HTML

NIH Publication No: 06-4512

BACKGROUND REVIEW DOCUMENT

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

Bovine Corneal Opacity and Permeability Test Method

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

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

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

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

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

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

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

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

The following Federal regulatory and research agencies are ICCVAM members:

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

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

Department of Transportation
Environmental Protection Agency

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

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

Bovine Corneal Opacity and Permeability Test Method

Background Review Document

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

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

March 2006
NIH Publication No. 06-4512

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

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

Printed: 08/08/2006

TABLE OF CONTENTS

Page Number

LIST OF TABLES

Page Number

LIST OF FIGURES

Page Number

LIST OF ACRONYMS AND ABBREVIATIONS

ANOVA	Analysis of variance
ASTM	American Society for Testing and Materials
BCOP	Bovine Corneal Opacity and Permeability
BIBRA	The British Industrial Biological Research Association
BRD	Background Review Document
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service Registry Number
CCT	Central corneal thickness
CGRP	Calcitonin Gene Related Peptide
CTAB	Cetyltrimethylammonium bromide
CTFA	Cosmetic, Toiletry, and Fragrance Association
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
EDTA	Ethylenediaminetetraacetic acid
EEC	European Economic Community
EPA	(U.S.) Environmental Protection Agency
EU	European Union
FBS	Fetal bovine serum
FDA	(U.S.) Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide and Rodenticide Act
FFDCA	Federal Food, Drug, and Cosmetic Act
FHSA	Federal Hazardous Substances Act
FR	*Federal Register*
g	Gram

GHS	Globally Harmonized System (of Classification and Labeling of Chemicals)
GLP	Good Laboratory Practices
HBSS	Hanks' Balanced Salt Solution
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
INVITTOX	*In Vitro* Techniques in Toxicology Database
IRAG	Interagency Regulatory Alternatives Group
IRE	Isolated Rabbit Eye
IVIS	*In Vitro* Irritancy Score
K-C	Kay and Calandra
kg	Kilogram
μg	Microgram
μL	Microliter
μm	Micrometer
MAS	Maximum average score
MEM	Minimum Essential Medium
MeSH	(National Library of Medicine) Medical subject heading
mg	Milligram
mL	Milliliter
mm	Millimeter
MMAS	Modified maximum average score
MMPs	Matrix metalloproteinases
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
PCOP	Porcine corneal opacity and permeability
P.L.	Public Law
QA	Quality assurance
QSAR	Quantitative structure activity relationship
r	rho (correlation coefficient)
SD	Standard deviation
SDS	Sodium dodecyl sulfate
STN	Science and Technology Information Network
TG	Test Guideline
TSCA	Toxic Substances Control Act
UN	United Nations
UV/VIS	Ultraviolet/visible
v/v	Volume to volute ratio
WHO	World Health Organization
w/v	Weight to volume ratio
w/w	Weight to weight ratio

ACKNOWLEDGEMENTS

The following individuals are acknowledged for their contributions to the in vitro ocular test method review process:

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

Consumer Product Safety Commission (CPSC)

Marilyn Wind, Ph.D.
Kailash Gupta, Ph.D.

Department of Defense (DOD)

Harry Salem, Ph.D.

Department of Transportation (DOT)

Steve Hwang, Ph.D.

Environmental Protection Agency (EPA)

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

Food and Drug Administration (FDA)

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

National Institute for Occupational Safety & Health (NIOSH)

Surender Ahir, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)

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

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

David Allen, Ph.D.
ILS, Inc.

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

Sue Brenzel
ILS, Inc.

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

Patricia Ceger, M.S.
ILS, Inc.

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

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

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

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

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

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

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

James Truax, M.A.
ILS, Inc.

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

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

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.

PREFACE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

EXECUTIVE SUMMARY

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

The information summarized in this BRD is based on publications obtained from the peer-reviewed literature, as well as unpublished information submitted to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in response to two *Federal Register* (*FR*) Notices requesting high quality *in vivo* rabbit eye test data and *in vitro* ocular irritation data for BCOP, the Isolated Chicken Eye (ICE), the Isolated Rabbit Eye (IRE), and the Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) test methods. An online literature search identified 18 publications that contained BCOP test method results and protocol information; of these publications, detailed *in vivo* data were obtained for five studies. Submitted BCOP and detailed *in vivo* data for three additional studies allowed for an evaluation of test method accuracy[2] and reliability[3] for a total of eight studies.

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

The BCOP assay is an *in vitro* eye irritation test method using isolated bovine eyes from cattle that have been slaughtered for meat or other purposes. In the BCOP assay, opacity is determined by the amount of light transmission 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. More recent additions/endpoints to the BCOP assay are assessment of corneal swelling or hydration, and histological assessment of morphological alterations in the

[1] Exposure of the isolated bovine cornea to irritants can produce opacity and/or an increase in permeability to sodium fluorescein dye. Both of these endpoints can be quantified and used to evaluate the potential eye irritation of substances.

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

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

cornea (Bruner et al. 1998; Ubels et al. 1998; Cooper et al. 2001; Jones et al. 2001). When histological assessment is added to the BCOP assay, the type and depth of corneal injury can be evaluated, as well as whether the tissue damage is permanent (e.g., damage to the endothelium) (Gran et al. 2003). Therefore, a histopathological assessment can be useful to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category) or to identify ocular damage that does not produce opacity or permeability in isolated cornea. Histology also is used for new chemistries or formulas that have not been well characterized in the BCOP assay, for known chemistries with delayed effects or where the mode of action cannot be easily predicted, and for known chemistries when a complete characterization of damage is needed.

