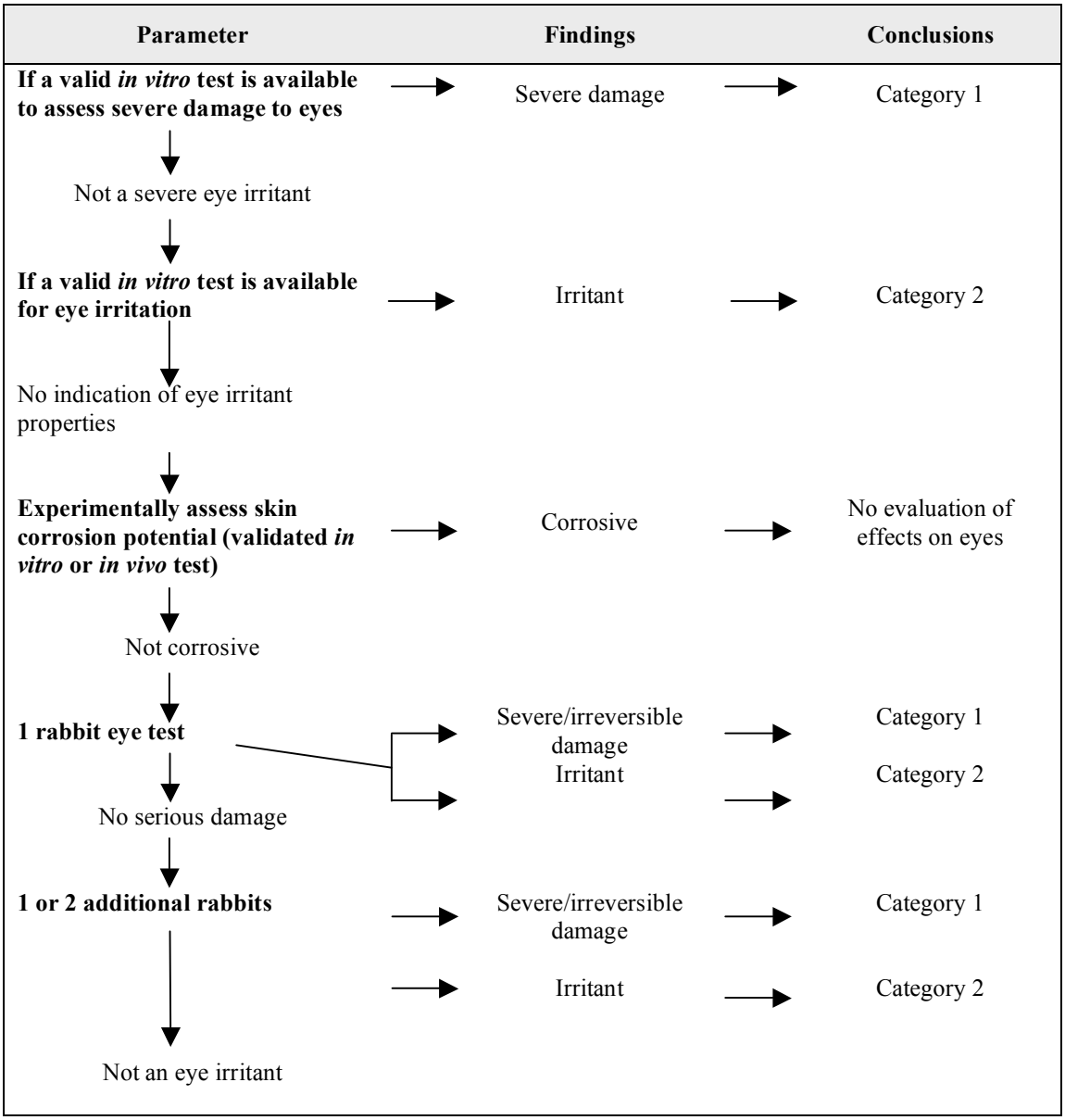
validation status of the BCOP test method, including what is known about its reliability and accuracy, its applicability domain, the numbers and types of substances tested and the availability of a standardized protocol. This BRD was prepared for use by an ICCVAM expert panel review of BCOP as a method to identify all ocular hazard categories. Parallel reviews of the IRE, BCOP, and ICE test methods are being conducted. Results of the Expert Panel Report, combined with the analyses presented in the BRDs, will be used to support ICCVAM recommendations on the proposed standardized test method protocols, proposed list of recommended reference substances, and additional optimization and/or validation studies that may be necessary to further develop and characterize the usefulness and limitations of these methods.

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

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

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

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



823 Adapted from UN (2003).

824

825 ## 1.3 Validation of the BCOP Test Method

826 The ICCVAM Authorization Act (Sec. 4(c)) mandates that "[e]ach Federal Agency … shall
827 ensure that any new or revised … test method … is determined to be valid for its proposed
828 use prior to requiring, recommending, or encouraging [its use]." (Public Law [P.L.] 106-
829 545).

Validation is the process by which the reliability and relevance of an assay for a specific purpose are established (ICCVAM 2003). Relevance is defined as the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 2003). For the BCOP test method described in the BCOP BRD (ICCVAM, 2006a), relevance is restricted to how well the test method identifies substances that are capable of producing corrosive or severe irritant effects to the eye. For the current BRD, relevance is based on how well the test method identifies substances that are capable of producing nonsevere ocular irritation or substances not labeled as irritants. Reliability is defined as the reproducibility of a test method within and among laboratories and should be based on performance with a diverse set of substances that are representative of the types of chemical and product classes that are expected to be tested and cover the range of responses that need to be identified. The validation process will provide data and information that will allow U.S. Federal agencies to develop guidance on the development and use of the BCOP test method as part of a tiered-testing approach to evaluating the eye irritation potential of substances.

The first stage in this evaluation process is the preparation of a BRD that presents and discusses the relevant data and information about the assay, including its mechanistic basis, proposed uses, reliability, and performance characteristics (ICCVAM 2003). This BRD summarizes the available information on the BCOP test method. Where adequate data is available, the qualitative and quantitative performance of the assay is evaluated.

1.4 Search Strategies and Selection of Citations for the BCOP BRD

The BCOP test method data summarized in this BRD are based on information found in the peer-reviewed scientific literature as detailed in the Background Review Document, Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability Test Method (ICCVAM, 2006a). NICEATM is currently evaluating a non-animal assessment approach for evaluating eye irritation potential and labeling requirements for antimicrobial cleaning products (AMCPs). Three *in vitro* test methods are proposed in the testing strategy including the bovine corneal opacity and permeability test method. The final AMCP BRD was provided to NICEATM by IIVS on July 21, 2008. The substances within the AMCP validation database tested in the BCOP test method have been added to the validation database of the BCOP BRD (ICCVAM, 2006a). A subsequent literature

search conducted in January 2009 revealed no new articles containing results utilizing the BCOP test method.

863 ***[This Page Intentionally Left Blank]***

864

865

866

2.0 BCOP Test Method Protocol Components

2.1 Overview of How the BCOP Test Method is Conducted

The BCOP test method is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the bovine cornea *in vitro*. In this test method, damage by the test substance is assessed by quantitative measurements of changes in corneal opacity and permeability with an opacitometer and a visible light spectrophotometer, respectively. Both measurements are used to calculate an IVIS, which is used to assign an *in vitro* irritancy hazard classification category for prediction of the *in vivo* ocular irritation potential of a test substance.

For a detailed description of how the BCOP test method is conducted, see ICCVAM (2006a). Briefly, isolated corneas are obtained from the eyes of freshly slaughtered cattle. Test substances are applied to the epithelial surface of the cornea using different treatment methods depending on the physical nature and chemical characteristics (e.g., solids, semi-solids [including creams and waxes], liquids, viscous [including gels] vs. non-viscous liquids) of the test substance. Liquids are tested undiluted, while surfactants are tested at a concentration of 10% in a 0.9% sodium chloride solution, distilled water, or other solvent that has been demonstrated to have no adverse effects on the test system. Corneas are exposed to liquids and surfactants for 10 minutes. Non-surfactant solids are typically tested as solutions or suspensions at 20% concentration in a 0.9% sodium chloride solution, distilled water, or other solvent that has been demonstrated to have no adverse effects on the test system. Solids may also be tested neat by direct application onto the corneal surface. Corneas are exposed to solids for four hours.

Corneal opacity is measured quantitatively as the amount of light passing through the cornea, resulting in opacity values measured on a continuous scale. Permeability is measured quantitatively as the amount of sodium fluorescein dye that passes across the full thickness of the cornea, as detected in the medium in the posterior chamber. The mean opacity and mean permeability (OD490) values for each treatment group were then used to calculate an *in vitro* score for each treatment group:

***In Vitro* Irritancy Score = mean opacity value + (15 x mean OD490 value)**

The *in vitro* irritation classification schemes used for this evaluation were based on two different predetermined ranges of *in vitro* scores. The differences between the two ranges are attributed to two different criteria used to identify ocular corrosives and severe irritants (i.e., EPA Category I, EU R41, GHS Category 1). One approach (**Table 2-1**) included the ICCVAM recommended decision criteria for identifying an ocular corrosive/severe irritant (i.e., IVIS ≥ 55.1, ICCVAM, 2006a).

Table 2-1 *In Vitro* Ocular Irritancy Classification Scheme for the BCOP Test Method (ICCVAM 2006)

In Vitro Score Range	*In Vitro* Classification
0-3.0	**Not Labeled**
3.1 - 25	**Mild irritant**
25.1 - 55	**Moderate irritant**
≥ 55.1	**Severe irritant**

The second approach (**Table 2-2**) included an alternative decision criteria used for identifying an ocular corrosive/severe irritant in the AMCP BRD submission (i.e., IVIS ≥ 75).

Table 2-2 *In Vitro* Ocular Irritancy Classification Scheme for the BCOP Test Method (AMCP BRD Submission)

In Vitro Score Range	*In Vitro* Classification
0-3.0	**Not Labeled**
3.1 - 25	**Mild irritant**
25.1 – 74.9	**Moderate irritant**
≥ 75	**Severe irritant**

3.0 Substances Used for Validation of the BCOP Test Method

In vitro ocular test method validation studies should, ideally, evaluate an adequate sample of test substances and products from chemical and product classes, which have also been evaluated using the *in vivo* rabbit eye test method. Test substances with a wide range of *in vivo* ocular responses (e.g., corrosive/severe irritant to Not Labeled) also should be assessed to determine limits to the range of responses that can be evaluated by the *in vitro* test method.

As noted in **Section 1.4**, the substances contained within the AMCP BRD tested in the BCOP test method were added to BCOP data employed in the ICCVAM evaluation of BCOP for identifying ocular corrosives and severe irritants (ICCVAM, 2006a). Therefore, the database in the current evaluation was composed of substances from the AMCP BRD along with previously evaluated published reports from the literature (i.e. Gautheron et al. [1994], Balls et al. [1995], Swanson et al. [1995], Southee [1998], Swanson and Harbell [2000], and Bailey et al. [2004]).

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

938 **Table 3-1 Chemical Classes Tested in the BCOP Test Method**

Chemical Class	# of Substances	Chemical Class	# of Substances
Acyl halide	3	Imide	2
Alcohol	22	Inorganic salt	6
Aldehyde	1	Ketone	12
Alkali	3	Lactone	3
Aluminum compound	1	Nitrile compound	1
Amide	2	Nitro compound	2
Amidine	6	Oil	1
Amine	10	Onium compound	12
Amino acid	4	Organic salt	3
Boron compound	1	Organic sulfur compound	5
Carboxylic acid	17	Organophosphate	1
Ester	12	Organosilicon compound	1
Ether/Polyether	9	Phenol	1
Formulation	69	Polycyclic compound	3
Heterocyclic compound	12	Terpene	1
Hydrocarbon	18	Wax	1

939

940 As shown in **Table 3-1**, the chemical classes with the greatest amount of *in vitro* BCOP data
941 are alcohols, carboxylic acids, esters, formulations, heterocyclic compounds, hydrocarbons,
942 ketones, and onium compounds. Other chemical classes tested include amines,
943 ethers/polyethers, inorganic and organic salts, and organic sulfur compounds. The
944 formulations tested include hair shampoos, personal care cleansers, detergents, bleaches,
945 insect repellents, petroleum products, and fabric softener.

946 As shown in **Table 3-2**, the most common product classes tested in the BCOP assay are
947 chemical/synthetic intermediates, cleaners, drugs/pharmaceuticals/therapeutic agents,
948 petroleum products, solvents, shampoos, and surfactants. Other product classes tested include
949 detergents, insect repellents, lubricants, personal care cleansers, pesticides, and plasticizers.

950

950 **Table 3-2 Product Classes Tested in the BCOP Test Method**

Product Class	# of Substances	Product Class	# of Substances
Adhesive	1	Fertilizer	1
Agricultural chemical	2	Flame retardant	1
Antifreeze agent	1	Flavor ingredient	3
Antimicrobial Cleaning Product	66	Food additive	1
Bactericide/Fungicide/ Disinfectant/Germicide	11	Herbicide	3
Beverage	1	Insect repellant	8
Bleach	3	Lubricant/lubricant additive	6
Chelating agent	2	Paint, lacquer, varnish (component)	1
Chemical/synthetic intermediate	28	Pesticide	8
Cleaner	15	Petroleum product	16
Cleanser (personal care)	13	Photographic chemical/ developing agent	2
Coupling agent	1	Plant growth regulator	2
Cutting fluid	2	Plasticizer	4
Degreaser	1	Preservative	2
Dessicant	1	Reagent	5
Detergent	11	Shampoo (hair)	14
Drug/Pharmaceutical/ Therapeutic agent and/or Metabolite	17	Soap	3
Dry cleaning preparation	1	Solvent	34
Dye, in manufacture of	3	Surfactant	39
Emulsifier	1	Anionic surfactant	3
Etching and/or electroplating	2	Cationic surfactant	6
Explosive	1	Nonionic surfactant	5
Fabric softener	1	Thermometer fluid	1

951

952 ***[This Page Intentionally Left Blank]***

953

954

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

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

Most of the BCOP studies evaluated in this BRD include *in vivo* reference data generated using the basic procedures for the *in vivo* rabbit eye test method described above. These data were used by the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to assign an ocular hazard classification according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix B**). Exceptions included the following:

- For Gautheron et al. (1994), the *in vivo* reference data were obtained from concurrent *in vivo* studies performed by Dr. J. Giroux at the Agence du Medicament in Montpelier, France. Studies were performed according to European Economic Committee (EEC) (1984 and 1991) guidelines with a few modifications. Three rabbits were used per test substance and MAS (Draize et al. 1944) were calculated. Only the MAS and day 1 scores for the 52 compounds are presented in the Gautheron et al. (1994) publication. The substances were classified by the study authors according to both EEC (1984) and Kay and Calandra (1962) systems. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal were provided by Dr. Philippe Vanparys in January 2005. Sufficient *in vivo* data were provided for 51 of these substances to be classified by NICEATM according to the EPA (EPA 1996), the EU (EU 2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix C**).

- For the EC/HO validation study (Balls et al. 1995), MMAS were calculated for the 59 test substances from existing and concurrently run *in vivo* studies, all of which were performed according to OECD TG 405 and following GLP guidelines. The data were generated since 1981 and met the following criteria:
  - Normally used at least 3 New Zealand White rabbits tested at the same time.
  - A volume of 0.1 mL or the equivalent weight of substance was instilled into the conjunctival sac.
  - Anesthesia was not used.
  - Observations were made at least at 1, 2, and 3 days after instillation.

  All 59 of these substances were classified by NICEATM according to the EU (2001) classification system, but due to lack of sufficient *in vivo* data, only 55 and 57 substances, respectively, were classified according to the EPA (1996) and the GHS (UN 2003) ocular irritancy classification systems, (**Appendix C**).
- For the Swanson et al. (1995) study, *in vivo* reference data were obtained from standard (100 µL of test material; 7 formulations) or modified (30 µL of test material; 13 formulations) Draize eye irritancy tests. A MAS(30) or a MAS(100) is reported for each test substance. *In vivo* categories reported in the publication are mild (2 substances), mild/moderate (2), moderate (4), moderate/severe (1), severe/corrosive (4), and corrosive (7), and are based on an internal classification scheme used at S.C. Johnson & Son, Inc. Subsequent to the publication, the sponsor of the study, S.C. Johnson & Son, Inc., assigned GHS (UN 2003) and EPA (1996) classifications to the substances and provided these classifications, along with detailed *in vivo* data for each test substance, to NICEATM. NICEATM verified these EPA and GHS ocular irritancy classifications for 13 of the substances, and also classified the same 13 test substances based on the EU (2001) ocular irritancy classification system (**Appendix C**). However, 11 of the test substances evaluated using a 30 µL test substance volume were not included in the accuracy analysis, since

definitive classifications could not be assigned for the three regulatory ocular irritancy classification systems.