U.S. Federal regulatory agencies were surveyed to determine whether BCOP test method data have been considered for regulatory use where submission of testing data is required. The EPA and the Food and Drug Administration (FDA) responded that BCOP data have been submitted to their respective agencies. The EPA Office of Pesticide Programs (OPP) received and reviewed BCOP data submitted in support of two new products (formulations). A labeling decision was made by the EPA for the two new products; however, hazard classification and labeling of these products was not based solely on the results of the submitted BCOP data. The FDA Center for Drug Evaluation and Research (CDER) has accepted BCOP data, on a case-by-case basis, for topically-applied products and more than 25 oral and inhalation products, but not for any ocular products. These substances or products were not formally classified for ocular irritation potential by the FDA.

The BCOP assay is currently used by several U.S. and European companies (e.g., pharmaceutical, personal care, and household cleaning product companies) as an in-house screen to assess the ocular irritation potential of a wide range of substances for which there could be accidental exposures in the workplace or home. The test method is used in the following ways:

1. For workplace safety applications to assess the irritancy of synthetic intermediates, various ingredients of a product, or the final product during the manufacturing process (Sina 1994).
2. For product safety applications to assess cosmetics, pharmaceuticals, soaps, household and industrial cleaners, personal care products, and other types of product formulations (Swanson et al. 1995; Casterton et al. 1996; Chamberlain et al. 1997; Harbell and Curren 1998; Cater et al. 2002; Cuellar et al. 2003; Bailey et al. 2004).

For example, it has been reported that some companies perform the assay as an in-house screen of industrial raw materials and intermediates; materials that give a BCOP score of 25 or higher are labeled as irritants with no further testing. Materials considered nonirritating based on the BCOP assay are tested *in vivo* to confirm the *in vitro* results (Chamberlain et al. 1997). In another company, the BCOP assay is used to evaluate both non-registered household products and registered household disinfectants, pesticides and repellents (Cuellar N and Swanson J, personal communication). For non-registered household products, BCOP data from new product formulations are usually matched with relevant benchmark formulations for which the ocular irritation potential is well characterized; *in vivo*

confirmatory testing is generally not performed. For registered products, use of the BCOP assay is limited to product development issues and worker safety at this company.

The BCOP test method protocols used in the various studies considered in this BRD are similar, but not identical. The essential principles of the test method protocol include isolating and culturing the bovine cornea, treating the isolated cornea with a test substance, collecting opacity and permeability data, and evaluating the data in relation to a prediction model. However, given the various uses and applications of the BCOP test method by different investigators and laboratories, and the evolution of the test method over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method. Variations in the publicly available BCOP protocols include different instrumentation to evaluate opacity, different prediction models or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations.

A total of 161 substances and formulations were evaluated in the eight studies, of which 69 were 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 detergents, pesticides, plasticizers, reagents, bactericides, and insect repellents.

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

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

The accuracy evaluation of the BCOP test method was limited to the substances evaluated in eight *in vitro-in vivo* comparative studies. The ability of the BCOP test method to correctly identify ocular corrosives and severe irritants, as defined by the EPA (1996), the EU (2001),

and the GHS (UN 2003) was evaluated using two approaches. In the first approach, the accuracy of BCOP was assessed separately for each *in vitro-in vivo* comparative study. In the second approach, the accuracy of BCOP was assessed after pooling data across *in vitro-in vivo* comparative studies that used similar protocols and the same method of data collection. While there were some differences in results among the three hazard classification systems evaluated (i.e, EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that BCOP test method performance was comparable among the three hazard classification systems. The overall accuracy of the BCOP test method ranged from 79% to 81%, depending on the classification system used. Sensitivity and specificity ranged from 75% to 84% and 79% to 81%, respectively. The false positive rate ranged from 19% to 21%, while the false negative rate ranged from 16% to 25%.

The accuracy analysis also indicated that alcohols are often overpredicted (50% to 56% [7/14 to 9/16] false positive rate, depending on the classification system used) in the BCOP test method. Ketones (40% [4/10]), carboxylic acids (38% to 44% [3/8 to 4/9]), and heterocyclic compounds (33% [2/6]) also had high false positive rates. Although there were a small number of underpredicted substances (4 to 5), alcohols (2) were most often underpredicted by the BCOP test method.

With regard to physical form of the substances overpredicted by the BCOP test method, 18 to 20 were liquids and two were solids. Considering the proportion of the total available database, liquids (90/120 to 92/124) appear more likely than solids (30/120 to 32/124) to be overpredicted by the BCOP test method.

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 indicated above, solids (42% to 50% false negative rate) appear more likely than liquids (4% to 5% false negative rate) to be underpredicted by the BCOP test method.

Exclusion of three discordant classes (i.e., alcohols, ketones and solids) from the data set resulted in an increased accuracy (from 81% to 92%), a decreased false positive rate (from 20% to 12%), and a decreased false negative rate (from 16% to 0%).

The 35 substances labeled as surfactants were rarely underpredicted by the BCOP test method for substances classified as severe by the EU (EU 2001) and GHS (UN 2003) classification systems (i.e., R41 or Category 1), as evidenced by the false negative rates ranging from 7% to 8%. Substances classified as severe (i.e., Category I) by the EPA classification system (EPA 1996) were more often underpredicted (false negative rate of 23%). However, although the available database was smaller (n = 7 to 9), substances labeled as pesticides were more often underpredicted by the BCOP test method (false negative rates ranging from 40% to 50%).

Considering the comparable proportion of acidic and basic underpredicted substances (18% to 30% [2/11 to 3/10] vs. 23% to 33% [3/13 to 3/9]), there was little difference among the underpredicted substances for which pH information was available. However, it is noted that

pH information was available for only a portion of the 40 to 43 severe irritant substances (i.e., Category 1, Category I, or R41) in the database for each classification system.

Finally, with respect to the GHS classification system only, the seven underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions (false negative rate of 23% [3/13]), rather than on severe lesions (false negative rate of 17% [4/24]).