- For the European Community prevalidation study (Southee 1998) of the BCOP assay, detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for each of these substances was available in the ECETOC Reference Chemicals data bank (ECETOC 1998). Fifteen of the substances have been classified by NICEATM according to the EU (2001) system; 14 of the substances have been classified according to the EPA (1996) and the GHS (UN 2003) ocular irritancy classification systems (**Appendix C**).
- S.C. Johnson and Son, Inc. provided detailed *in vivo* reference data for nine of the 13 test substances evaluated in the Swanson and Harbell (2000) study of ethanol containing insect repellent formulations. The standard Draize eye irritancy test protocol was used for these nine test substances; each test included six animals.
- ExxonMobil Biomedical Sciences, Inc. provided detailed *in vivo* reference data for the 16 petrochemical products evaluated by Bailey et al. (2004). All 16 substances had been tested previously using the standard Draize eye irritancy test protocol; each test included either three or six animals.

4.1 *In Vivo* Classification Criteria Used for BRD Analysis

As described in ICCVAM (2006a), the *in vivo* rabbit eye database used to conduct a retrospective analysis of the accuracy of the BCOP test method includes studies that were conducted using one to six rabbits. However, some of the *in vivo* classification systems considered for the accuracy analyses are currently devised to be applied to studies using no more than three rabbits. Thus, to maximize the amount of data used for the evaluation of BCOP, the decision criteria for each classification system were expanded to include studies that used more than three rabbits in their evaluation.

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

- At least three rabbits were tested in the study, unless a severe effect (e.g., corrosion of the cornea) was noted in a single rabbit. In such cases, substance classification could proceed based on the effects observed in less than three rabbits.
- A volume of 0.1 mL or 0.1 g was tested in each rabbit. A study in which a lower quantity was applied to the eye was accepted for substance classification, provided that a severe effect (e.g., corrosion of the cornea, lesion persistence) was observed in a rabbit.
- Observations of the eye must have been made, at minimum, at 24-, 48-, and 72-hours following test substance application, if no severe effect was observed.
- Observations of the eye must have been made until reversibility was assessed, typically meaning that all endpoint scores were cleared. Results from a study terminated early were not used, unless the reason for the early termination was documented.

If any of the above criteria were not fulfilled, then the data for that substance were not used for the accuracy analyses. The rules used for classification according to the EPA, EU, or GHS classification systems are detailed in ICCVAM (2006a).

4.2 *In Vivo* Data Quality

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

Although an attempt was made to obtain the original study records, such records could not be obtained. Therefore, the extent to which the *in vivo* rabbit eye studies used to provide the comparative data in the published BCOP validation studies were compliant with GLP

guidelines is based on the information provided in the reports. Based on the available information, Balls et al. (1995) and Southee (1998) explicitly state GLP guidelines were followed. For the Bailey et al. (2004) report, approximately half of the *in vivo* studies were conducted according to GLP guidelines, while GLP compliance was not explicitly stated for the remaining half of substances. For Gautheron et al. (1994), the *in vivo* studies were conducted according to European Economic Community (EEC) 1984 and 1991 test guidelines (predecessors of the current EU test guideline for eye irritation), but this information alone does not give enough information about GLP compliance. For the remaining reports (Swanson et al. 1995 and Swanson and Harbell 2000), the extent of GLP compliance is not known.

1083 ***[This Page Intentionally Left Blank]***

1084

1085

1086

5.0 BCOP Test Method Data and Results

A total of eight reports, seven published and one unpublished obtained for this evaluation were useful for an accuracy analysis. These data were extracted from seven publications, data submissions, or study reports including: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Southee (1998), Swanson and Harbell (2000), Bailey et al. (2004) and the AMCP BRD and contained sufficient data for an accuracy analysis of the BCOP test method for the identification of all categories of ocular irritation. As detailed in **Section 6.0**, the 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

NICEATM staff made several attempts to obtain original *in vitro* and *in vivo* data from BCOP test method studies. In addition, authors of published BCOP studies were contacted to request original BCOP data and *in vivo* reference data from their respective publications. As a result of these efforts, some original BCOP test method data (i.e., corrected opacity and OD_{490} values for individual corneas) were obtained. ECVAM provided corrected opacity and OD_{490} values in a written report for 16 substances evaluated in the European Community Prevalidation Study of the BCOP (Southee 1998). Dr. Joseph Sina also submitted corrected opacity and OD_{490} values electronically for 43 compounds; however, corresponding *in vivo* reference data was not obtained. ECVAM subsequently provided the mean opacity values, mean permeability values, and mean *in vitro* scores obtained for the 59 substances evaluated in the Balls et al. (1995) study. Dr. Freddy Van Goethem provided a summary table and individual cornea data for 52 compounds tested in the EEC validation study (Gautheron et al. 1994). S.C. Johnson & Son, Inc. provided transformed BCOP data (mean opacity, permeability, and *in vitro* scores) for the Swanson et al. (1995) and Swanson and Harbell (2000) studies, and ExxonMobil Biomedical Sciences, Inc. provided detailed study reports for the Bailey et al. (2004) study.

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

The majority of other published BCOP reports, which are discussed in **Section 9.0,** did not contain sufficient *in vitro* or *in vivo* data with which to conduct an accuracy analysis.

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

The BCOP studies included in the accuracy analysis in this document (**Section 6.0**) evaluated variability in the BCOP assay by calculating the mean (± SD) for the opacity values and the OD490 values for each treatment group and control group. The mean opacity and mean permeability (OD490) values for each treatment group were then used to calculate an *in vitro* score for each treatment group as follows:

In Vitro Irritancy Score = mean opacity value + (15 x mean OD490 value)

Sina et al. (1995) reported that this formula was derived empirically during in-house and interlaboratory studies. The data generated for a series of 36 compounds in a multilaboratory study were subjected to a multivariate analysis to determine the equation of best fit between *in vivo* and *in vitro* data. This analysis was performed by scientists at two separate companies, who generated nearly identical derived equations. The *In Vitro* Irritancy Score provides a numerical value that can be used to compare the relative irritancy of test substances.

The accuracy analysis in this document is focused on evaluating the ability of the BCOP test method to identify ocular corrosives and severe irritants as defined by the EPA (1996), EU (2001), and the GHS (UN 2003). A review of the BCOP test method protocols indicates that the decision criteria applied to *in vitro* data to classify a test substance as a severe ocular irritant or a nonsevere ocular irritant (i.e., mild irritant, moderate irritant) and/or Not Labeled are similar for four BCOP protocols (Gautheron et al. 1994; Balls et al. 1995; Southee 1998; Bailey et al. 2004). The *in vitro* irritation classification scheme used in these studies is similar to the decision criteria first proposed by Gautheron et al. (1994), for which *in vitro* irritancy categories were based on predetermined ranges of *in vitro* scores (see **Section 2.0**).

5.3 Summary of Results

Where provided, the specific information extracted for each substance included its name, CASRN (if available), the concentration tested, the available BCOP data (e.g., mean opacity value, mean OD_{490} value, standard deviation, number of replicates, mean *in vitro* score), the

in vitro irritation classification of the test substance (based on the *in vitro* irritation classification scheme applied or noted by the study author), and the reference. Other supporting information, such as the source, purity and physicochemical characteristics of the test substances, was included to the extent this information was available. If not provided, the CASRN was obtained from various sources, including the National Library of Medicine's ChemID database. Chemical and product classes were assigned based on the MeSH classification system (available at http://www.nlm.nih.gov/mesh). **Appendix A** provides information on the names, synonyms, CASRN, and chemical/product class, where available, for each substance while **Appendix B** contains the *in vitro* BCOP test method data sorted by reference and alphabetically by substance name.

5.4 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines and with the use of coded chemicals (OECD 1998; EPA 2003a, 2003b; FDA 2003). The data quality was evaluated by a review of the methods section in literature references and the submitted reports. The data quality presented in the reviewed literature references can be evaluated to the extent this information was provided in the published reports. Based on the available information, the reports that specifically identified following GLP guidelines or used data obtained according to GLP guidelines were Balls et al. (1995), Swanson et al. (1995), Swanson and Harbell (2000), and Bailey et al. (2004). The reports that identified using coded chemicals were Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004).

6.0 BCOP Test Method Accuracy

6.1 Accuracy of the BCOP Test Method

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

- Accuracy (concordance): the proportion of correct outcomes (positive and negative) of a test method
- Sensitivity: the proportion of all positive substances that are classified as positive
- Specificity: the proportion of all negative substances that are classified as negative
- Positive predictivity: the proportion of correct positive responses among substances testing positive
- Negative predictivity: the proportion of correct negative responses among substances testing negative
- False positive rate: the proportion of all negative substances that are falsely identified as positive
- False negative rate: the proportion of all positive substances that are falsely identified as negative.

The ability of the BCOP test method to identify all categories of ocular irritation potential, as defined by the GHS, EPA, and EU classification systems (EPA 1996; EU 2001; UN 2003), was evaluated. This same analysis was also performed with specific chemical classes and/or physical properties excluded based on them previously being identified as discordant in BCOP (ICCVAM, 2006a).

The evaluations were conducted on the overall data set by combining results from the reports indicated in **Section 5.0** then assigning an overall ocular irritancy classification for each substance (**Appendix B** and **C**). When the same substance was evaluated in multiple

laboratories, an overall BCOP classification was based on the majority classification among all of the studies. When there was an equal number of differing irritancy classifications for substances (e.g., two tests classified a substance as a Not Labeled and two tests classified a substance as a mild irritant), the more severe irritancy classification was used for the overall classification of the substance (mild irritant, in this case).

As described in **Section 2.0**, the *in vitro* irritation classification schemes used for this evaluation were based on two different predetermined ranges of *in vitro* scores. The differences between the two ranges are attributed to two different criteria used to identify ocular corrosives and severe irritants (i.e., EPA Category I, EU R41, GHS Category 1). One approach (**Table 2-1**) included the ICCVAM recommended decision criteria for identifying an ocular corrosive/severe irritant (i.e., IVIS ≥ 55.1, ICCVAM, 2006b). The second approach (**Table 2-2**) included an alternative decision criteria used for identifying an ocular corrosive/severe irritant in the AMCP BRD submission (i.e., IVIS ≥ 75).

6.1.1 Ability to Identify Ocular Corrosives and Severe Irritants from All Other Classes

The BCOP test method has been previously recommended for use in identifying ocular corrosives and severe irritants (i.e., EPA Category I, EU R41, and GHS Category 1, ICCVAM, 2006b). In the original ICCVAM evaluation of BCOP, which was based on 145 substances, overall accuracy, false positive, and false negative rates were 79% (113/143) to 81% (119/147), 19% (20/103) to 21% (22/103), 16% (7/43) to 25% (10/40) respectively, depending on the hazard classification system evaluation (i.e., EPA, EU, or GHS). Because additional substances with sufficient BCOP and *in vivo* data were added to the BCOP test method validation database, this evaluation was repeated to verify that similar performances were achieved. Based on the current BCOP validation database, which has increased to 211 substances, overall accuracy, false positive, and false negative rates are 77% (91/118) to 79% (147/186), 24% (20/85 to 29/123), 15% (10/65) to 21% (7/33) depending on the hazard classification system evaluation (i.e., EPA, EU, or GHS) (**Table 6-1**).

Table 6-1 Accuracy of the BCOP Test Method for Distinguishing Corrosives/Severe Irritants from All Other Categories as Defined by GHS, EPA and EU Classification Systems[1]

BCOP	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.			%	No.[3]	%	No.
GHS	187	79	148/187	85	55/65	76	93/122	24	29/122	15	10/65
EPA	187	79	148/187	84	53/63	77	95/124	23	29/124	16	10/63
EU	118	77	91/118	79	26/33	76	65/85	24	20/85	21	7/33

Abbreviations: GHS = Globally Harmonized System; EPA = U.S. Environmental Protection Agency; EU = European Union; BCOP= Bovine Corneal Opacity and Permeability; NA = Not Applicable

[1]GHS classification system (UN 2003), EPA classification system (EPA 1996); EU classification system (EU 2001)

The following sections provide detailed analyses and results of the performance of the BCOP test method for each of the ocular hazard classification systems (i.e., EPA, EU, and GHS).

6.2 GHS Classification System: BCOP Test Method Accuracy

The seven reports used in the accuracy evaluation (Gautheron et al. [1994], Balls et al. [1995], Swanson et al. [1995], Southee [1998], Swanson and Harbell [2000], Bailey et al. [2004] and the Antimicrobial Cleaning Products BRD submission) included BCOP data on 211 substances, 187 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the GHS classification system (UN [2003]) (see **Appendix C**). Among these studies, Gautheron et al. (1994), Balls et al. (1995), and Southee (1998) provided BCOP data for substances tested in multiple laboratories and thus required that a consensus i*n vitro* classification be assigned to each substance. Based on results from *in vivo* rabbit eye experiments, 35% (65/187)[2] were classified as Category 1, 14% (26/187[3]) were classified as Category 2A, 3% (6/187) were classified as Category 2B, and 48% (90/187)

[2] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2B classification. The accuracy analysis was performed with the substance classified as Category 1. 1% sodium hydroxide was duplicated in the database. Sodium hydroxide (Prinsen and Koëter, 1993) was removed because the *in vivo* classification corresponded to a 10% solution.

[3] Triton X-100 (10%) and dibenzyl phosphate were excluded because they were classified *in vitro* as 2A/2B.

were classified as Not Labeled. Twenty-four substances could not be classified according to the GHS classification system due to the lack of adequate animal data and are so noted in **Appendix C**.

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

The BCOP test method correctly identified 85% (55/65) and 78% (51/65) of the Category 1 substances using decision criteria of IVIS ≥ 55.1 and IVIS ≥ 75, respectively (**Table 6-2**). Among the Category 1 substances that were underpredicted by BCOP (based on IVIS ≥ 55.1), 9% (6/65) were classified as Category 2A and 6% (4/65) were classified as Category 2B. Among the Category 1 substances that were underpredicted by BCOP (based on IVIS ≥ 75), 15% (10/65) were classified as Category 2A and 6% (4/65) were classified as Category 2B.

Table 6-2 Evaluation of the Performance of the BCOP Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by GHS, EPA and EU Classification Systems[1]

Severe using ≥55.1											
	Overall Correct Classification	**Severe[2]**		**Moderate[3]**			**Mild[4]**			**Not Labeled[5]**	
		actual	**under**	**over**	**actual**	**under**	**over**	**actual**	**under**	**over**	**actual**
GHS	49% (91/187)	85% (55/65)	15% (10/65)	62% (16/26)	27% (7/26)	11% (3/26)	67% (4/6)	33% (2/6)	0% (0/6)	70% (63/90)	30% (27/90)
EPA	55% (101/187)	84% (53/63)	16% (10/63)	50% (11/22)	32% (7/22)	18% (4/22)	50% (28/57)	36% (21/57)	14% (8/57)	53% (24/45)	47% (21/45)
EU	50% (59/118)	79% (26/33)	21% (7/33)	48% (10/21)	52% (11/21)	0% (0/21)	NA	NA	NA	66% (42/64)	34% (22/64)
Severe using ≥75											
		Severe		**Moderate**			**Mild**			**Not Labeled**	
		actual	**under**	**over**	**actual**	**under**	**over**	**actual**	**under**	**over**	**actual**
GHS	50% (94/187)	78% (51/65)	22% (14/65)	31% (8/26)	54% (14/26)	15% (4/26)	67% (4/6)	33% (2/6)	0% (0/6)	70% (63/90)	30% (27/90)
EPA	49% (92/187)	78% (49/63)	22% (14/63)	36% (8/22)	45% (10/22)	19% (4/22)	47% (27/57)	39% (22/57)	14% (8/57)	53% (24/45)	47% (21/45)
EU	51% (60/118)	73% (24/33)	27% (9/33)	29% (6/21)	67% (14/21)	4% (1/21)	NA	NA	NA	66% (42/64)	34% (22/64)

Abbreviations: GHS = Globally Harmonized System; EPA = U.S. Environmental Protection Agency; EU = European Union; BCOP= Bovine Corneal Opacity and Permeability; NA = Not Applicable

[1]GHS classification system (UN 2003), EPA classification system (EPA 1996); EU classification system (EU 2001)

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

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

[4]Mild = GHS Category 2B; EPA Category III; EU R36.

[5]Not Labeled = Not Classified.

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

For the 26 substances that could be evaluated, the BCOP test method correctly identified 27% (7/26) as moderate irritants while 62% (16/26) were overpredicted and 11% (3/26) were underpredicted using decision criteria defining ocular corrosives/severe irritants ≥55.1 (**Table 6-2**). Using decision criteria defining ocular corrosives/severe irritants ≥75, the BCOP test method correctly identified 54% (14/26) as moderate irritants while 31% (8/26) were overpredicted and 15% (4/26) were underpredicted (**Table 6-2**).

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

Regardless of the decision criteria used for defining ocular corrosives/severe irritants, for the six substances that could be evaluated, the BCOP test method correctly identified 33% (2/6) as mild irritants while 67% (4/6) were overpredicted (**Table 6-2**).

6.2.4 Identification of Not Classified Substances

Regardless of the decision criteria used for defining ocular corrosives/severe irritants, for the 90 substances that could be evaluated, the BCOP test method correctly identified 30% (27/90) as Not Classified while 70% (63/90) were overpredicted (**Table 6-2**).

As indicated in Table 6-1, the use of the alternative decision criteria proposed in the AMCP BRD (2008) in which ocular corrosives/severe irritants ≥75, did not improve the overall performance of BCOP hazard classification. Therefore, the remaining analyses will present results utilizing the ICCVAM recommended decision criteria for ocular corrosives/severe irritants (≥55.1).

6.2.5 Overall Correct Classification

As indicated in **Table 6-2**, the use of the alternative decision criteria proposed in the AMCP BRD (2008) in which ocular corrosives/severe irritants ≥75, did not improve the overall performance of BCOP hazard classification. Therefore, the remaining analyses will present results utilizing the ICCVAM recommended decision criteria for ocular corrosives/severe irritants (≥55.1). Overall, correct classification for the entire database of 187 substances was 49% (91/187), but ranged from 25% (2/8) to 75% (6/8) when each of the eight individual validation databases was evaluated (**Table 6-3**).

Table 6-3 Evaluation of the Performance of the BCOP 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[2]		Moderate[3]			Mild[4]			Not Classified[5]	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Gautheron et al. (1994)	45% (18/40)	100% (4/4)	0% (0/4)	67% (2/3)	33% (1/3)	0% (0/3)	100% (1/1)	0% (0/1)	0% (0/1)	59% (19/32)	41% (13/32)
Balls et al. (1995)	48% (20/42)	75% (12/16)	25% (4/16)	50% (5/10)	30% (3/10)	20% (2/10)	67% (2/3)	33% (1/3)	0% (0/3)	69% (9/13)	31% (4/13)
Swanson et al. (1995)	75% (6/8)	100% (6/6)	0% (0/6)	0% (0/0)	0% (0/0)	0% (0/0)	0% (0/0)	0% (0/0)	0% (0/0)	100% (2/2)	0% (0/2)
Southee (1998)	50% (5/10)	50% (3/6)	50% (3/6)	50% (1/2)	50% (1/2)	0% (0/2)	0% (0/1)	100% (1/1)	0% (0/1)	100% (1/1)	0% (0/1)
Swanson & Harbell (2000)	25% (2/8)	100% (1/1)	0% (0/1)	50% (2/4)	25% (1/4)	25% (1/4)	0% (0/0)	0% (0/0)	0% (0/0)	100% (3/3)	0% (0/3)
Bailey et al. (2004)	50% (7/14)	67% (2/3)	33% (1/3)	0% (0/0)	0% (0/0)	0% (0/0)	100% (1/1)	0% (0/1)	0% (0/1)	50% (5/10)	50% (5/10)
AMCP BRD	51% (33/65)	93% (27/29)	7% (2/29)	86% (6/7)	14% (1/7)	0% (0/7)	0% (0/0)	0% (0/0)	0% (0/0)	83% (24/29)	17% (5/29)
Overall	49% (91/187)	85% (55/65)	15% (10/65)	62% (16/26)	27% (7/26)	11% (3/26)	67% (4/6)	33% (2/6)	0% (0/6)	70% (63/90)	30% (27/90)

Abbreviations: GHS = Globally Harmonized System; BCOP = Bovine Corneal Opacity and Permeability;
[1]GHS classification system (UN 2003)
[2]Severe = Category 1.
[3]Moderate = Category 2A.
[4]Mild = Category 2B.
[5]Not Classified = Not Classified.

6.2.6 Ability to Identify Not Classified from All Other Classes

In addition to evaluating the ability of the BCOP test method to identify each individual ocular hazard category according to GHS classification system, ICCVAM also evaluated the ability of the BCOP test method to distinguish not classified substances from all irritant classes. Using this approach of identifying not classified substances from all other classes for the 187 substances considered, the BCOP test method has an accuracy of 66% (124/187), a sensitivity of 100% (97/97), a specificity of 30% (27/90), a false positive rate of 70% (63/90) and a false negative rate of 0% (0/97) (**Table 6-4**).

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

Gautheron et al. (1994): Based upon the *in vivo* rabbit data forty substances could be assigned a GHS classification. Based on these 40 substances, the BCOP test method has an accuracy of 53% (21/40), sensitivity of 100% (8/8), specificity of 41% (13/32), false positive rate of 59% (19/32), and a false negative rate of 0% (0/8) (**Table 6-4**).

Balls et al. (1995): Based upon the *in vivo* rabbit data 42 substances could be assigned a GHS classification. Based on these 42 substances, the BCOP test method has an accuracy of 79% (33/42), sensitivity of 100% (29/29), specificity of 31% (4/13), false positive rate of 69% (9/13), and a false negative rate of 0% (0/29) (**Table 6-4**).

Swanson et al. (1995): Based upon the *in vivo* rabbit data eight substances could be assigned a GHS classification. Based on these eight substances, the BCOP test method has an accuracy of 75% (6/8), sensitivity of 100% (6/6), specificity of 0% (0/2), false positive rate of 100% (2/2), and a false negative rate of 0% (0/6) (**Table 6-4**).

Southee (1998): Based upon the *in vivo* rabbit data 10 substances could be assigned a GHS classification. Based on these ten substances, the BCOP test method has an accuracy of 90% (9/10), sensitivity of 100% (9/9), specificity of 0% (0/1), false positive rate of 100% (1/1), and a false negative rate of 0% (0/9) (**Table 6-4**).

Swanson and Harbell (2000): Based upon the *in vivo* rabbit data eight substances could be assigned a GHS classification. Based on these eight substances, the BCOP test method has an accuracy of 63% (5/8), sensitivity of 100% (5/5), specificity of 0% (0/3), false positive rate of 100% (3/3), and a false negative rate of 0% (0/5) (**Table 6-4**).

Bailey et al. (2004): Based upon the *in vivo* rabbit data 14 substances could be assigned a GHS classification. Based on these fourteen substances, the BCOP test method has an accuracy of 64% (9/14), sensitivity of 100% (4/4), specificity of 50% (5/10), false positive rate of 50% (5/10), and a false negative rate of 0% (0/4) (**Table 6-4**).

AMCP BRD (2008): Based upon the *in vivo* rabbit data 65 substances could be assigned a GHS classification. Based on these 65 substances, the BCOP test method has an accuracy of 63% (41/65), sensitivity of 100% (36/36), specificity of 17% (5/29), false positive rate of 83% (24/29), and a false negative rate of 0% (0/36) (**Table 6-4**).

Table 6-4 Accuracy of the BCOP Test Method for Distinguishing Not Classified from All Other Irritant Classes as Defined by the GHS Classification System[1], by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Gautheron et al. (1994)	40	53	21/40	100	8/8	41	13/32	59	19/32	0	0/8
Balls et al. (1995)	42	79	33/42	100	29/29	31	4/13	69	9/13	0	0/29
Swanson et al. (1995)	8	75	6/8	100	6/6	0	0/2	100	2/2	0	0/6
Southee (1998)	10	90	9/10	100	9/9	0	0/1	100	1/1	0	0/9
Swanson & Harbell (2000)	8	63	5/8	100	5/5	0	0/3	100	3/3	0	0/5
Bailey et al. (2004)	14	64	9/14	100	4/4	50	5/10	50	5/10	0	0/4
AMCP BRD	65	63	41/65	100	36/36	17	5/29	83	24/29	0	0/36
Overall	187	66	124/187	100	97/97	30	27/90	70	63/90	0	0/97

[1]GHS = Globally Harmonized System (UN 2003). NC vs. Cat 1/2A/2B.

[2]N = Number of substances included in this analysis/the total number of substances in the study.

[3]No. = Data used to calculate the percentage.

6.2.7 Discordant Results According to the GHS Classification System

In order to evaluate discordant responses of the BCOP test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. These included specific classes of chemicals and certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, physical form), with sufficiently robust numbers of substances ($n \geq 5$).

As indicated in **Table 6-5**, there were some notable trends in the performance of the BCOP test method among these subgroups of substances. The chemical classes of substances that were most consistently overpredicted according to the GHS classification system (i.e., were false positives) by the BCOP test method were alcohols and hydrocarbons. Of the 53 overpredicted substances, eight were alcohols and eight were hydrocarbons. Additional chemical classes represented among the overpredicted substances were carboxylic acids (6), heterocyclic compounds (4), and esters (4). Among the 23 substances labeled as surfactants, 22% (5/23) were overpredicted by the BCOP test method.

With regard to the physical form of the substances overpredicted by the BCOP test method, 44 were liquids and nine were solids. Considering the proportion of the total available database, liquids (90/122; 74%) appear more likely than solids (32/122; 26%) to be overpredicted by the BCOP test method.

Alcohols (2) and carboxylic acids (2) were most often underpredicted (i.e., were false negatives[4]) by the BCOP test method according to the GHS classification system (see **Appendix C**). As can be seen in **Table 6-5**, the 16 substances labeled as surfactants were rarely underpredicted by the BCOP test method (13% [1/8] Category 1 substances was underpredicted; none of the Category 2A or 2B substances were underpredicted).

[4] False negative in this context refers to a substance that was classified as a nonsevere (mild or moderate) irritant or Not Labeled by the BCOP test method, but as a severe irritant based on *in vivo* data.

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

Category	N	Under Prediction (*In Vivo*/*In Vitro*)						Over Prediction (*In Vivo*/*In Vitro*)					
		1 (Severe)[2]			2A (Moderate)[3]		2B (Mild)[4]	2A (Mod)	2B (Mild)		NL (Not Labeled)[5]		
		NL	2B	2A	NL	2B	NL	1	1	2A	1	2A	2B
Overall	147	0% (0/36)	11% (4/36)	11% (4/36)	0% (0/19)	16% (3/19)	0% (0/6)	53% (10/19)	17% (1/6)	50% (3/6)	15% (9/61)	11% (7/61)	38% (23/61)
Chemical Class[6]													
Alcohol	18	0% (0/3)	33% (1/3)	33% (1/3)	0% (0/6)	0% (0/6)	0% (0/1)	67% (4/6)	0% (0/1)	100% (1/1)	43% (3/7)	0% (0/7)	0% (0/7)
Amine\Amidine	7	0% (0/5)	0% (0/5)	0% (0/5)	0% (0/2)	0% (0/2)	0/0	0/0	0/0	0/0	0% (0/4)	0% (0/4)	25% (1/4)
Carboxylic Acid	14	0% (0/6)	33% (2/6)	0% (0/6)	0% (0/2)	0% (0/2)	0/0	50% (1/2)	0/0	0/0	33% (2/6)	33% (2/6)	17% (1/6)
Ester	10	0% (0/2)	0% (0/2)	0% (0/2)	0% (0/3)	33% (1/3)	0% (0/1)	33% (1/3)	0% (0/1)	0% (0/1)	0% (0/4)	5% (2/4)	25% (1/4)
Ether	6	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0/0	100% (1/1)	0/0	0/0	25% (1/4)	0% (0/4)	0% (0/4)
Heterocyclic	13	0% (0/6)	17% (1/6)	0% (0/6)	0% (0/1)	0% (0/1)	0/0	0% (0/1)	0/0	0/0	17% (1/6)	0% (0/6)	50% (3/6)
Hydrocarbon	11	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	9% (1/11)	18% (2/11)	45% (5/11)
Inorganics	7	0% (0/4)	0% (0/4)	0% (0/4)	0/0	0/0	0% (0/1)	0/0	0% (0/1)	0% (0/1)	0% (0/2)	0% (0/2)	50% (1/2)
Ketone	9	0/0	0/0	0/0	0% (0/2)	0% (0/2)	0% (0/1)	0% (0/2)	0% (0/1)	0% (0/1)	33% (2/6)	0% (0/6)	17% (1/6)
Onium Compound	11	13% (1/8)	0% (0/8)	0% (0/8)	0/0	0/0	0% (0/1)	0/0	0% (0/1)	0% (0/1)	0% (0/2)	0% (0/2)	50% (1/2)
Polyether	2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0% (0/2)	0% (0/2)	0% (0/2)

Category	N	Under Prediction (*In Vivo*/*In Vitro*)						Over Prediction (*In Vivo*/*In Vitro*)					
		1 (Severe)[2]			2A (Moderate)[3]		2B (Mild)[4]	2A (Mod)	2B (Mild)		NL (Not Labeled)[5]		
		NL	2B	2A	NL	2B	NL	1	1	2A	1	2A	2B
Properties of Interest													
Liquids	90	0% (0/24)	4% (1/24)	8% (2/24)	0% (0/17)	18% (3/17)	0% (0/5)	53% (9/17)	20% (1/5)	60% (3/5)	16% (7/44)	16% (7/44)	39% (17/44)
Solids	32	0% (0/12)	25% (3/12)	17% (2/12)	0% (0/2)	0% (0/2)	0/0	50% (1/2)	0/0	0/0	12% (2/17)	0% (0/17)	35% (6/17)
Pesticide	8	0% (0/5)	20% (1/5)	20% (1/5)	0% (0/1)	0% (0/1)	0/0	100% (1/1)	0/0	0/0	50% (1/2)	0% (0/2)	50% (1/2)
Surfactant-Total	23	0% (0/14)	7% (1/14)	0% (0/14)	0% (0/1)	0% (0/1)	0% (0/1)	100% (1/1)	0% (0/1)	0% (0/1)	0% (0/7)	14% (1/7)	43% (3/7)
-nonionic	10	0% (0/5)	0% (0/5)	0% (0/5)	0% (0/1)	0% (0/1)	0/0	100% (1/1)	0/0	0/0	0% (0/4)	0% (0/4)	0% (0/4)
-anionic	9	0% (0/5)	20% (1/5)	0% (0/5)	0/0	0/0	0% (0/1)	0/0	0% (0/1)	0% (0/1)	0% (0/3)	33% (1/3)	67% (2/3)
-cationic	7	0% (0/6)	0% (0/6)	0% (0/6)	0/0	0/0	0/0	0/0	0/0	0/0	0% (0/1)	0% (0/1)	100% (1/1)

Abbreviations: GHS = Globally Harmonized System; BCOP = Bovine Corneal Opacity and Permeability
[1]GHS classification system (UN 2003)
[2]Severe = Category 1.
[3]Moderate = Category 2A.
[4]Mild = Category 2B.
[5]Not Labeled = Not Labeled
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based upon MeSH categories (www.nlm.nih.gov/mesh) as defined in Appendix A.

With regard to physical form of the substances underpredicted by the BCOP test method, six were liquids and five were solids. Given the proportion of the total available database, solids (32/122; 26%) appear more likely than liquids (90/122; 74%) to be underpredicted by the BCOP test method.

Table 6-6 shows the effects on the BCOP test method performance statistics of excluding from the data set problematic classes (i.e., those which gave the most discordant results according to the GHS classification system) identified in the BCOP BRD (ICCVAM, 2006a). In general, exclusion of alcohols, ketones or solids individually resulted in small changes in the performance statistics. Slight increases in the overall correct classification were noted with the exclusion of problematic classes, with the highest correct classification 51% (49/97) noted when both alcohols and ketones were excluded. The exclusion of problematic classes had little impact on the ability to identify Not Classified substances (see **Table 6-7**; accuracy was 68% (83/122) with the entire database, but ranged from 64% to 69% with the exclusion of problematic classes or combinations of those classes.