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from three studies (Southee 1998; Dr. Sina's submission; Dr. Van Goethem's submission) provides an indication of the extent of intralaboratory repeatability of the BCOP test method for substances predicted as severe eye irritants. For the 16 substances evaluated in the Southee (1998) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas ranged from 11.8 to 14.2 for the three laboratories. For the 29 substances evaluated by Dr. Sina, the within experiment mean and median %CV values for *In Vitro* Irritancy Scores were 71 and 35, respectively. The dataset provided by Dr. Sina included 10 substances with low *In Vitro* Irritancy Scores around the background range of the assay (< 3.5), contributing to the increased variability of this dataset. However, the range of %CV values for the five substances predicted as severe irritants (*In Vitro* Scores > 55.1) in this study is 1.1 to 13. For the 52 substances evaluated by Dr. Van Goethem in the Gautheron et al. (1994) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas was 18.1%, comparable to the results obtained with the data from Southee (1998).

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from two studies (Gettings et al. 1996; Southee 1998) indicates the extent of intralaboratory reproducibility of the BCOP test method for substances predicted as severe eye irritants. For the Gettings et al. (1996) study, the between experiment (n = 3) mean and median %CV values for permeability values were 33.4 and 29.0, respectively, for 25 surfactant-based personal care cleaning formulations. For the Southee (1998) study, the mean %CV values for *In Vitro* Irritancy Scores for the 16 substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

A qualitative assessment of the data provided for multiple laboratories in three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) provides an indication of the extent of interlaboratory reproducibility. In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the five participating laboratories for the Balls et al. (1995) study were in 100% agreement in regard to the ocular irritancy classification for 40 to 41 (67% to 68%) of the 60 substances tested *in vitro* in the study, depending on the classification system used. 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% to 86% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). For the study by Gautheron et al. (1994), regardless of the classification system used, there was 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 substances, which were tested in either 11 or 12

laboratories. For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances, regardless of the classification system used. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as surfactants, organic solvents, chemical intermediates, detergents, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) by performing a %CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. For the Gautheron et al. (1994) study, the 17 substances predicted as severe in the BCOP assay had mean and median %CV values of 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories. For the Balls et al. (1995) study, the 32 substances predicted as severe in the BCOP assay had mean and median %CV values of 25% and 22%, respectively, for results obtained in five laboratories. For the Southee (1998) study, the mean and median %CV values for the *In Vitro* Irritancy Scores of the 16 substances were 32.4% and 22.8%, respectively, for three laboratories.

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

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

1.1 Introduction

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

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

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

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

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

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

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

Exposure of the eye of a rabbit to a test substance is the primary method for assessing the hazard potential of substances that may come in contact with or be placed near the eye of a human. The rabbit eye test method currently accepted by U.S. Federal and international regulatory agencies (CPSC 1995; EPA 1998; OECD 2002) is based on a method developed by Draize and colleagues in 1944 (Draize et al. 1944). This technique involves placing a test substance into the lower conjunctival sac of one eye of a rabbit. The contralateral eye serves as a negative control. The rabbit is then observed at selected intervals for up to 21 days after exposure for adverse effects to the conjunctiva, cornea, and iris.

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

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

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

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

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

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

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

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

This BRD evaluates existing data to determine the accuracy and reliability of the BCOP test method for identifying ocular corrosives and severe irritants. The BCOP assay is an *in vitro* eye irritation test method developed by Gautheron et al. (1992) as a modification of an earlier ocular irritation assay using isolated bovine eyes from cattle that have been slaughtered for meat or other purposes (Muir 1985). Gautheron et al. (1992) was interested in developing a reproducible, predictive *in vitro* test to evaluate the ocular irritancy of substances representing a variety of chemical and product classes. This test method developer focused on a cornea-based assay because the cornea is one of the main targets during accidental eye exposures, and damage to the cornea can result in visual impairment or loss. In addition, corneal effects are weighted heavily in the original *in vivo* ocular irritancy scoring systems (e.g., 80 out of a possible 110 points in the Draize eye test scoring system), and continue to be an ocular tissue observation on which current ocular hazard classification systems are based. Measurement of opacity in the isolated bovine cornea was initially investigated since it is the only corneal endpoint graded in many *in vivo* ocular irritancy assays. Opacity in the cornea, which is normally a transparent tissue, is a significant adverse effect of some irritants that can lead to a loss of vision. However, some known irritant substances, such as sodium lauryl sulfate and certain medium-length chained alcohols, destroy the corneal epithelium without producing significant opacity. Damage to the epithelium was subsequently quantified for these substances by measuring penetration of the dye sodium fluorescein through the isolated cornea, which is an adaptation of an *in vitro* technique previously described by Tchao (1988). Gautheron and colleagues refined the BCOP assay to measure both opacity and permeability, two important components of ocular irritation, and concluded that use of the two endpoints better predicted ocular irritancy (Gautheron et al. 1992; see **Section 9.0** for a review of these data).

In the BCOP assay, opacity is determined by the amount of light transmission through the cornea, and permeability is determined by the amount of sodium fluorescein dye that passes through all corneal cell layers. While these *in vitro* toxicity measurements using the isolated

cornea are correlated with *in vivo* ocular irritation corneal effects, they represent only one aspect of the overall complex response of the eye to irritants, which involves other tissues such as the iris and conjunctiva. More recent additions/endpoints to the BCOP assay are assessment of corneal swelling or hydration, and histological assessment of morphological alterations in the cornea (Bruner et al. 1998; Ubels et al. 1998; Cooper et al. 2001; Jones et al. 2001). When histological assessment is added to the BCOP assay, the type and depth of corneal injury can be evaluated, as well as whether the tissue damage is permanent (e.g., damage to the endothelium) (Curren et al. 2000).

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

The BCOP assay is currently used by some U.S. and European companies (e.g., pharmaceutical, cosmetic, and personal care product companies) as an in-house method to assess the ocular irritation potential of a wide range of substances or products (Gautheron et al. 1994; Sina 1994; Sina et al. 1995; Casterton et al. 1996; Chamberlain et al. 1997; Bailey et al. 2004; Cuellar et al. 2004; Swanson et al. 2004). For example, in some companies, materials that induce a high BCOP score are labeled as severe irritants (based on an internal hazard classification scheme) with no further testing. Materials that are predicted as nonirritants based on the BCOP assay are tested *in vivo* to confirm the *in vitro* results (Chamberlain et al. 1997). In another company, the BCOP assay is used to evaluate non-registered household products and registered household disinfectants, pesticides and repellents (Cuellar N and Swanson J, personal communication). For non-registered household products, the BCOP assay is used to predict the relative eye irritation potential of new consumer product formulations compared to benchmark substances, such as products on the market or substances for which the eye irritation potential is well characterized; *in vivo* confirmatory testing is generally not performed. For registered products, use of the BCOP assay is limited to product development issues and worker safety at this company.

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

1.1.2 Peer Reviews of the BCOP Test Method

Studies have been conducted in recent years to assess the validity of the BCOP test method as a complete replacement for the *in vivo* ocular irritation and corrosion test method (e.g., Balls et al. 1995). Additionally, Gautheron et al. (1994) assessed the ability of the BCOP test method to identify severe ocular irritants as classified by the European Economic Community (EEC 1984) classification system. Previous validation efforts may have failed because: 1) they attempted to support the utility of an *in vitro* alternative as a full replacement for the *in vivo* rabbit test, rather than as a component in a tiered testing strategy; and/or, 2) data generated with the *in vitro* test method(s) have typically been compared to *in vivo* maximum average scores (MAS).

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

1.2 Scientific Basis for the BCOP Test Method

1.2.1 Purpose and Mechanistic Basis of the BCOP Test Method

The BCOP is an organotypic model (i.e., isolated whole organ, or component thereof) that provides short-term maintenance of normal physiological and biochemical function of the cornea in an isolated system (Chamberlain et al. 1997). As noted above, the BCOP was developed as an alternative eye irritation test method in order to obviate the need for laboratory animals as the source for test eyes.

The most commonly used endpoints evaluated in the BCOP assay to measure the extent of damage to the cornea following exposure to a chemical substance are corneal opacity and permeability. Opacity is quantitatively measured by the amount of light transmission through the cornea, and permeability is quantitatively measured as the amount of the small molecule, sodium fluorescein, that penetrates all corneal cell layers. Irritant-induced opacity in the cornea indicates denaturation/precipitation of proteins in the epithelial or stromal layers and/or swelling, vacuolization, or damage to the cells in the stromal layer (Millichamp 1999). Development of opacity in the cornea, which is normally a transparent tissue, is a significant adverse effect of some irritants that can lead to vision loss. Increased corneal permeability results from damage to the corneal epithelium, which normally serves a barrier function. In addition, histopathological evaluation of the treated cornea provides useful descriptive information of corneal damage (Curren et al. 2000; Cooper et al. 2001).

Histopathology or confocal microscopy would allow for a more accurate assessment of the extent of corneal injury. Maurer et al. (2002) proposed that the extent of ocular injury, as

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

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

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

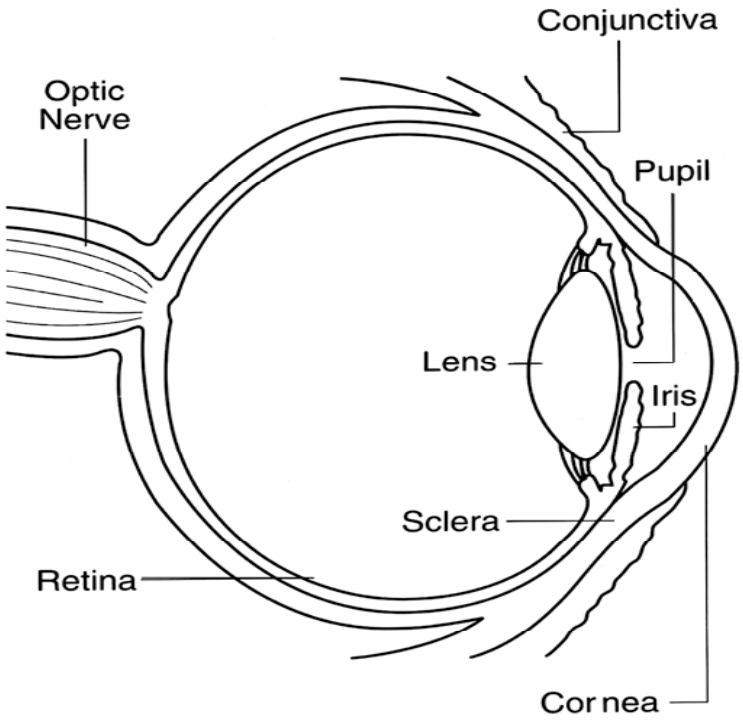
The eyeball is a fibrovascular globe, which is surrounded by a bony orbit that is impenetrable to light (Bruner 1992). The anterior portion of the eyeball is the only portion that is exposed to the environment, while the remainder of the eye is protected by the eyelids and the bony orbit. The eyeball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

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

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

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

Figure 1-1 Anatomy of the Human Eye



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

cell proliferation) can cause corneal opacity and concomitant loss of function (Parish 1985; Wilkie and Wyman 1991; Bruner 1992).

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

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

1.2.2.2 *Differences Between Human, Rabbit and Bovine Eyes*

There are several anatomical and physiological differences between the rabbit eye and the human eye. One difference is the presence of a nictitating membrane, or third eyelid, in the rabbit. As this membrane slides horizontally across the eye, it is proposed that it aids removing and/or excluding irritating substances from the corneal surface (Calabrese 1983). It also is proposed that the kinetic removal of a substance from a rabbit eye may occur at a

rate different than in humans, due to the presence of the nictitating membrane, although this has not been documented in comparative studies (Curren and Harbell 1998). Another difference is the larger conjunctival sac in the rabbit, which allows for larger test volumes to be instilled, perhaps more than could be accounted for on accidental exposure (Curren and Harbell 1998).