As indicated in **Table 6-5**, hydrocarbons were also noted as discordant when evaluating BCOP for its ability to identify all hazard categories. Among the 11 hydrocarbons in the validation database, 73% (8/11) were overpredicted by BCOP (**Table 6-5**). Compared to the entire database, exclusion of hydrocarbons resulted in only modest improvements of overall correct classification [50% (55/111) versus 48% (58/122)] and identification of Not Classified substances [38% (19/50) versus 36% (22/61)] (**Table 6-6**). Exclusion of hydrocarbons also resulted in modest improvement in overall performance for identifying Not Classified substances [see **Table 6-7**; increased accuracy from 68% (83/112) to 72% (80/111) and decreased false positive rate from 64% (39/61) to 62% (31/50), while false negative rates remained 0% (0/61 versus 0/61).

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

BCOP	Overall Correct Classification	Severe[2]		Moderate[3]			Mild[4]			Non-irritant[5]	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Overall	48% (58/122)	78% (28/36)	22% (8/36)	53% (10/19)	32% (6/19)	15% (3/19)	67% (4/6)	33% (2/6)	0% (0/6)	64% (39/61)	36% (22/61)
w/o Alcohols	49% (52/106)	82% (27/33)	18% (6/33)	46% (6/13)	31% (4/13)	23% (3/13)	60% (3/5)	40% (2/5)	0% (0/5)	65% (36/55)	35% (19/55)
w/o Ketones	49% (55/113)	78% (28/36)	22% (8/36)	47% (8/17)	35% (6/17)	18% (3/17)	80% (4/5)	20% (1/5)	0% (0/5)	64% (35/55)	36% (20/55)
w/o Solids	44% (40/90)	88% (21/24)	13% (3/24)	53% (9/17)	29% (5/17)	18% (3/17)	80% (4/5)	20% (1/5)	0% (0/5)	70% (31/44)	30% (13/44)
w/o Alcohols and Ketones	51% (49/97)	82% (27/33)	18% (6/33)	36% (4/11)	36% (4/11)	27% (3/11)	75% (3/4)	25% (1/4)	0% (0/4)	65% (32/49)	35% (17/49)
w/o Alcohols, Ketones, and Solids	47% (31/66)	91% (20/22)	9% (2/22)	33% (3/9)	34% (3/9)	33% (3/9)	100% (3/3)	0% (0/3)	0% (0/3)	75% (24/32)	25% (8/32)
w/o Hydrocarbons	50% (55/111)	78% (28/36)	22% (8/36)	53% (10/19)	32% (6/19)	15% (3/19)	67% (4/6)	33% (2/6)	0% (0/6)	62% (31/50)	38% (19/50)

Abbreviations: GHS = Globally Harmonized System; BCOP = Bovine Corneal Opacity and Permeability

[1]GHS classification system (UN 2007).

[2]Severe = Category 1.

[3]Moderate = Category 2A.

[4]Mild = Category 2B.

[5]Non-irritant = Not Classified.

Table 6-7 Accuracy of the BCOP Test Method for Distinguishing Not Classified Substances from All Other Irritant Classes as Defined by the GHS Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

BCOP	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall	122	68	83/122	100	61/61	36	22/61	64	39/61	0	0/61
w/o Alcohols	106	66	70/106	100	51/51	35	19/55	65	36/55	0	0/51
w/o Ketones	113	69	78/113	100	58/58	36	20/55	64	65/55	0	0/58
w/o Solids	90	66	59/90	100	46/46	30	13/44	70	31/44	0	0/46
w/o Alcohols and Ketones	97	67	65/97	100	48/48	35	17/49	65	32/49	0	0/48
w/o Alcohols, Ketones and Solids	66	64	42/66	100	34/34	25	8/32	75	24/32	0	0/34
w/o Hydrocarbons	111	72	80/111	100	61/61	38	19/50	62	31/50	0	0/61

[1]GHS = Globally Harmonized System (UN 2003). NC vs. Cat 1/2A/2B.
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]No. = Data used to calculate the percentage

6.3 EPA Classification System: BCOP Test Method Accuracy

The seven reports used in the accuracy evaluation (Gautheron et al. [1994], Balls et al. [1995], Swanson et al. [1995], Southee [1998], Swanson and Harbell [2000], Bailey et al. [2004] and the Antimicrobial Cleaning Products BRD submission) included BCOP data on 211 substances, 187 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the EPA classification system (UN [2003]) (see **Appendix C**). Among these studies, Gautheron et al. (1994), Balls et al. (1995), and Southee (1998) provided BCOP data for substances tested in multiple laboratories and thus required that a consensus i*n vitro* classification be assigned to each substance. Based on results from *in vivo* rabbit eye experiments, 35% (65/187)[5] were classified as Category I, 14% (26/187[6]) were classified as Category II, 3% (6/187) were classified as Category III, and 48% (90/187) were classified as Category IV. Twenty-four substances could not be classified according to the GHS classification system due to the lack of adequate animal data and are so noted in **Appendix C**.

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

The BCOP test method correctly identified 84% (53/63) and 78% (49/63) of the Category I substances using decision criteria defining ocular corrosives/severe irritants ≥55.1 and ocular corrosives/severe irritants ≥75, respectively (**Table 6-1**). Using decision criteria defining *in vitro* scores ≥55.1 as ocular corrosives/severe irritants, of the Category I substances that were under predicted by BCOP, 10% (6/63) were classified as Category II and 6% (4/63) were classified as Category III. Using decision criteria defining *in vitro* scores ≥75 as ocular corrosives/severe irritants, of the Category I substances that were under predicted by BCOP, 16% (10/63) were classified as Category II and 6% (4/63) were classified as Category III.

[5] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2B classification. The accuracy analysis was performed with the substance classified as Category 1. 1% sodium hydroxide was duplicated in the database. Sodium hydroxide (Prinsen and Koëter, 1993) was removed because the *in vivo* classification corresponded to a 10% solution.

[6] Triton X-100 (10%) and dibenzyl phosphate were excluded because they were classified *in vitro* as 2A/2B.

Category I substances that were underpredicted by BCOP, 10% (6/63) were classified as Category II and 6% (4/63) were classified as Category III. Using decision criteria defining *in vitro* scores ≥75 as ocular corrosives/severe irritants, of the Category I substances that were underpredicted by BCOP, 16% (10/63) were classified as Category II and 6% (4/63) were classified as Category III.

6.3.2 Identification of Category II Substances (Moderate Ocular Irritants)

For the 22 substances that could be evaluated, the BCOP test method correctly identified 32% (7/22) as moderate irritants while 50% (11/22) were overpredicted and 18% (4/22) were underpredicted using decision criteria defining *in vitro* scores ≥55.1 ocular corrosives/severe irritants (**Table 6-8**). Using decision criteria defining *in vitro* scores ≥75 as ocular corrosives/severe irritants, the BCOP test method correctly identified 45% (10/22) as moderate irritants while 36% (8/22) were overpredicted and 19% (4/22) were underpredicted (**Table 6-1**).

Table 6-8 Evaluation of the Performance of the BCOP 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[2]		Moderate[3]			Mild[4]			Non-irritant[5]	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Gautheron et al. (1994)	53% (21/40)	100% (4/4)	0% (0/4)	67% (2/3)	33% (1/3)	0% (0/3)	45% (9/20)	35% (7/20)	20% (4/20)	31% (4/13)	69% (9/13)
Balls et al. (1995)	45% (19/42)	69% (9/13)	31% (4/13)	50% (5/10)	30% (3/10)	20% (2/10)	59% (10/17)	29% (5/17)	12% (2/17)	0% (0/2)	100% (2/2)
Swanson et al. (1995)	75% (6/8)	100% (6/6)	0% (0/6)	0% (0/0)	0% (0/0)	0% (0/0)	100% (1/1)	0% (0/1)	0% (0/1)	100% (0/1)	0% (0/1)
Southee (1998)	50% (5/10)	40% (2/5)	60% (3/5)	50% (1/2)	50% (1/2)	0% (0/2)	33% (1/3)	67% (2/3)	0% (0/3)	0% (0/0)	0% (0/0)
Swanson & Harbell (2000)	50% (4/8)	100% (3/3)	0% (0/3)	0% (0/2)	50% (1/2)	50% (1/2)	100% (1/1)	0% (0/1)	0% (0/1)	100% (2/2)	0% (0/2)
Bailey et al. (2004)	46% (6/13)	0% (0/1)	100% (1/1)	0% (0/0)	0% (0/0)	0% (0/0)	33% (1/3)	33% (1/3)	33% (1/3)	44% (4/9)	56% (5/9)
AMCP BRD	62% (41/66)	94% (29/31)	6% (2/31)	60% (3/5)	20% (1/5)	20% (1/5)	42% (5/12)	50% (6/12)	8% (1/12)	72% (13/18)	28% (5/18)
Overall	54% (102/187)	84% (53/63)	16% (10/63)	50% (11/22)	32% (7/22)	18% (4/22)	50% (28/57)	36% (21/57)	14% (8/57)	53% (24/45)	47% (21/45)

Abbreviations: EPA = U.S. Environmental Protection Agency; BCOP= Bovine Corneal Opacity and Permeability;
[1]EPA classification system (EPA 1996)
[2]Severe = Category I.
[3]Moderate = Category II.
[4]Mild = Category III.
[5]Not Labeled = Category IV.

6.3.3 Identification of Category III Substances (Mild Ocular Irritants)

Using decision criteria defining *in vitro* scores ≥55.1 as ocular corrosives/severe irritants, for the 56 substances that could be evaluated, the BCOP test method correctly identified 36% (21/57) as mild irritants while 50% (28/57) were overpredicted and 14% (8/57) were underpredicted (**Table 6-8**). Using decision criteria defining *in vitro* scores ≥75 as ocular corrosives/severe irritants, for the 57 substances that could be evaluated, the BCOP test method correctly identified 39% (22/57) as mild irritants while 47% (27/57) were overpredicted and 14% (8/57) were underpredicted (**Table 6-1**).

6.3.4 Identification of Category IV Substances

Regardless of decision criteria used for defining *in vitro* scores as ocular corrosives/severe irritants, for the 45 substances that could be evaluated, the BCOP test method correctly identified 47% (21/45) as Category IV while 53% (24/45) were overpredicted (**Table 6-8**).

6.3.5 Ability to identify Category IV from All Other Classes

In addition to evaluating the ability of the BCOP test method to identify each individual ocular hazard category according to the GHS classification system, ICCVAM also evaluated the ability of the BCOP test method to distinguish Category IV from all irritant classes. Using this approach of identifying Category IV from all other classes for the 187 substances considered, the BCOP test method has an accuracy of 83% (155/187), a sensitivity of 94% (134/142), a specificity of 47% (21/45), a false positive rate of 53% (24/45) and a false negative rate of 6% (8/142) (**Table 6-9**). Of the eight false negative compounds, six were from discordant classes; five solids and one ketone. Chemical class information was unavailable for the one substance that was from the AMCP BRD.

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

Gautheron et al. (1994): Based upon the *in vivo* rabbit data, 40 substances could be assigned an EPA classification. Based on these 40 substances, the BCOP test method has an accuracy of 80% (32/40), sensitivity of 85% (23/27), specificity of 69% (9/13), false positive rate of 31% (4/13), and a false negative rate of 15% (4/27) (**Table 6-9**).

Balls et al. (1995): Based upon the *in vivo* rabbit data 42 substances could be assigned an EPA classification. Based on these 42 substances, the BCOP test method has an accuracy of 95% (40/42), sensitivity of 95% (38/40), specificity of 100% (2/2), false positive rate of 0% (0/2), and a false negative rate of 5% (2/40) (**Table 6-9**).

Swanson et al. (1995): Based upon the *in vivo* rabbit data eight substances could be assigned an EPA classification. Based on these eight substances, the BCOP test method has an accuracy of 88% (7/8), sensitivity of 100% (7/7), specificity of 0% (0/1), false positive rate of 100% (1/1), and a false negative rate of 0% (0/7) (**Table 6-9**).

Southee (1998): Based upon the *in vivo* rabbit data 10 substances could be assigned an EPA classification. Based on these 10 substances, the BCOP test method has an accuracy of 100% (10/10), sensitivity of 100% (10/10), specificity of 0% (0/10), false positive rate of 0% (0/0), and a false negative rate of 0% (0/0) (**Table 6-9**).

Swanson and Harbell (2000): Based upon the *in vivo* rabbit data eight substances could be assigned an EPA classification. Based on these eight substances, the BCOP test method has an accuracy of 67% (6/8), sensitivity of 100% (6/6), specificity of 0% (0/2), false positive rate of 100% (2/2), and a false negative rate of 0% (0/6) (**Table 6-9**).

Bailey et al. (2004): Based upon the *in vivo* rabbit data thirteen substances could be assigned an EPA classification. Based on these thirteen substances, the BCOP test method has an accuracy of 62% (8/13), sensitivity of 75% (3/4), specificity of 56% (5/9), false positive rate of 44% (4/9), and a false negative rate of 25% (1/4) (**Table 6-9**).

AMCP BRD: Based upon the *in vivo* rabbit data 66 substances could be assigned an EPA classification. Based on these 66 substances, the BCOP test method has an accuracy of 79% (52/66), sensitivity of 98% (47/48), specificity of 28% (5/18), false positive rate of 72% (13/18), and a false negative rate of 2% (1/48) (**Table 6-9**).

Table 6-9 Accuracy of the BCOP Test Method for Distinguishing Category IV Ocular Irritants from All Other Irritant Classes as Defined by the EPA Classification System[1], by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Gautheron et al. (1994)	40	80	32/40	85	23/27	69	9/13	31	4/13	15	4/27
Balls et al. (1995)	42	95	40/42	95	38/40	100	2/2	0	0/2	5	2/40
Swanson et al. (1995)	8	88	7/8	100	7/7	0	0/1	100	1/1	0	0/7
Southee (1998)	10	100	10/10	100	10/10	0	0/0	0	0/0	0	0/10
Swanson & Harbell (2000)	8	67	6/8	100	6/6	0	0/2	100	2/2	0	0/6
Bailey et al. (2004)	13	62	8/13	75	3/4	56	5/9	44	4/9	25	1/4
AMCP BRD	66	79	52/66	98	47/48	28	5/18	72	13/18	2	1/48
Overall	187	83	155/187	94	134/142	47	21/45	53	24/45	6	8/142

[1] EPA classification system (EPA 1996). Cat IV vs. Cat I/II/III.
[2]N = Number of substances included in this analysis/the total number of substances in the study.
3No. = Data used to calculate the percentage

6.3.6 Discordant Results According to the EPA Classification System

In order to evaluate discordant responses of the BCOP test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., pesticides, surfactants, pH, physical form). As indicated in **Table 6-10**, there were some notable trends in the performance of the BCOP test method among these subgroups of substances. The chemical class that was most consistently overpredicted according to the EPA classification system (i.e., were false positives) by the BCOP test method is alcohols. Nine out the 41 overpredicted substances were alcohols. Additional chemical classes represented among the overpredicted substances were hydrocarbons (6), carboxylic acids (5), ketones (4), esters (4), ethers (3) inorganic salts (1) and onium compounds (1). Among the substances labeled as surfactants only 17% (2/12) were overpredicted by the BCOP test method.

Among the eight false negatives for the EPA system, 100% (8/8) were EPA Category III substances based on Draize data. For 38% (3/8) of these substances, the categorization was based on at least one rabbit with a corneal opacity score of one that was not resolved until day three of the study. Another substance was categorized based on all six rabbits with a conjunctival redness score of three that was not resolved until day seven of the study.