There are also some species differences in morphology of the cornea that could have an effect on the response of the isolated cornea to irritants. In different species, the cornea is known to vary in thickness. For example, the corneal thickness of the bovine eye is 0.8 mm, while that of the human eye is approximately 0.5 mm , and the rabbit eye is about 0.37 mm (Chan and Hayes 1985). The number of epithelial cell layers in the cornea ranges from five to seven in rabbits, compared to an average of five in humans and 10 to 14 in cattle (Cooper et al. 2001). The thicknesses of structural components of the cornea also are different between species. For example, Descemet's membrane is proposed to be about 5 to 10 μm in humans and 7 to 8 μm in rabbits (Calabrese 1983). Furthermore, the area of the cornea in relation to the total surface of the globe varies significantly between species; in humans, the relationship is 7%, while in rabbits the relationship is 25% (Swanston 1985). The Bowman's layer is well developed in humans, but it is not present to any great degree in cattle or rabbits. Finally, young rabbits have the ability to regenerate damaged corneal endothelium, while humans do not (Chambers W, personal communication). While there are known anatomical differences between human, rabbit, and bovine corneas, studies have not been found that compare the response of bovine, rabbit, and human corneas to irritants.

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

1.2.2.3 *The In Vivo Rabbit Eye Test Method*

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

the hazard classification systems used by some regulatory agencies (CPSC 1995; EPA 1998; EU 2001; UN 2003).

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

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

The tear film consists of an inner layer of mucous, a middle layer of water, and an outer film of oil. The tear film contains lactoferrin, peroxidase, lysozyme, immunoglobulins and complement factors to eliminate potentially offensive material (Unger 1992). In conjunction with the neurogenically controlled blink reflex and tear producing cells, the tear film serves as a protective barrier against an ocular irritant for the corneal epithelium. The physicochemical properties (e.g., hydrophilicity, hydrophobicity, hypertonicity, hypotonicity, oxididation, reduction) in addition to the chemical and biochemical properties of an applied test substance impact its ability to breach the tear film, or interact with its components and impact the corneal epithelium. The tear film and the aqueous humor also provide nourishment (e.g., glucose and oxygen) to the nonvascularized cornea. The extent of damage to the tear film by an applied substance therefore impacts the ability of the tear film to nourish dependent corneal tissue. Changes in the distribution, physical structure, or secretion rate of the tear film by an applied test substance might have significant nutritional, refractory, chemical and physical impacts on corneal tissue (Mishima and Hedbys 1968a, 1968b).

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

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

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

Corneal damage also may be characterized by morphological changes (e.g., described as stippling, ulceration, mottling, pannus, neovascularization). Corneal injury also is dependent on the type and concentration of applied chemical. Alkalis penetrate more readily than acids do, and the depth of penetration is dependent on alkali concentration (McCulley 1987). With alkali injury, the hydroxyl ion saponifies the fatty acid components of the cell membrane, disrupting cellular contents and resulting in cell death. The cation is responsible for the penetration process (Grant 1974). Acids tend to penetrate less deeply than alkalis, with the exception of hydrofluoric and sulfuric acids. The hydrogen ion causes damage due to pH alteration, while the anion precipitates and denatures protein in the corneal epithelium and superficial stroma (Freidenwald et al. 1946). Limbal ischemia is a significant consequence of even mild alkali or acid burns (Kuckelkorn et al. 2002).

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

Neutrophils are recruited in response to acid and alkali injury as well as in response to other ocular toxicants (Pfister 2005). Neutrophil migration is stimulated by the release of chemotatic factors (e.g., interleukins, growth factors, etc.) from damaged or chemically activated local resident epithelial cells or stromal keratocytes (Wilson et al. 2001). Loss of keratocytes following either chemical or mechanical epithelial injury may be mediated by apoptosis, perhaps by release of IL-1 and TNFα (Wilson et al. 2001). Resident mast cells may release biogenic amines that perturb the hydrostatic balance and permit inflammatory or edemagenic mediators into the locally inflamed area. Migrated neutrophils release additional cytokines (e.g., IL-1 and TNF-α) and enzymes such as proteases, collagenases, kinases, and phospholipaseA2 (PLA2). PLA2 produces edemagenic and vasoactive mediators such as prostaglandins and leukotrienes from arachidonic acid in cellular membranes.

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

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

Applied test substances also can adversely affect homeostasis within the cornea. As oxygen is absorbed into the cornea from the atmosphere, interference with oxygen uptake may lead to corneal swelling (Mishima and Hedbys 1968a, 1968b). The cellular respiratory needs of the endothelium and epithelium are similar, both requiring carbohydrate metabolism. Glucose metabolism in the cornea occurs by glycolysis and oxidation through the tricarboxylic acid cycle as well as through the hexose-monophosphate shunt (Kinoshita 1962). Glucose within the cornea is used to supply glycogen, which is stored in the epithelium. Applied substances that modulate any of these processes may be associated with ocular toxicity.

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

In the BCOP test method, damage to the isolated cornea is assessed by measuring corneal opacity and permeability in a short-term test that typically takes less than 8 hours to perform. The two endpoints are measured quantitatively with an opacitometer and an ultraviolet/visible (UV/VIS) spectrophotometer, respectively, at two or four hours after exposure to a test substance, depending on the physical properties of the substance tested.

Depending on the physicochemical properties of the test substance, post-exposure measurements may be extended to 24 hours (e.g., for substances with delayed responses). In contrast, the *in vivo* rabbit eye test involves a qualitative visual evaluation of the severity of adverse effects on the cornea, the iris, and the conjunctiva, as well as the reversibility of any ocular effects detected at selected intervals up to 21 days after exposure. In BCOP, liquids are usually applied undiluted for 10 minutes, then rinsed off the cornea, followed by a 2-hour incubation of the cornea in assay medium. Solids are usually applied as a suspension or solution (20%) for four hours, then rinsed off the cornea before opacity and permeability measurements are performed. Whether the test substance is a liquid or a solid, the entire cornea is exposed for a specified duration. In the *in vivo* rabbit eye test, liquid and solid test substances are applied to the conjunctival sac, usually in an undiluted form. Because the rabbit eye can blink and/or tear, exposure of the cornea to the test substance will be affected by these factors in terms of coverage or duration. The neurogenic components that drive tear film production are not present in the BCOP. When compared with an *in vivo* rabbit eye study, application of a test substance in the absence of this protective barrier might be expected to cause an increase in false positive outcomes. One of the conclusions from a workshop on mechanisms of eye irritation highlighted the need for additional research on the impact of chemicals on tear film and the consequences of tear film disruption (Bruner et al. 1998). Protective mechanisms for the eye (e.g., blinking, tear film) are built into *in vivo* testing, but are absent in *in vitro* testing. However, note that for some test substances (e.g., solids), blinking can also induce mechanical damage *in vivo*, contributing to a higher degree of irritation. Thus, the BCOP test method differs from the *in vivo* rabbit eye test method in the following significant ways:

- The BCOP evaluates only corneal effects and does not assess effects on the iris and the conjunctiva as performed in the *in vivo* rabbit eye test. Measurements are performed quantitatively in the BCOP assay, while they are assessed with qualitative observations in the *in vivo* rabbit eye test.
- Corneal exposure conditions, including test substance concentration and exposure duration, are well controlled in the BCOP assay, but subject to potentially greater variation *in vivo,* due in part to the blink response and natural tearing of the eye in a live animal.
- Reversibility/irreversibility of corneal effects induced by a test substance cannot be observed in the BCOP assay, *per se*, but histological evaluation of the exposed cornea may provide additional information about the depth and type of injury that could aid predictions, as to whether damage is irreversible (Harbell J, personal communication). Maurer et al. (2002) have shown that that type and depth of ocular injury are good predictors of the degree and duration of injury.
- The observation period of the BCOP assay is typically less than 24 hours, whereas ocular effects are typically evaluated in the *in vivo* rabbit eye test for a minimum of 72 hours and can extend up to 21 days.
- Protective mechanisms of the eye, such as tear production and blinking, are built into *in vivo* testing, but are absent in *in vitro* testing.
- The BCOP assay does not account for systemic effects following ocular instillation that may be noted with the *in vivo* rabbit eye test (e.g., toxicity or lethality as in the case of certain pesticides). However, these effects are

typically predicted from other acute toxicity test methods, and may not be relevant for the many consumer products that are formulated with well-characterized raw materials of known systemic toxicity.

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

Studies indicate that the BCOP test method is amenable to use with a broad range of substances with a few limitations. Substances amenable to testing include, but are not limited to, inorganic chemicals; aliphatic, aromatic, and heterocyclic chemicals; and mixtures/formulations (Gautheron et al. 1994; Balls et al. 1995; Sina et al. 1995; Gettings et al. 1996). While a wide range of substances with various physicochemical characteristics can be tested in the BCOP assay, water insoluble solid substances that are less dense than water (i.e., float on top of the solvent) do not adequately contact the cornea during treatment (Sina and Gautheron 1998). Colored test substances may be problematic as they could interfere with the opacity and/or permeability measurements.

Chamberlain et al. (1997) noted some false negative responses in the BCOP assay for substances with a delayed onset of irritation *in vivo*. However, these BCOP data were obtained using a 10-minute exposure/2-hour post-exposure protocol for liquids and a 4-hour exposure/post-exposure protocol for solids. It has been noted by some investigators that extending the post-exposure incubation time of the BCOP assay to 24 hours, and adding histopathological evaluation identifies some chemicals and formulations that produce a delayed onset of corneal damage (e.g., reactive chemicals, such as sodium percarbonate and hydrogen peroxide; Gran et al. 2003).

Additionally, some false positive responses have been noted for certain highly volatile solvents when tested using a 10-minute exposure/2-hour post-exposure protocol for liquids (Gautheron et al. 1994). More recent studies show that using a 3-minute/2-hour post-exposure protocol for volatile solvents provides a better prediction of *in vivo* results for some of these substances (Cuellar et al. 2004). Thus, as experience has been gained with the BCOP assay, practitioners have found that modifying the exposure and post-exposure times for certain substances improves the assay's predictive capability relative to results from the *in vivo* rabbit eye test.

1.3 Regulatory Rationale and Applicability

1.3.1 Current Regulatory Testing Requirements and ICCVAM Prioritization Criteria

The following section reviews and summarizes the extent to which the five ICCVAM prioritization criteria apply to the BCOP assay (ICCVAM 2003).

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

The BCOP assay has been proposed as a method to identify ocular corrosives or severe irritants, as is required by several U.S. laws. **Table 1-1** identifies the U.S. agencies and programs, which classify and label substances for eye irritation and corrosion. These agencies are the FDA, the EPA, Department of Agriculture, Department of Labor, the

Consumer Products Safety Commission (CPSC), and the Chemical Safety and Hazard Investigation Board. Therefore, the proposed use of the BCOP test method is applicable to the regulatory testing needs of multiple U.S. Federal agencies and programs.

Criteria 2. Warranted, based on the extent of expected use or application and impact on human, animal, or ecological health.
Current regulatory testing needs require the *in vivo* assessment of the eye irritancy or corrosivity hazard associated with the use of chemicals/products for labeling purposes. These testing needs require the use of laboratory rabbits. Alternative *in vitro* eye irritation and corrosion test methods could be applied to these testing needs.