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

Category	N	Under Prediction (*In Vivo*/*In Vitro*)						Over Prediction (*In Vivo*/*In Vitro*)					
		I (Severe)[2]			II (Moderate)[3]		III (Mild)[4]	II (Mod)	III (Mild)		IV (Not Labeled)[5]		
		IV	III	II	IV	III	IV	I	I	II	I	II	III
Overall	121	0% (0/32)	13% (4/32)	13% (4/32)	0% (0/17)	18% (3/17)	16% (7/45)	47% (8/17)	29% (13/45)	20% (9/45)	4% (1/27)	0% (0/27)	37% (10/27)
Chemical Class[6]													
Alcohol	17	0% (0/2)	50% (1/2)	50% (1/2)	0% (0/6)	0% (0/6)	0% (0/5)	67% (4/6)	80% (4/5)	20% (1/5)	0% (0/4)	0% (0/4)	0% (0/4)
Amine\Amidine	7	0% (0/2)	0% (0/2)	0% (0/2)	0/0	0/0	50% (2/4)	0/0	0% (0/4)	25% (1/4)	0% (0/1)	0% (0/1)	0% (0/1)
Carboxylic Acid	15	0% (0/7)	0% (0/7)	0% (0/7)	0% (0/2)	0% (0/2)	20% (1/5)	50% (1/2)	20% (1/5)	40% (2/5)	100% (1/1)	0% (0/1)	0% (0/1)
Ester	10	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/4)	25% (1/4)	20% (1/5)	50% (2/4)	0% (0/5)	40% (2/5)	0/0	0/0	0/0
Ether	6	0/0	0/0	0/0	0% (0/4)	0% (0/4)	0% (0/2)	100% (1/1)	67% (2/3)	0% (0/3)	0% (0/4)	0% (0/4)	0% (0/4)
Heterocyclic	12	0% (0/5)	20% (1/5)	0% (0/5)	0/0	0/0	25% (1/4)	0% (0/1)	20% (1/5)	0% (0/5)	0% (0/1)	0% (0/1)	0% (0/1)
Hydrocarbon	11	0/0	0/0	0/0	0% (0/4)	0% (0/4)	0% (0/2)	0/0	20% (1/5)	40% (2/5)	0% (0/6)	0% (0/6)	50% (3/6)
Inorganics	7	0% (0/3)	0% (0/3)	0% (0/3)	0% (0/1)	0% (0/1)	33% (1/3)	100% (1/1)	0% (0/3)	0% (0/3)	0/0	0/0	0/0
Ketone	10	0/0	0/0	0/0	0% (0/1)	0% (0/1)	14% (1/7)	100% (1/1)	43% (3/7)	0% (0/7)	0% (0/1)	0% (0/1)	0% (0/1)
Onium Compound	10	0% (0/6)	0% (0/6)	17% (1/6)	0% (0/1)	0% (0/1)	0% (0/2)	100% (1/1)	0% (0/2)	0% (0/2)	0% (0/1)	0% (0/1)	0% (0/1)
Polyether	2	0/0	0/0	0/0	0/0	0/0	100% (1/1)	0/0	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)	0% (0/1)

Category	N	Under Prediction (*In Vivo*/*In Vitro*)						Over Prediction (*In Vivo*/*In Vitro*)					
		I (Severe)[2]			II (Moderate)[3]		III (Mild)[4]	II (Mod)	III (Mild)		IV (Not Labeled)[5]		
		IV	III	II	IV	III	IV	I	I	II	I	II	III
Properties of Interest													
Liquids	89	0% (0/21)	5% (1/21)	10% (2/21)	0% (0/15)	20% (3/15)	9% (3/33)	47% (7/15)	36% (12/33)	27% (9/33)	0% (0/20)	0% (0/20)	45 (9/20)
Solids	32	0% (0/11)	27% (3/11)	18% (2/11)	0% (0/2)	0% (0/2)	36% (4/11)	50% (1/2)	9% (1/11)	0% (0/11)	14% (1/7)	0% (0/7)	14% (1/7)
Pesticide	9	0% (0/5)	20% (1/5)	20% (1/5)	0/0	0/0	0% (0/4)	0/0	67% (2/3)	0% (0/3)	0/0	0/0	0/0
Surfactant-Total	22	0% (0/11)	9% (1/11)	0% (0/11)	0% (0/2)	0% (0/2)	17% (1/6)	100% (2/2)	33% (2/6)	33% (2/6)	0% (0/3)	0% (0/3)	33% (1/3)
-nonionic	11	0% (0/4)	0% (0/4)	0% (0/4)	0% (0/1)	0% (0/1)	33% (1/3)	100% (1/1)	67% (2/3)	0% (0/3)	0% (0/3)	0% (0/3)	33% (1/3)
-anionic	8	0% (0/5)	20% (1/5)	0% (0/5)	0/0	0/0	0% (0/2)	0/0	0% (0/2)	100% (2/2)	0% (0/1)	0% (0/1)	100% (1/1)
-cationic	6	0% (0/4)	0% (0/4)	0% (0/4)	0% (0/1)	0% (0/1)	0% (0/1)	100% (1/1)	0% (0/1)	0% (0/1)	0/0	0/0	0/0

Abbreviations: EPA classification system (EPA 1996); BCOP = Bovine Corneal Opacity and Permeability

[1] EPA classification system (EPA 1996)

[2]Severe = Category I.

[3]Moderate = Category II.

[4]Mild = Category III.

[5]Not Labeled = Category IV

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

With regard to physical form of the substances overpredicted by the BCOP test method, 37 were liquids and four were solids. Considering the proportion of the total available database, liquids (89/121; 74%) appear more likely than solids (32/121; 26%) to be overpredicted by the BCOP test method. Among the 22 substances labeled as surfactants, 32% (7/22) were overpredicted by the BCOP test method.

A relatively small number of substances were underpredicted (i.e., were false negatives) by the BCOP test method according to the EPA classification system (see **Appendix C**). Alcohols (2), esters (2) and heterocyclic compounds were most often underpredicted. As can be seen in **Table 6-10**, some of the 21 substances labeled as surfactants were underpredicted by the BCOP test method (56% [5/9] false negative rate).

With regard to physical form of the substances underpredicted by the BCOP test method, nine were solids and nine were liquids. Given the proportion of the total available database, solids (32/121; 26%) appear more likely than liquids (89/121; 74%) to be underpredicted by the BCOP test method.

Table 6-11 shows the effects on the BCOP test method performance statistics of excluding from the data set problematic classes (i.e., those that gave the most discordant results according to the EPA classification system) identified in the BCOP BRD (ICCVAM, 2006a). In general, the exclusion of alcohols, ketones or solids individually resulted in small changes in the performance statistics. Exclusion of both alcohols and ketones improved the overall classification rate; 56% (54/96) versus 51% (62/121) for all compounds in the database. The classification of ocular corrosives/severe irritants was most improved by the exclusion of problematic classes. Using the entire database, 75% (24/32) of severe ocular corrosives/severe irritants were accurately classified while removal of solids resulted in 86% (18/21) correct classification and removal of alcohols, ketones and solids resulted in correct classification of 90% (18/20) Category I.

As indicated in **Table 6-10**, hydrocarbons were also noted as discordant when evaluating BCOP for its ability to identify all hazard categories. Among the 11 hydrocarbons in the validation database, 55% (6/11) were overpredicted by BCOP (**Table 6-10**). Compared to the entire database, exclusion of hydrocarbons resulted in only modest improvements of overall correct classification [52% (57/110) versus 51% (62/121)] and identification of Category IV

substances [62% (13/21) versus 59% (16/27)] (**Table 6-11**). Exclusion of hydrocarbons also resulted in modest improvement in overall performance for identifying Category IV substances [see **Table 6-12**; increased accuracy from 85% (103/121) to 86% (95/110), decreased false positive rate from 41% (11/27) to 38% (8/21). However, exclusion of hydrocarbons slightly increased the false negative rate 7% (7/94) to 8% (7/89).

Table 6-11 Evaluation of the Performance of the BCOP Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

BCOP	Overall Correct Classification	Severe[2]		Moderate[3]			Mild[4]			Non-irritant[5]	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Overall	51% (62/121)	75% (24/32)	25% (8/32)	47% (8/17)	35% (6/17)	18% (3/17)	49% (22/45)	36% (16/45)	15% (7/45)	41% (11/27)	59% (16/27)
w/o Alcohols	54% (57/105)	73% (24/33)	43% (14/33)	36% (4/11)	36% (4/11)	27% (3/11)	43% (17/40)	40% (16/40)	18% (7/40)	46% (11/24)	54% (13/24)
w/o Ketones	53% (59/112)	75% (24/32)	25% (8/32)	44% (7/16)	38% (6/16)	19% (3/16)	47% (18/38)	37% (14/38)	16% (6/38)	42% (11/26)	58% (15/26)
w/o Solids	48% (43/89)	86% (18/21)	14% (3/21)	47% (7/15)	33% (5/15)	20% (3/15)	64% (21/33)	27% (9/33)	9% (3/33)	45% (9/20)	55% (11/20)
w/o Alcohols and Ketones	56% (54/96)	80% (24/30)	20% (6/30)	30% (3/10)	40% (4/10)	30% (3/10)	39% (13/33)	42% (14/33)	18% (6/33)	48% (11/23)	52% (12/23)
w/o Alcohols, Ketones, and Solids	54% (35/65)	90% (18/20)	10% (2/20)	25% (2/8)	38% (3/8)	37% (3/8)	57% (12/21)	33% (7/21)	10% (2/21)	56% (9/16)	44% (7/16)
w/o Hydrocarbons	52% (57/110)	75% (24/32)	25% (8/32)	47% (8/17)	35% (6/17)	18% (3/17)	48% (19/40)	35% (14/40)	17% (7/40)	38% (8/21)	62% (13/21)

Abbreviations: EPA = Environmental Protection Agency; BCOP = Bovine Corneal Opacity and Permeability;
[1]EPA classification system (EPA 1996).
[2]Severe = Category 1.
[3]Moderate = Category II.
[4]Mild = Category III.
[5]Non-irritant = Category IV.

Table 6-12 shows the effects on the ability of the BCOP test method to distinguish Category IV substances based upon exclusion of problematic classes from the data set. Exclusion of problematic classes individually or in combination, had a minimal effect on accuracy 85% versus 82% to 87%, sensitivity 91% to 96% or specificity 44% to 63%. The overall false positive rate of 7% (7/94) showed the largest decrease following the exclusion of solids where the false positive rate is reduced to 4% (3/69).

Table 6-12 Accuracy of the BCOP Test Method for Distinguishing Category IV Ocular Irritants from All Other Irritant Classes as Defined by the EPA Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

BCOP	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall	121	85	103/121	93	87/94	59	16/27	41	11/27	7	7/94
w/o Alcohols	105	83	87/105	91	74/81	63	13/24	46	11/24	9	7/81
w/o Ketones	112	85	95/112	93	80/86	58	15/26	42	11/26	7	6/86
w/o Solids	89	87	77/89	96	66/69	55	11/20	45	9/20	4	3/69
w/o Alcohols and Ketones	96	82	79/96	92	67/73	52	12/23	48	11/23	8	6/73
w/o Alcohols, Ketones and Solids	65	82	53/65	96	47/49	44	7/16	56	9/16	4	2/49
w/o Hydrocarbons	110	86	95/110	92	82/89	62	13/21	38	8/21	8	7/89

[1] EPA classification system (EPA 1996). Cat IV vs. Cat I/II/III.
[2] N = Number of substances included in this analysis/the total number of substances in the study.
[3] No. = Data used to calculate the percentage

6.4 EU Classification System: BCOP Test Method Accuracy

The six reports used in the accuracy evaluation (Gautheron et al. [1994], Balls et al. [1995], Swanson et al. [1995], Southee [1998], Swanson and Harbell [2000], and Bailey et al. [2004]) included BCOP data on 118 substances that had sufficient *in vivo* data to be assigned an ocular irritancy classification according to the EU classification system (EU [2004]) (see **Appendix C**). Among these studies, Gautheron et al. (1994), Balls et al. (1995), and Southee (1998) provided BCOP data for substances tested in multiple laboratories and thus required that a consensus i*n vitro* classification be assigned to each substance. Based on results from *in vivo* rabbit eye experiments, 28% (33/118)[7] were classified as R41, 14% (21/118[8]) were classified as R36, 54% (64/118) were classified as Not Labeled.

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

The BCOP test method correctly identified 79% (26/33) and 73% (24/33) of the R41 substances using decision criteria defining *in vitro* scores ≥55.1 as R41 and *in vitro* scores ≥75 as R41, respectively (**Table 6-1**). Using decision criteria defining *in vitro* scores ≥55.1 as R41, among the seven substances that were underpredicted by BCOP, all were classified as R36. Using decision criteria defining *in vitro* scores ≥75 as R41, among the seven substances that were underpredicted by BCOP, all were classified as R36.

6.4.2 Identification of R36 Substances (Irritants)

For the 21 substances that could be evaluated, the BCOP test method correctly identified 52% (11/21) as R36 while 48% (10/21) were overpredicted using decision criteria defining *in vitro* scores ≥55.1 as R41 (**Table 6-13**). Using *in vitro* scores defining decision criteria ≥75 as R41, the BCOP test method correctly identified 67% (14/21) as R36 while 29% (6/21) were overpredicted and 4% (1/21) were underpredicted (**Table 6-1**).

---

[7] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2B classification. The accuracy analysis was performed with the substance classified as Category 1. 1% sodium hydroxide was duplicated in the database. Sodium hydroxide (Prinsen and Koëter, 1993) was removed because the *in vivo* classification corresponded to a 10% solution.

[8] Triton X-100 (10%) and dibenzyl phosphate were excluded because they were classified *in vitro* as 2A/2B.

Table 6-13 Evaluation of the Performance of the BCOP 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[2]		Moderate[3]			Mild			Non-irritant[4]	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Gautheron et al. (1994)	45% (18/40)	100% (4/4)	0% (0/4)	67% (2/3)	33% (1/3)	0% (0/3)	NA	NA	NA	61% (20/33)	39% (13/33)
Balls et al. (1995)	50% (19/38)	71% (10/14)	29% (4/14)	50% (5/10)	50% (5/10)	0% (0/10)	NA	NA	NA	71% (10/14)	29% (4/14)
Swanson et al. (1995)	67% (6/9)	100% (6/6)	0% (0/6)	0% (0/0)	0% (0/0)	0% (0/0)	NA	NA	NA	100% (3/3)	0% (0/3)
Southee (1998)	60% (6/10)	60% (3/5)	40% (2/5)	25% (1/4)	75% (3/4)	0% (0/4)	NA	NA	NA	100% (1/1)	0% (0/1)
Swanson & Harbell (2000)	38% (3/8)	100% (1/1)	0% (0/1)	50% (2/4)	50% (2/4)	0% (0/4)	NA	NA	NA	100% (3/3)	0% (0/3)
Bailey et al. (2004)	54% (7/13)	67% (2/3)	33% (1/3)	0% (0/0)	0% (0/0)	0% (0/0)	NA	NA	NA	50% (5/10)	50% (5/10)
AMCP BRD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Overall	50% (59/118)	79% (26/33)	21% (7/33)	48% (10/21)	52% (11/21)	0% (0/21)	NA	NA	NA	66% (42/64)	34% (22/64)

Abbreviations: EU = European Union; BCOP = Bovine Corneal Opacity and Permeability;
NA = Not Applicable
[1]EU classification system (EU 2001)
[2]Severe = R41.
[3]Moderate = R36.
[4]Not Labeled = Not classified.

6.4.3 Identification of Not Labeled Substances

Regardless of decision criteria used for defining R41, for the 64 substances that could be evaluated, the BCOP test method correctly identified 34% (22/64) as Not Labeled, while 66% (42/64) were overpredicted (**Table 6-13**).

6.4.4 Ability to Distinguish Not Labeled Substances from All Other Classes

In addition to evaluating the ability of the BCOP test method to identify each individual ocular hazard category according to the EU classification system, ICCVAM also evaluated the ability of the BCOP test method to distinguish Not Labeled substances from all irritant classes. Using this approach of identifying Not Labeled substances from all other classes for the 118 substances considered, the BCOP test method has an accuracy of 64% (76/118), a sensitivity of 100% (54/54), a specificity of 34% (22/64), a false positive rate of 66% (42/64) and a false negative rate of 0% (0/54) (**Table 6-14**).

Table 6-14 Accuracy of the BCOP Test Method for Distinguishing Not Labeled Substances from All Other Irritant Classes as Defined by the EU Classification System[1], by Study and Overall

Data Source	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Gautheron et al. (1994)	40	50	20/40	100	7/7	39	13/33	61	20/33	0	0/7
Balls et al. (1995)	38	74	28/38	100	24/24	29	4/14	61	10/14	0	0/24
Swanson et al. (1995)	9	67	6/9	100	6/6	0	0/3	100	3/3	0	0/6
Southee (1998)	10	90	9/10	100	9/9	0	0/1	100	1/1	0	0/9
Swanson & Harbell (2000)	8	63	5/8	100	5/5	0	0/3	100	3/3	0	0/5
Bailey et al. (2004)	13	62	8/13	100	3/3	50	5/10	50	5/10	0	0/3
AMCP BRD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Overall	118	64	76/118	100	54/54	34	22/64	66	42/64	0	0/54

[1]EU classification system (EU 2001). NC vs. R41/R36.
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]No. = Data used to calculate the percentage

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

Gautheron et al. (1994): Based upon the *in vivo* rabbit data 40 substances could be assigned an EU classification. Based on these 40 substances, the BCOP test method has an accuracy of 50% (20/40), sensitivity of 100% (7/7), specificity of 39% (13/33), false positive rate of 61% (20/33), and a false negative rate of 0% (0/7) (**Table 6-14**).

Balls et al. (1995): Based upon the *in vivo* rabbit data thirty-eight substances could be assigned an EU classification. Based on these thirty-eight substances, the BCOP test method has an accuracy of 74% (28/38), sensitivity of 100% (24/24), specificity of 29% (4/14), false positive rate of 61% (10/14), and a false negative rate of 0% (0/24) (**Table 6-14**).

Swanson et al. (1995): Based upon the *in vivo* rabbit data nine substances could be assigned an EPA classification. Based on these nine substances, the BCOP test method has an accuracy of 67% (6/9), sensitivity of 100% (6/6), specificity of 0% (0/3), false positive rate of 100% (3/3), and a false negative rate of 0% (0/6) (**Table 6-14**).

Southee (1998): Based upon the *in vivo* rabbit data 10 substances could be assigned an EPA classification. Based on these 10 substances, the BCOP test method has an accuracy of 90% (9/10), sensitivity of 100% (9/9), specificity of 0% (0/1), false positive rate of 100% (1/1), and a false negative rate of 0% (0/9) (**Table 6-14**).

Swanson and Harbell (2000): Based upon the *in vivo* rabbit data eight substances could be assigned an EPA classification. Based on these eight substances, the BCOP test method has an accuracy of 63% (5/8), sensitivity of 100% (5/5), specificity of 0% (0/3), false positive rate of 100% (3/3), and a false negative rate of 0% (0/5) (**Table 6-14**).

Bailey et al. (2004): Based upon the *in vivo* rabbit data thirteen substances could be assigned an EPA classification. Based on these thirteen substances, the BCOP test method has an accuracy of 62% (8/13), sensitivity of 100% (3/3), specificity of 50% (5/10), false positive rate of 50% (5/10), and a false negative rate of 0% (0/3) (**Table 6-14**).

6.4.5 Discordant Results According to the EU Classification System

In order to evaluate discordant responses of the BCOP test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain

properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, physical form).

As indicated in **Table 6-15**, there were some notable trends in the performance of the BCOP test method among these subgroups of substances. The chemical class of substance that was most consistently overpredicted according to the EU classification system (i.e., were false positives) by the BCOP test method is alcohols. Seven out the 42 overpredicted substances were hydrocarbons. Additional chemical classes represented among the overpredicted substances were ketones (5), esters (5), carboxylic acids (4), alcohols (3), and heterocyclic compounds (3). Among the 24 substances labeled as surfactants, 25% (6/24) were overpredicted by the BCOP test method.

With regard to physical form of the substances overpredicted by the BCOP test method, 35 were liquids and seven were solids. Considering the proportion of the total available database, liquids (88/118; 75%) appear more likely than solids (29/118; 25%) to be overpredicted by the BCOP test method.

Alcohols (2) were most often underpredicted (i.e., were false negatives) by the BCOP test method according to the EU classification system (see **Appendix C**). As can be seen in **Table 6-15**, none of the 24 substances labeled as surfactants were underpredicted by the BCOP test method (0% [0/24]).

With regard to physical form of the substances underpredicted by the BCOP test method, five were solids and one was a liquid. Despite the proportion of the total available database, solids (29/118; 25%) appear more likely than liquids (88/118; 75%) to be underpredicted by the BCOP test method.

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

Category	N	Under Prediction (*In Vivo/In Vitro*)			Over Prediction (*In Vivo/In Vitro*)		
		R41[2]		R36[3]	R36	NL[4]	
		R36	NL	NL	R41	R41	R36
Overall	118	21% (7/33)	0% (0/33)	0% (0/21)	48% (10/21)	13% (8/64)	38% (24/64)
Chemical Class[5]							
Alcohol	16	67% (2/3)	0% (0/3)	0% (0/6)	50% (3/6)	0% (0/7)	0% (0/7)
Amine\Amidine	6	0% (0/2)	0% (0/2)	0/0	0/0	0% (0/4)	25% (1/4)
Carboxylic Acid	13	25% (1/4)	0% (0/4)	0% (0/3)	33% (1/3)	33% (2/6)	17% (1/6)
Ester	10	0% (0/2)	0% (0/2)	0% (0/3)	33% (1/3)	40% (2/5)	40% (2/5)
Ether	6	0% (0/1)	0% (0/1)	0% (0/1)	100% (1/1)	0% (0/2)	0% (0/2)
Heterocyclic	13	17% (1/6)	0% (0/6)	0% (0/1)	0% (0/1)	0% (0/6)	50% (3/6)
Hydrocarbon	11	0/0	0/0	0/0	0/0	18% (2/11)	45% (5/11)
Inorganics	7	0% (0/5)	0% (0/5)	0% (0/1)	0% (0/1)	0% (0/2)	50% (1/2)
Ketone	9	0/0	0/0	0% (0/2)	100% (2/2)	14% (1/7)	28% (2/7)
Onium Compound	11	13% (1/8)	0% (0/8)	0% (0/1)	0% (0/1)	0% (0/2)	50% (1/2)
Polyether	2	0/0	0/0	0/0	0/0	0% (0/2)	0% (0/2)

Category	N	Under Prediction (*In Vivo/In Vitro*)			Over Prediction (*In Vivo/In Vitro*)		
		R41[2]		R36[3]	R36	NL[4]	
		R36	NL	NL	R41	R41	R36
Properties of Interest							
Liquids	88	4% (1/23)	0% (0/23)	0% (0/18)	50% (9/18)	17% (8/47)	38% (18/47)
Solids	30	50% (5/10)	0% (0/10)	0% (0/2)	50% (1/2)	0% (0/17)	35% (6/17)
Pesticide	7	50% (2/4)	0% (0/4)	0% (0/1)	100% (1/1)	0% (0/2)	50% (1/2)
Surfactant-Total	24	0% (0/13)	0% (0/13)	0% (0/2)	50% (1/2)	22% (2/9)	33% (3/9)
-nonionic	11	0% (0/5)	0% (0/5)	0% (0/1)	100% (1/1)	0% (0/5)	20% (1/5)
Anionic	9	0% (0/4)	0% (0/4)	0% (0/1)	0% (0/1)	50% (2/4)	50% (2/4)
Cationic	7	0% (0/6)	0% (0/6)	0/0	0/0	0% (0/1)	100% (1/1)

Abbreviations: EU classification system (EU 2001); BCOP= Bovine Corneal Opacity and Permeability

[1] EU classification system (EU 2001)

[2]Severe = R41.

[3]Moderate = R36.

[4]Not Labeled = Not labeled as irritant

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

Table 6-16 shows the effects on the BCOP test method performance statistics of excluding from the data set problematic classes (i.e., those that gave the most discordant results, according to the EU classification system) identified in the BCOP BRD (ICCVAM, 2006a). In general, the exclusion of alcohols, ketones or solids individually resulted in small changes in the performance statistics. Exclusion of both alcohols and ketones improved the overall classification rate; 53% (50/94) versus 50% (59/118) for all compounds in the database. The classification of ocular corrosives/severe irritants was most improved by the exclusion of problematic classes. Using the entire database, 79% (26/33) of severe ocular corrosives/severe irritants were accurately classified while removal of solids resulted in 91% (21/23) correct classification and removal of alcohols, ketones and solids resulted in correct classification of 95% (20/21) ocular corrosives/severe irritants. Evaluation of overpredicted substances shows 64% (7/11) of hydrocarbons were overpredicted (**Table 6-15**). Compared to the entire database, exclusion of hydrocarbons improved overall correct classification [52% (56/107) versus 50% (62/121)] and slightly improved identification of Not Labeled substances [36% (19/53) versus 34% (22/64)] (**Table 6-16**).

Table 6-17 shows the effects on the ability of the BCOP test method to distinguish Not Labeled substances based upon exclusion of problematic classes from the data set. Exclusion of problematic classes individually or in combination, had a minimal effect on accuracy 64% versus 60% to 66% or specificity 24% to 35%. Sensitivity was 100% using the overall database and therefore unchanged by the exclusion of problematic classes. None of the R41 substances were classified by BCOP as Not Labeled. Exclusion of hydrocarbons resulted in modest improvement in overall performance for identifying Not Labeled substances [see **Table 6-17**; increased accuracy from 64% (76/118) to 68% (73/107), decreased false positive rate from 66% (42/64) to 64% (34/53) while the false negative rate remained 0% (0/54 versus 0/54).

Table 6-16 Evaluation of the Performance of the BCOP Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

BCOP	Overall Correct Classification	Severe[2]		Moderate[3]			Mild[4]			Non-irritant[5]	
		Actual	Under	Over	Actual	Under	Over	Actual	Under	Over	Actual
Overall	50% (59/118)	79% (26/33)	21% (7/33)	48% (10/21)	52% (11/21)	0% (0/21)	NA	NA	NA	66% (42/64)	34% (22/64)
w/o Alcohols	50% (52/103)	83% (25/30)	17% (5/30)	47% (7/15)	53% (8/15)	0% (0/15)	NA	NA	NA	67% (39/58)	33% (19/58)
w/o Ketones	52% (59/109)	79% (26/33)	21% (7/33)	42% (8/19)	58% (11/19)	0% (0/19)	NA	NA	NA	65% (37/57)	35% (20/57)
w/o Solids	49% (43/88)	91% (21/23)	9% (2/23)	50% (9/18)	50% (9/18)	0% (0/18)	NA	NA	NA	72% (34/47)	28% (13/47)
w/o Alcohols and Ketones	53% (50/94)	83% (25/30)	17% (5/30)	38% (5/13)	62% (8/13)	0% (0/13)	NA	NA	NA	67% (34/51)	33% (17/51)
w/o Alcohols, Ketones, and Solids	52% (34/65)	95% (20/21)	5% (1/21)	40% (4/10)	60% (6/10)	0% (0/10)	NA	NA	NA	76% (26/34)	24% (8/34)
w/o Alcohols	52% (56/107)	79% (26/33)	21% (7/33)	48% (10/21)	52% (11/21)	0% (0/21)	NA	NA	NA	64% (34/53)	36% (19/53)

Abbreviations: EU = European Union; BCOP = Bovine Corneal Opacity and Permeability

[1]EU classification system (EU 2001).

[2]Severe = R41.

[3]Moderate = R36.

[4]Mild = NA.

[5]Not Labeled = Not Classified.

Table 6-17 Accuracy of the BCOP Test Method for Distinguishing Not Labeled Substances from All Other Irritant Classes as Defined by the EU Classification System[1], with Exclusion of Discordant Chemical and Physical Classes

BCOP	N[2]	Accuracy		Sensitivity		Specificity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.
Overall	118	64	76/118	100	54/54	34	22/64	66	42/64	0	0/54
w/o Alcohols	103	62	64/103	100	45/45	33	19/58	67	39/58	0	0/45
w/o Ketones	109	66	72/109	100	52/52	35	20/57	65	37/57	0	0/52
w/o Solids	88	61	54/88	100	41/41	28	13/47	72	34/47	0	0/41
w/o Alcohols and Ketones	94	64	60/94	100	43/43	33	17/51	67	34/51	0	0/43
w/o Alcohols, Ketones and Solids	65	60	39/65	100	31/31	24	8/34	76	26/34	0	0/31
w/o Hydrocarbons	107	68	73/107	100	54/54	36	19/53	64	34/53	0	0/54

[1] EU classification system (EU 2001). NC vs. R41/R36.
[2]N = Number of substances included in this analysis/the total number of substances in the study.
[3]No. = Data used to calculate the percentage.

7.0 BCOP Test Method Reliability

An assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM 2003). Quantitative and qualitative evaluations of BCOP test method reliability have been conducted previously (ICCVAM, 2006a). However, additional qualitative analyses of test method reproducibility were conducted to evaluate the extent of agreement of BCOP hazard classifications among the laboratories.

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

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998).

Balls et al. (1995): Of 14 substances classified by the GHS as Not Labeled 4/14 (29%) were correctly identified while 2/4 (50%) GHS Category 2B substances were correctly identified, 3/14 (21%) substances classified as GHS Category 2A were correctly identified, and 17/22 (77%) GHS Category 1 substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing Not Classified substances from all other classes of 55/59 (93%) substances (**Table 7-1**).
- All five participating laboratories agreed on the classification of 13/17 (76%) substances that were correctly identified as GHS Category 1, 0/3 (0%) substances correctly classified as GHS Category 2A, 1/2 (50%) substances correctly classified as GHS Category 2B and 2/4 (50%) substances correctly classified as GHS Not Classified (**Table 7-2**).
- The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (76% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories) (**Table 7-2**).

- There was 100% agreement on the 10 false positive substances among the 5 laboratories.

Gautheron et al. (1994): Of 34 substances classified by the GHS as Not Labeled 13/34 (38%) were correctly identified while 0/2 (0%) GHS Category 2B substances were correctly identified, 1/3 (33%) substances classified as GHS Category 2A were correctly identified, and 6/8 (75%) GHS Category 1 substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 39/52 (75%) substances (**Table 7-1**).
- All five participating laboratories agreed on the classification of 4/6 (67%) substances that were correctly identified as GHS Category 1, 0/1 (0%) substances correctly classified as GHS Category 2A, 1/2 (50%) substances correctly classified as GHS Category 2B and 7/13 (54%) substance correctly classified as GHS Not labeled (**Table 7-2**).
- The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (67% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories) (**Table 7-2**).
- Of the 21 false positive substances, 17 (81%) were shown to have 100% agreement among the 5 laboratories.
-

Table 7-1 Evaluation of the Reliability of the BCOP Test Method in Predicting Not Labeled Ocular Substances or Corrosives/Severe/Moderate/Mild Irritants as Defined by the GHS Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73-75% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤55% Agreement among Labs
Balls et al. (1995)	+/+	5	38	37 (97%)			1 (3%)				
	+/-	5	1				1 (100%)				
	-/+	5	10	10 (100%)							
	-/-	5	4	2 (50%)			1 (25%)			1 (25%)	
	?/-	5	1	1 (100%)							
	?/+	5	5	5 (100%)							
	Total		59	55 (93%)			3 (5%)			1 (2%)	
Gautheron et al. (1994)	+/+	11 12	12	11 (92%)	1 (9%)						
	+/-	11 12	1								1 (100%)
	-/+	11 12	21	17 (81%)	1 (5%)				1 (5%)	2 (10%)	
	-/-	11 12	13	7 (54%)	1 (8%)			4 (31%)	1 (8%)		
	?/-	11 12	1		1 (100%)						
	?/+	11	4	4 (100%)							
	Total		52	39 (75%)	4 (8%)			4 (8%)	2 (4%)	2 (4%)	1 (2%)
Southee (1998)	+/+	3	10	10 (100%)							
	+/-	3	2								2 (100%)
	-/+	3	1	1 (100%)							
	-/-	3	2	2 (100%)							
	?/-	3	0								
	?/+	3	1	1 (100%)							
	Total		16	14 (88%)							2 (12%)

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

[2]n indicates number of substances.

[3]Number in parentheses indicates percentage of tested chemicals.