Criteria 3. The potential for the proposed test method, compared to current test methods accepted by regulatory agencies, to (a) refine animal use (decreases or eliminates pain and distress), (b) reduce animal use, or (c) replace animal use.[2]
The BCOP test method has the potential to refine or reduce animal use in eye irritation testing. The BCOP test method was designed to use an animal species that is routinely used in the food industry (cattle) and that are routinely slaughtered for other purposes (e.g., food consumption). Substances that are identified as ocular corrosives or severe irritants would be excluded from testing *in vivo*, which would reduce the number of rabbits used for ocular testing and also spare animals the pain and distress of exposure to severe eye irritants.

Criteria 4. The potential for the proposed test method to provide improved prediction of adverse health or environmental effects, compared to current test methods accepted by regulatory agencies.
Based on its long history of use and acceptance by U.S. Federal and international regulatory agencies, the current system of ocular hazard assessment, which is based on the rabbit eye test (i.e., CPSC 1995; EPA 1998; OECD 2002), appears to have adequately protected public health. However, use of the rabbit eye test to predict the ocular irritation potential of substances for humans is not without controversy (e.g., intra- and inter-laboratory variability, qualitative evaluation of ocular lesions). The accuracy of the currently used *in vivo* rabbit eye test for predicting severe eye irritants in humans and the limitations of the method for predicting the irritancy of specific chemical and/or product classes are not known due to the lack of comparative data. Therefore, the potential of the proposed test method to provide improved prediction of adverse human health effects is unknown.

Criteria 5. The extent to which the test method provides other advantages (e.g., reduced cost and time to perform) compared to current methods.
Under certain circumstances, the BCOP test method could reduce the time needed to assess a substance, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days and can

---

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

be extended up to 21 days, while the standard BCOP test method can be completed in about five hours for liquid substances and seven hours for solids. However, it should be noted that the rabbit eye test may be completed within four hours for corrosive or severe irritants that produce severe lesions shortly after application to the rabbit eye, since animals should be killed for humane reasons. Additionally, the time required to perform the BCOP test method may be increased up to 24 hours when extended exposure or post-exposure times are used, or up to a week or more when histopathology is conducted. Histopathology significantly increases the time required to complete the BCOP assay, since additional time is needed for technicians to fix, process, section, and stain the corneal tissue, and for a qualified pathologist to evaluate and grade the corneal lesions.

Regarding comparative costs (based on conducting GLP compliant studies), the standard BCOP assay conducted with concurrent positive and negative controls costs $1400 per test substance at IIVS (Harbell J, personal communication). A histological evaluation, which includes photographs of tissue sections of treated corneas, as well as negative and control corneas, can be added for an additional $650-$850 per sample. A more involved GLP compliant BCOP study for one sample with benchmarks and histology costs about $4,500, which includes two time courses and one benchmark (Cuellar N and Swanson J, personal communication). The current cost of a GLP compliant EPA OPPTS Series 870 Acute Eye Irritation test (EPA 1998) or OECD Test Guideline 405 test (OECD 2002) at MB Research Laboratories (Spinnerstown, Pennsylvania) ranges from $765 for a 3 day/3 animal study up to $1665 for a 21 day/3 animal study (MB Research Laboratories, personal communication). While the cost of the BCOP assay includes concurrent positive controls, the *in vivo* rabbit test method does not include equivalent controls. One company notes that the turnaround time from initiation of the study to receipt of the final report is similar for the BCOP assay and the *in vivo* rabbit eye test (Cuellar N and Swanson J, personal communication).

1.3.2 Intended Uses of the Proposed BCOP Test Method

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

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

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

As detailed in **Section 1.2.1**, the BCOP test method evaluates only corneal effects to measure the extent of an irritant response. Corneal opacity is the only common endpoint shared between the BCOP and the *in vivo* rabbit eye test. However, this shared endpoint is evaluated differently in the two test methods. Corneal opacity is measured quantitatively with the aid of instrumentation (i.e., opacitometer or spectrophotometer) in the BCOP assay, while it is evaluated qualitatively by trained laboratory personnel in the *in vivo* rabbit eye test method. For the BCOP test method, opacity is measured on a continuous scale (e.g., 0 to 500), while for the *in vivo* rabbit eye test, opacity is graded on a discrete scale for which the only possible values are 0 for no opacity, 1 for scattered or diffuse areas of opacity, 2 for easily discernible translucent areas, 3 for nacreous areas, and 4 for complete corneal opacity.