1812 **Table 7-2 Evaluation of the Interlaboratory Variability of the BCOP Test Method In Predicting Ocular Irritant Classes**
1813 **Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the GHS Classification System, by Study**

Study	*In vivo* Classification (No.)[1]	Classification *(in vitro)*	No. of Substances	Number of Testing Labs	Substances with 100% Agreement Among Laboratories (%)	Substances with 70-95% Agreement Among Laboratories (%)	Substances with 60-69% Agreement Among Laboratories (%)	Substances with <60% Agreement Among Laboratories (%)
Balls et al. (1995)	NI (14)	Actual	4	5	2 (50%)	1 (25%)	1 (25%)	-
		Over	10	5	10 (100%)	-	-	-
	2B (4)	Under	0	5	-	-	-	-
		Actual	2	5	1 (50%)	1 (50%)	-	-
		Over	2	5	1 (50%)	1 (50%)	-	-
	2A (14)	Under	2	5	2 (100%)	-	-	-
		Actual	3	5	-	1 (33%)	1 (33%)	1 (33%)
		Over	9	5	3 (33%)	3 (33%)	3 (33%)	-
	1 (22)	Under	5	5	3 (60%)	1 (20%)	1 (20%)	-
		Actual	17	5	13 (76%)	3 (18%)	1 (6%)	-
Gautheron et al. (1994)	NI (34)	Actual	13	11	7 (54%)	4 (31%)	2 (15%)	-
		Over	21	11	17 (81%)	1 (5%)	1 (5%)	2 (10%)
	2B (2)	Under	0	11	-	-	-	-
		Actual	0	11	-	-	-	-
		Over	2	11	1 (50%)	1 (50%)	-	-
	2A (3)	Under	0	11	-	-	-	-
		Actual	1	11	-	1 (100%)	-	-
		Over	2	11	1 (50%)	1 (50%)	-	-
	1 (8)	Under	2	11	1 (50%)	1 (50%)	-	-
		Actual	6	11	4 (67%)	1 (17%)	-	1 (17%)
Southee (1998)	NI (3)	Actual	2	3	2 (100%)	-		-
		Over	1	3	1 (100%)	-		-
	2B (3)	Under	1	3	-	-		1 (100%)
		Actual	1	3	1 (100%)	-		-
		Over	1	3	1 (100%)	-		-
	2A (2)	Under	0	3	-	-		-
		Actual	2	3	1 (50%)	1 (50%)		-

Study	*In vivo* Classification (No.)[1]	Classification *(in vitro)*	No. of Substances	Number of Testing Labs	Substances with 100% Agreement Among Laboratories (%)	Substances with 70-95% Agreement Among Laboratories (%)	Substances with 60-69% Agreement Among Laboratories (%)	Substances with <60% Agreement Among Laboratories (%)
		Over	0	3	-	-		-
	1 (7)	Under	3	3	3 (100%)	-		-
		Actual	4	3	4 (100%)	-		-

1814

1815

1816

Southee (1998): Of 3 substances classified by the GHS as Not Labeled, 2/3 (67%) were correctly identified while 1/3 (33%) GHS Category 2B substances were correctly identified, 2/2 (100%) substances classified as GHS Category 2A were correctly identified, and 4/7 (57%) GHS Category 1 substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 14/16 (88%) substances (**Table 7-1**).
- All five participating laboratories agreed on the classification of 4/4 (100%) substances that were correctly identified as GHS Category 1, 1/2 (50%) substances correctly classified as GHS Category 2A, 1/1 (100%) substances correctly classified as GHS Category 2B and 2/2 (100%) substances correctly classified as GHS Not labeled (**Table 7-2**).
- Of the 1 false positive substance, there was 100% agreement among the 5 laboratories.

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

Balls et al. (1995): Of the two substances classified by the EPA as Category IV, 2/2 (100%) were correctly identified while 6/20 (30%) EPA Category III substances were correctly identified, 4/13 (31%) substances classified as EPA Category II were correctly identified, and 13/18 (72%) EPA Category I substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 55/59 (93%) substances (**Table 7-3**).
- All five participating laboratories agreed on the classification of 10/13 (77%) substances that were correctly identified as EPA Category I, 0/4 (0%) substances correctly classified as EPA Category II, 4/6 (67%) substances correctly classified as EPA Category III and 1/2 (50%) substances correctly classified as EPA Category IV (**Table 7-4**).

Table 7-3 Evaluation of the Reliability of the BCOP Test Method In Predicting Not Labeled Ocular Substances or Corrosives/Severe/Moderate/Mild Irritants as Defined by the EPA Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with≤ 55% Agreement among Labs
Balls et al. (1995)	+/+	5	47	47 (1000%)							
	+/-	5	4	1 (25%)			1 (25%)			1 (25%)	1 (25%)
	-/+	5	0								
	-/-	5	2	1 (50%)			1 (50%)				
	?/-	5	1	1 (100%)							
	?/+	5	5	5 (100%)							
	Total		59	55 (93%)			2 (3%)			1 (2%)	1 (2%)
Gautheron et al. (1994)	+/+	11 12	28	26 (93%)	1 (4%)					1 (4%)	
	+/-	11 12	7	1 (14%)				2 (29%)	1 (14%)		3 (43%)
	-/+	11 12	3	3 (100%)							
	-/-	11 12	10	6 (60%)	1 (10%)			2 (20%)			1 (10%)
	?/-	11 12	1	-	1 (100%)						
	?/+	11	3	3 (100%)							
	Total		52	39 (75%)	3 (6%)			4 (8%)	1 (2%)	1 (2%)	4 (8%)
Southee (1998)	+/+	3	10	10 (100%)							
	+/-	3	3	1 (33%)							2 (67%)
	-/+	3	0								
	-/-	3	1	1 (33%)							
	?/-	3	0								
	?/+	3	2	2 (67%)							
	Total		16	14 (88%)							

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or Not Labeled (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too

1849 early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 6.1** for a description of the rules followed to classify the
1850 ocular irritancy of test substances tested multiple times *in vitro*.
1851 [2]n indicates number of substances.
1852 [3]Number in parentheses indicates percentage of tested chemicals.
1853

1854

Table 7-4 Evaluation of the Interlaboratory Variability of the BCOP Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the EPA Classification System, by Study

Study	*In vivo* Classification (No.)[1]	Classification *(in vitro)*	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 80-92% Agreement Among Laboratories (%)	Substances with 61-79% Agreement Among Laboratories (%)	Substances with 50-60% Agreement Among Laboratories (%)	Substances with <50% Agreement Among Laboratories (%)
Balls et al. (1995)	IV (2)	Actual	2	5	1 (50%)	1 (50%)	-	-	-
		Over	0	5	-	-	-	-	-
	III (20)	Under	2	5	1 (50%)	-	-	1 (50%)	-
		Actual	6	5	4 (67%)	1 (17%)	-	1 (17%)	-
		Over	4	5	1 (25%)	1 (25%)	-	2 (50%)	-
	II (13)	Under	2	5	2 (100%)	-	-	-	-
		Actual	4	5	-	1 (25%)	-	2 (50%)	1 (25%)
		Over	7	5	4 (57%)	1 (14%)	-	2 (28%)	-
	I (18)	Under	5	5	3 (60%)	1 (20%)	-	1 (20%)	-
		Actual	13	5	10 (77%)	2 (15%)	-	1 (8%)	-
Gautheron et al. (1994)	IV (13)	Actual	10	11/12	9 (90%)	-	-	-	1 (10%)
		Over	3	11/12	3 (100%)	-	-	-	-
	III (23)	Under	5	11/12	-	-	-	-	-
		Actual	7	11/12	2 (29%)	3 (43%)	1 (14%)	1 (14%)	-
		Over	11	11/12	9 (82%)	2 (18%)	-	-	-
	II (5)	Under	1	11/12	-	-	-	-	-
		Actual	1	11/12	-	1 (100%)	-	-	-
		Over	3	11/12	1 (33%)	2 (67%)	-	-	-
	I (7)	Under	2	11/12	1 (50%)	1 (50%)	-	-	-
		Actual	5	11/12	3 (60%)	1 (20%)	-	1 (20%)	-
Southee (1998)	IV (1)	Actual	1	5	1 (100%)	-	-	-	-
		Over	0	5	-	-	-	-	-
	III (6)	Under	2	5	1 (50%)		-	1 (50%)	-
		Actual	2	5	2 (100%)	-	-	-	-
		Over	2	5	2 (100%)	-	-	-	-
	II (2)	Under	0	5	-	-	-	-	-

Study	***In vivo* Classification (No.)[1]**	**Classification *(in vitro)***	**Number of Substances**	**Number of Testing Laboratories**	**Substances with 100% Agreement Among Laboratories (%)**	**Substances with 80-92% Agreement Among Laboratories (%)**	**Substances with 61-79% Agreement Among Laboratories (%)**	**Substances with 50-60% Agreement Among Laboratories (%)**	**Substances with <50% Agreement Among Laboratories (%)**
		Actual	1	5	1 (100%)	-	-	-	-
		Over	1	5	1 (100%)	-	-	-	-
	I (5)	Under	3	5	3 (100%)	-	-	-	-
		Actual	2	5	2 (100%)	-	-	-	-

[1]Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), a EPA classification could not be made for 2 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*.

- The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (92% of the accurately identified severe substances were shown to have 80% - 100% classification agreement among testing laboratories) (**Table 7-4**).

Gautheron et al. (1994): Of 13 substances classified by the EPA as Category IV, 10/13 (77%) were correctly identified while 7/23 (30%) EPA Category III substances were correctly identified, 1/5 (20%) substances classified as EPA Category II were correctly identified, and 5/7 (71%) EPA Category I substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 39/52 (75%) substances (**Table 7-3**).
- All five participating laboratories agreed on the classification of 3/5 (60%) substances that were correctly identified as EPA Category I, 0/1 (0%) substances correctly classified as EPA Category II, 2/7 (29%) substances correctly classified as EPA Category III and 9/10 (90%) substances correctly classified as EPA Category IV (**Table 7-4**).
- Of the 3 false positive substances, 3 (100%) were shown to have 100% agreement among the 5 laboratories.

Southee (1998): Of the 1 substance classified by the EPA as Category IV, 1/1 (100%) were correctly identified while 2/6 (33%) EPA Category III substances were correctly identified, 1/2 (50%) substances classified as EPA Category II were correctly identified, and 2/5 (40%) EPA Category I substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 14/16 (88%) substances (**Table 7-3**).

- All five participating laboratories agreed on the classification of 2/2 (100%) substances that were correctly identified as EPA Category I, 1/1 (100%) substances correctly classified as EPA Category II, 2/2 (100%) substances correctly classified as EPA Category III and 1/1 (100%) substance correctly classified as EPA Category IV (**Table 7-4**).

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

Balls et al. (1995): Of 17 substances classified by the EU as Not Labeled, 4/17 (24%) were correctly identified while 6/14 (43%) EU Category R36 substances were correctly identified, and 14/22 (64%) EU R41 substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 55/59 (93%) substances (**Table 7-5**).
- All five participating laboratories agreed on the classification of 12/14 (86%) substances that were correctly identified as EU R41, 2/6 (33%) substances correctly classified as EU Category R36, and 2/4 (50%) substances correctly classified as EU Not Labeled (**Table 7-6**).
- The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (100% of the accurately identified severe substances were shown to have 95% - 100% classification agreement among testing laboratories) (**Table 7-6**).
- Of the 13 false positive substances, 13 (100%) were shown to have 100% agreement among the 5 laboratories.

Gautheron et al. (1994): Of 36 substances classified by the EU as Not Labeled, 13/36 (36%) were correctly identified while 2/4 (50%) EU Category R36 substances were correctly identified, and 6/8 (75%) EU R41 substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 39/52 (75%) substances (**Table 7-5**).
- All five participating laboratories agreed on the classification of 4/6 (67%) substances that were correctly identified as EU R41, 0/2 (0%) substances correctly classified as EU Category R36, and 7/13 (54%) substances correctly classified as EU Not Labeled (**Table 7-6**).
- Of the 23 false positive substances, 20/23 (87%) were shown to have 91% - 100% agreement among the 5 laboratories.

Southee (1998): Of the 4 substances classified by the EU as Not Labeled, 2/4 (50%) were correctly identified while 2/4 (50%) EU Category R36 substances were correctly identified, and 4/6 (67%) EU R41 substances were correctly identified.

- The five participating laboratories were in 100% agreement to the ocular irritancy classification when assessing non labeled substances from all other classes of 14/16 (88%) substances (**Table 7-5**).
- All five participating laboratories agreed on the classification of 4/4 (100%) substances that were correctly identified as EU R41, 2/2 (100%) substances correctly classified as EU Category R36, and 2/2 (100%) substances correctly classified as EU Not Labeled (**Table 7-6**).
- Of the 2 false positive substances, all were shown to have 100% agreement among the 5 laboratories (**Table 7-6**).

Table 7-5 Evaluation of the Reliability of the BCOP Test Method In Predicting Not Labeled Ocular Substances or Corrosives/Severe/Moderate Irritants as Defined by the EU Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤55% Agreement among Labs
Balls et al. (1995)	+/+	5	31	100% (31/31)							
	+/-	5	2				100% (2/2)				
	-/+	5	13	100% (13/13)							
	-/-	5	4	50% (2/4)			25% (1/4)			25% (1/4)	
	?/-	5	1	100% (1/1)							
	?/+	5	8	100% (8/8)							
	Total		59	93% (55/59)			5% (3/60)			2% (1/60)	
Gautheron et al. (1994)	+/+	11 12	11	10 (9%)	1 (9%)						
	+/-	11 12	1								1 (100%)
	-/+	11 12	23	19 (95%)	1 (4%)					2 (9%)	1 (4%)
	-/-	11 12	13	7 (44%)				4 (57%)			2 (15%)
	?/-	11 12	1		1 (100%)						
	?/+	11	3	3 (100%)							
	Total		52	39 (75%)	3 (6%)			4 (8%)		2 (4%)	4 (8%)
Southee (1998)	+/+	3	8	8 (100%)							
	+/-	3	2								2 (100%)
	-/+	3	2	2 (100%)							
	-/-	3	2	2 (100%)							
	?/-	3	0								
	?/+	-	2	2 (100%)							
	Total		16	14 (88%)							2 (13%)

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

[2]n indicates number of substances.

[3]Number in parentheses indicates percentage of tested chemicals.

Table 7-6 Evaluation of the Interlaboratory Variability of the BCOP Test Method In Predicting Ocular Irritant Classes Compared to the *In Vivo* Rabbit Eye Test Method as Defined by the EU Classification System, by Study

Study	*In vivo* Classification (No.)[1]	Classification *(in vitro)*	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 76-95% Agreement Among Laboratories (%)	Substances with 50-75% Agreement Among Laboratories (%)
Balls et al. (1995)	NI (17)	Actual	4	5	2 (50%)	1 (25%)	1 (25%)
		Over	13	5	13 (100%)	-	-
	R36 (14)	Under	0	5	-	-	-
		Actual	6	5	2 (33%)	2 (33%)	2 (33%)
		Over	8	5	4 (50%)	1 (13%)	3 (38%)
	R41 (22)	Under	5	5	3 (60%)	1 (20%)	1 (20%)
		Actual	14	5	12 (86%)	2 (14%)	-
Gautheron et al. (1994)	NI (36)	Actual	13	11	7 (54%)	2 (15%)	4 (31%)
		Over	23	11	19 (83%)	1 (4%)	3 (13%)
	R36 (4)	Under	0	11	-	-	-
		Actual	2	11	-	1 (50%)	1 (50%)
		Over	2	11	1 (50%)	1 (50%)	-
	R41 (8)	Under	2	11	1 (50%)	1 (50%)	-
		Actual	6	11	4 (67%)	1 (17%)	1 (17%)
Southee (1998)	NI (4)	Actual	2	3	2 (100%)	-	-
		Over	2	3	2 (100%)	-	-
	R36 (4)	Under	1	3	-	-	1 (100%)
		Actual	2	3	2 (100%)	-	-
		Over	1	3	1 (100%)	-	-
	R41 (6)	Under	2	3	2 (100%)	-	-
		Actual	4	3	4 (100%)	-	-

8.0 BCOP Test Method Data Quality

8.1 Adherence to National and International GLP Guidelines

The evaluation of BCOP test method data quality included in the original evaluation of the BCOP is detailed in ICCVAM (2006a). As indicated in Section 8.0 of the AMCP BRD submission, it could not be ascertained as to whether all of the *in vitro* data contained in this BRD were generated under full GLP compliance, but where possible, that information is contained in the spreadsheets that form the database from which this BRD was generated. All of the new *in vitro* data that were generated during the course of constructing this BRD were conducted with full GLP compliance.

1957 ***[This Page Intentionally Left Blank]***

1958

1959

1960

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Reports in the Peer Reviewed Literature

Since the previous evaluation of the BCOP method in identifying ocular corrosives and severe irritants (ICCVAM, 2006a), a total of four BCOP studies have been located among the peer reviewed literature. A search of MEDLINE, TOXLINE and Web of Science showed 14 additional scientific publications with BCOP test method results and four additional references containing BCOP data (Debbasch et al. [2005], Van Goethem et al. [2005], Cater and Harbell –[2006], and Cater and Harbell [2008]). A total of four publications were identified containing BCOP test method analyses, however, none of these publications included raw data and as such were not added to the database.

In Debbasch et al. (2005), 12 make-up removers were tested both in the BCOP and in a clinical in-use test under ophthalmological control after their application to the external eyelid. The undiluted test product (750 μL) was pipetted onto the corneas and exposure conducted for four hours. Corneal opacity was determined using an adapted spectrophotometer and barrier disruption by fluorescein uptake using OD490 nm. *In vitro* scores were classified according to Gautheron et al. (1994) and Harbell and Curren (1998), but no *in vivo* rabbit eye data were reported, and these data have not be obtained. For this reason, this study was not included in the BCOP performance analyses detailed in this BRD.

In Cater and Harbell (2006), surfactant based "rinse-off" personal care formulations were tested in the BCOP test method, using slight modifications of the BCOP protocol reported by Sina et al. (1995). Corneas were exposed to the test substances (750 μL) for 10, 30 or 60 minutes either undiluted or diluted in deionized water. Corneas were evaluated for opacity, fluorescein uptake and histological alterations. No *in vivo* rabbit reference data were reported and thus this study was not included in the BCOP performance analyses detailed in this BRD.

In Goethem et al. (2006), 20 substances (7 compounds classified as GHS Not Classified and 13 GHS Category 1), were tested in the BCOP test method. These results were previously published in Vanparys et al. (1993) and Gautheron et al. (1994), which were included in the previous BCOP BRD (ICCVAM, 2006a).

In Cater and Harbell (2008), the BCOP test method was conducted on four commercial and one unregistered body wash developed for children or as mild bath products. The purpose of this testing was to determine if the BCOP test method could be used as a prediction model for relative ranking of human eye irritation responses under conditions of a standard human eye sting test to surfactant-based formulations. Test articles were prepared as 25% solutions in deionized water, 750 μL applied to the corneas and exposure conducted for 30 minutes. Following exposure, opacity and fluorescein uptake determined *in vitro*, but no *in vivo* rabbit eye data was reported.

10.0 Animal Welfare Considerations (Refinement, Reduction, And Replacement)

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

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

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

[This Page Intentionally Left Blank]

11.0 Practical Considerations

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

[This Page Intentionally Left Blank]

12.0 References

AMCP Background Review Document (BRD). 2008. Submission to ICCVAM: An *in vitro* approach of EPA toxicity labeling of anti-microbial cleaning products and annexes A-G (http://iccvam.niehs.nih.gov/methods/ocutox/antimicro/BRD.pdf).

Bailey PT, Freeman JJ, Phillips RD, Merrill JC. 2004. Validation of the BCOP assay as a predictor of ocular irritation of various petrochemical products. Poster presentation at the Society of Toxicology 2004 meeting.

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.

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

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

Carpenter CP, Smyth HF. 1946. Chemical burns of the rabbit cornea. Am J Opthalmol. 29:60-73.

Cater KC, and Harbell JW. 2006. Prediction of eye irritation potential of surfactant-based rinse-off personal care formulations by the bovine corneal opacity and permeability (BCOP) assay. Cut and Ocular Toxicol 25:217-233.

Cater KC, and Harbell JW. 2008. Comparison of *in vitro* eye irritation potential by bovine corneal opacity and permeability (BCOP) assay to erythema scores in human eye sting test of surfactant-based formulations. Cut and Ocular Toxicol 27:77-85.

CPSC. 2003. Commercial practices. Hazardous Substances and Articles. 16CFR1500.42.

Debbasch C, Ebenhahn C, Dami N, Pericoi M, Van den Berghe C, Cottin M, and Nohynek GJ. 2005. Eye irritation of low-irritant cosmetic formulations: correlation of *in vitro* results with clinical data and product comparison. Food and Chem Toxicol. 43:155-165.

Draize J, Woodard G, Calvery H. 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J Pharmacol Exp Ther 82:377-390.

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

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

EEC. 1991. Classification of irritant substances and preparations. Official Journal of the European Community. L180, 52.

EPA. 1996. Label Review Manual. 2^{nd} Edition. EPA737-B-96-001. U.S. Environmental Protection Agency:Washington, DC.

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

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

EPA. 2003b. Good laboratory practice standards. 40CFR160.

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

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

Gautheron P, Giroux J, Cottin M, Audegond L, Morilla A, Mayordomo-Blanco L, Tortajada A, Haynes G, Vericat JA, Pirovano R, Tos EG, Hagemann C, Vanparys P, Deknudt G, Jacobs G, Prinsen M, Kalweit S, Spielmann H. 1994. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) assay. Toxicol In Vitro 8(3):381-392.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park:National Toxicology Program.

ICCVAM Authorization Act of 2000. 2000. Public Law 106-545 [114 Stat. 2721]. Available: http://iccvam.niehs.nih.gov/about/overview.htm [accessed 21 October 2004].

ICCVAM. 2006a. Background Review Document: Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability Test Method. NIH Publication No: 06-4512. Research Triangle Park: National Toxicology Program.

ICCVAM. 2006b. ICCVAM Test Method Evaluation Report: *In Vitro* Ocular Toxicity Methods for Identifying Severe Irritants and Corrosives. NIH Publication No: 07-4517. Research Triangle Park: National Toxicology Program.

Kay JH, Calandra JC. 1962. Interpretation of eye irritation tests. J Soc Cosmet Chem 13:281-289.

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD principles on Good Laboratory Practice (as revised in 1997). ENV/MC/CHEM(98)17. OECD, Paris.

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

Prinsen MK, and Koeter HB. Justification of the enucleated eye test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. 1993. Food Chem Toxicol. 31(1):69-76.

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

Russell WMS, Burch RL. 1992. The principles of humane experimental technique. 14th ed. Universities Federation for Animal Welfare: South Mimms, Potters Bar, Herts, England.

Sina JF, Galer DM, Sussman RG, Gautheron PD, Sargent EV, Leong B, Shah PV, Curren RD, Miller K. 1995. A collaborative evaluation of seven alternatives to the Draize eye irritation test using pharmaceutical intermediates. Fundam Appl Toxicol 26:20-31.

Southee JA. 1998. Evaluation of the Prevalidation Process. Part 2, final report. Volume 2. The Bovine Corneal Opacity and Permeability (BCOP) Assay. European Community Contract No. 11279-95-10F 1ED ISP GB.

Swanson JE, Harbell JW. 2000. Evaluating the eye irritancy potential of ethanolic test materials with the bovine corneal opacity and permeability assay. The Toxicologist 54(1):188-189.

Swanson JE, Lake LK, Donnelly TA, Harbell JW, Huggins J. 1995. Prediction of ocular irritancy of full-strength cleaners and strippers by tissue equivalent and bovine corneal assays. J Toxicol - Cut and Ocular Toxicol 14(3):179-195.

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

Van Goethem F, Adriaens E, Alepee N, Straube F, De Wever B, Cappadoro M, Catoire S, Hansen E, Wolf A, and Vanparys P. 2006. Prevalidation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. Toxicol in Vitro. 20:1-17.

2128 ***[This Page Intentionally Left Blank]***

2129

2130

2131

13.0 GLOSSARY[9]

Accuracy[10]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance" (see also "two-by-two" table). Accuracy is highly dependent on the prevalence of positives in the population being examined.

Assay[2]: The experimental system used. Often used interchangeably with "test" and "test method."

Benchmark substance: A substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties:
a consistent and reliable source(s)
structural and functional similarity to the class of substances being tested
known physical/chemical characteristics
supporting data on known effects
known potency in the range of the desired response

Benchmark control: A sample containing all components of a test system and treated with a known substance (i.e., the benchmark substance) to induce a known response. The sample is processed with test substance-treated and other control samples to compare the response produced by the test substance to the benchmark substance to allow for an assessment of the sensitivity of the test method to assess a specific chemical class or product class.

Blepharitis: Inflammation of the eyelids.

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

---

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

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

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

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

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

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

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

Concordance[2]: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of "relevance". The term is often used interchangeably with "accuracy" (see also "two-by-two" table). Concordance is highly dependent on the prevalence of positives in the population being examined.

Conjunctiva: The mucous membrane that lines the inner surfaces of the eyelids and folds back to cover the front surface of the eyeball, except for the central clear portion of the outer eye (the cornea). The conjunctiva is composed of three sections: palpebral conjunctiva, bulbar conjunctiva, and fornix.

Conjunctival sac: The space located between the eyelid and the conjunctiva-covered eyeball. Substances are instilled into the sac to conduct an *in vivo* eye test.

Cornea: The transparent part of the coat of the eyeball that covers the iris and pupil and admits light to the interior.

Corneal opacity: Measurement of the extent of opaqueness of the cornea following exposure to a test substance. Increased corneal opacity is indicative of damage to the cornea. Opacity

can be evaluated subjectively as done in the Draize rabbit eye test, or objectively with an instrument such as an "opacitometer."

Corneal permeability: Quantitative measurement of damage to the corneal epithelium by a determination of the amount of sodium fluorescein dye that passes through all corneal cell layers.

Corrosion: Destruction of tissue at the site of contact with a substance.

Corrosive: A substance that causes irreversible tissue damage at the site of contact.

Endpoint[2]: The biological process, response, or effect assessed by a test method.

False negative[2]: A substance incorrectly identified as negative by a test method.

False negative rate[2]: The proportion of all positive substances falsely identified by a test method as negative (see "two-by-two" table). It is one indicator of test method accuracy.

False positive[2]: A substance incorrectly identified as positive by a test method.

False positive rate[2]: The proportion of all negative substances that are falsely identified by a test method as positive (see "two-by-two" table). It is one indicator of test method accuracy.

Fibrous tunic: The outer of the three membranes of the eye, comprising the cornea and the sclera; also called *tunica fibrosa oculi*.

Globally Harmonized System (GHS): A classification system presented by the United Nations that provides (a) a harmonized criteria for classifying substances and mixtures according to their health, environmental and physical hazards, and (b) harmonized hazard communication elements, including requirements for labeling and safety data sheets.

Good Laboratory Practices (GLP)[2]: Regulations promulgated by the U.S. Food and Drug Administration and the U.S. Environmental Protection Agency, and principles and procedures adopted by the Organization for Economic Cooperation and Development, and Japanese authorities that describe record keeping and quality assurance procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Hazard[2]: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Interlaboratory reproducibility[2]: A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test method can be transferred successfully among laboratories.

Intralaboratory repeatability[2]: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Intralaboratory reproducibility[2]: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

In vitro: In glass. Refers to assays that are carried out in an artificial system (e.g., in a test tube or petri dish) and typically use single-cell organisms, cultured cells, cell-free extracts, or purified cellular components.

***In Vitro* Irritancy Score:** An empirically-derived formula used in the BCOP assay whereby the mean opacity and mean permeability values for each treatment group are combined into a single *in vitro* score for each treatment group. The *In Vitro* Irritancy Score = mean opacity value + (15 x mean permeability value).

In vivo : In the living organism. Refers to assays performed in multicellular organisms.

Iris: The contractile diaphragm perforated by the pupil and forming the colored portion of the eye.

Negative control: An untreated sample containing all components of a test system, except the test substance solvent, which is replaced with a known nonreactive material, such as water. This sample is processed with test substance-treated samples and other control samples to determine whether the solvent interacts with the test system.

Negative predictivity[2]: The proportion of correct negative responses among substances testing negative by a test method (see "two-by-two" table). It is one indicator of test method accuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

Neuroectodermal tunic: The innermost of three membranes of the eye, comprising the retina.

Nictating (nictitating) membrane: The membrane that moves horizontally across the eye in some animal species (e.g., rabbit, cat) to provide additional protection in particular circumstances. It may be referred to as the "third eyelid."

Not Labeled: (a) A substance the produces no changes in the eye following application to the anterior surface of the eye. (b) Substances that are not classified as GHS Category 1, 2A, or 2B; or EU R41 or R36 ocular irritants.

Nonsevere irritant: (a) A substance that causes tissue damage in the eye following application to the anterior surface of the eye; the tissue damage is reversible within 21 days of application and the observed adverse effects in the eye are less severe than observed for a severe irritant. (b) Substances that are classified as GHS Category 2A or 2B; EPA Category II, III, or IV; or EU R36 ocular irritants.

Ocular: Of or relating to the eye.

Ocular corrosive: A substance that causes irreversible tissue damage in the eye following application to the anterior surface of the eye.

Ocular irritant: A substance that produces a reversible change in the eye following application to the anterior surface of the eye.

Opacitometer: An instrument used to measure "corneal opacity" by quantitatively evaluating light transmission through the cornea. The instrument has two compartments, each with its own light source and photocell. One compartment is used for the treated cornea, while the other is used to calibrate and zero the instrument. The difference between photocell signals in the two compartments is measured electronically as a change in voltage, and is displayed digitally, generating numerical opacity values with arbitrary units.

Palpebral conjunctiva: The part of the conjunctiva that covers the inner surface of the eyelids.

Pannus: A specific type of corneal inflammation that begins within the conjunctiva, and with time spreads to the cornea. Also referred to as "chronic superficial keratitis."

Performance[2]: The accuracy and reliability characteristics of a test method (see "accuracy", "reliability").

pH: A measure of the acidity or alkalinity of a solution. pH 7.0 is neutral; higher pHs are alkaline, lower pHs are acidic.

Positive control: A sample containing all components of a test system and treated with a substance known to induce a positive response, which is processed with the test substance-treated and other control samples to demonstrate the sensitivity of each experiment and to allow for an assessment of variability in the conduct of the assay over time.

Positive predictivity[2]: The proportion of correct positive responses among substances testing positive by a test method (see "two-by-two" table). It is one indicator of test method accuracy. Positive predictivity is a function of the sensitivity of the test method and the prevalence of positives among the substances tested.

Prevalence[2]: The proportion of positives in the population of substances tested (see "two-by-two" table).

Protocol[2]: The precise, step-by-step description of a test method, including a listing of all necessary reagents, criteria and procedures for evaluation of the test data.

Quality assurance[2]: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures is assessed independently by individuals other than those performing the testing.

Reduction alternative[2]: A new or modified test method that reduces the number of animals required.

Reference test method[2]: The accepted *in vivo* test method used for regulatory purposes to evaluate the potential of a test substance to be hazardous to the species of interest.

Refinement alternative[2]: A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals, or enhances animal well-being.

Relevance[2]: The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the "accuracy" or "concordance" of a test method.

Reliability[2]: A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

Replacement alternative[2]: A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility[2]: The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

Sclera: The tough, fibrous tissue that extends from the cornea to the optic nerve at the back of the eye.

Sensitivity[2]: The proportion of all positive substances that are classified correctly as positive in a test method. It is a measure of test method accuracy (see "two-by-two" table).

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

Severe irritant: (a) A substance that causes tissue damage in the eye following application to the anterior surface of the eye that is not reversible within 21 days of application or causes serious physical decay of vision. (b) Substances that are classified as GHS Category 1, EPA Category I, or EU R41 ocular irritants.

Solvent control: An untreated sample containing all components of a test system, including the solvent that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the

same solvent. When tested with a concurrent negative control, this sample also demonstrates whether the solvent interacts with the test system.

Specificity[2]: The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy (see "two-by-two" table).

Test[2]: The experimental system used; used interchangeably with "test method" and "assay."

Test method[2]: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay." See also "validated test method" and "reference test."

Test method component: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures.

Tiered testing: A testing strategy where all existing information on a test substance is reviewed, in a specified order, prior to *in vivo* testing. If the irritancy potential of a test substance can be assigned, based on the existing information, no additional testing is required. If the irritancy potential of a test substance cannot be assigned, based on the existing information, a step-wise animal testing procedure is performed until an unequivocal classification can be made.

Toxic keratoconjunctivitis: Inflammation of the cornea and conjunctiva due to contact with an exogenous agent. Used interchangeably with "contact keratoconjunctivitis, irritative keratoconjunctivitis, and chemical keratoconjunctivitis."

Transferability[2]: The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two table[2]: The two-by-two table can be used for calculating accuracy (concordance) ([a+d]/[a+b+c+d]), negative predictivity (d/[c+d]), positive predictivity (a/[a+b]), prevalence ([a+c]/[a+b+c+d]), sensitivity (a/[a+c]), specificity (d/[b+d]), false positive rate (b/[b+d]), and false negative rate (c/[a+c]).

		New Test Outcome		
		Positive	**Negative**	**Total**
Reference Test Outcome	**Positive**	a	c	a + c
	Negative	b	d	b + d
	Total	a + b	c + d	a + b + c + d

Uvea tract: The middle of three membranes of the eye, comprising the iris, ciliary body, and choroid. Also referred to as the "vascular tunic."

Validated test method[2]: An accepted test method for which validation studies have been completed to determine the relevance and reliability of this method for a specific proposed use.

Validation[2]: The process by which the reliability and relevance of a procedure are established for a specific purpose.

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

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