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

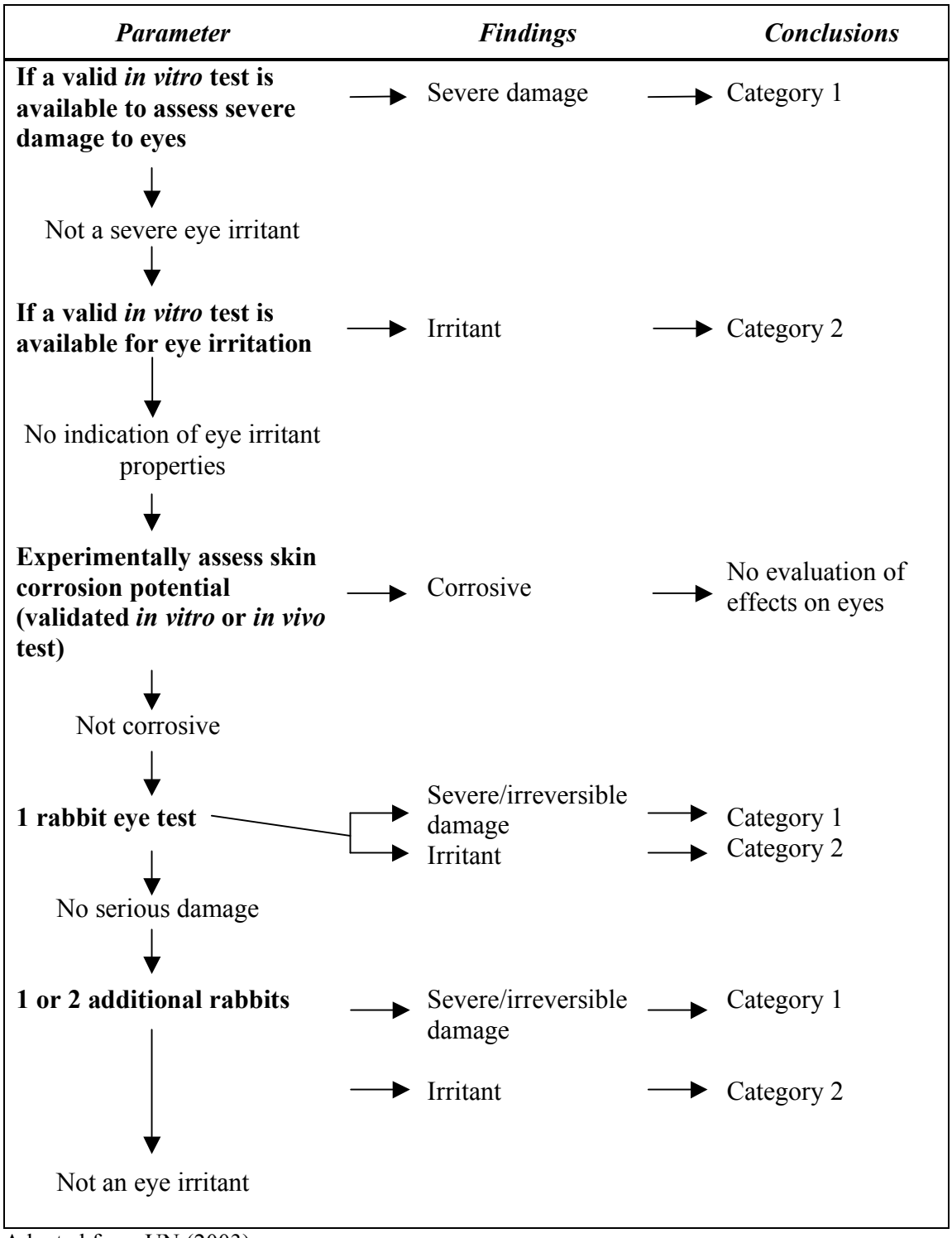
The BCOP test method is being considered for use in the identification of ocular corrosives and severe irritants in a tiered testing strategy (e.g., GHS; UN 2003). The GHS proposes a tiered testing and evaluation strategy for serious eye damage and eye irritation using available data from dermal irritation studies, knowledge of structure activity relationships, and pH screening. As shown in **Figure 1-2**, the GHS also allows for use of validated and accepted *in vitro* methods to identify severe ocular irritants/corrosives without further testing. If a test substance is classified in a validated *in vitro* method as an ocular corrosive or severe irritant, then no further testing would be required and the test substance would be appropriately labeled. If a test substance is not classified as an ocular corrosive or severe irritant using a validated *in vitro* method (i.e., the test substance remains unclassified), then current regulatory agency regulations for ocular testing would be followed. It is noted that the current testing strategy is proposed for use for regulatory classification and labeling purposes.

1.4 Validation of the *In Vitro* BCOP Test Method

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

Validation is the process by which the reliability and relevance of an assay for a specific purpose are established (ICCVAM 1997). Relevance is defined as the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 1997). For the BCOP test method described in this BRD, relevance is restricted to how well the assay identifies substances that are capable of producing corrosive or severe irritant effects to the eye. Reliability is defined as the reproducibility of a test method within and among laboratories and should be based on performance with a diverse set of substances that are representative of the types of chemical and product classes that are expected to be tested and the range of responses that needs 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.

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



Adapted from UN (2003).

The first stage in this evaluation is the preparation of a BRD that presents and evaluates the relevant data and information about the assay, including its mechanistic basis, proposed uses,

reliability, and performance characteristics (ICCVAM 1997). This BRD summarizes the available information on the various versions of the BCOP test method that have been published. Where adequate data are available, the qualitative and quantitative performances of the assays are evaluated and the reliability of each version of the test method is compared with the reliability of the other versions. If there are insufficient data to support the recommendation of a standardized protocol for BCOP, this BRD will aid in identifying essential test method components that should be considered during its development and validation.

1.5 Search Strategies and Selection of Citations for the BCOP BRD

An online search of entries in MEDLINE, TOXLINE, Web of Science, and STN International was conducted to retrieve database records on publications reporting on *in vitro* testing of substances for their ocular irritancy potential using the BCOP test method. The search was conducted in the database basic index, which includes words in the title and abstract, and indexing words. Specifically, records were sought containing the keywords "bovine" and "cornea or corneal" and "opacity" and "permeability" or "BCOP". Each database record included authors, bibliographic citation, and indexing terms. Most records also included abstracts. Of the 58 records obtained from the literature search in November 2003 (last updated in October 2004), 18 contained results and protocol information from a BCOP test method, nine were review articles, and seven were background articles related to the BCOP test method. Abstracts of selected titles were reviewed, and the relevant articles were selected and retrieved from the literature for analysis. A database of the literature citations was established using bibliographic database software. Subsequent to the initial search, additional articles with relevant information were identified and retrieved; many of these were identified from the bibliographies of the articles that were selected initially.

2.0 BCOP TEST METHOD PROTOCOL COMPONENTS

2.1 Overview of How the BCOP Test Method is Conducted

The basic procedures used to assess the effects of a test substance on an isolated bovine cornea were first reported by Gautheron et al. (1992). As described by Sina and Gautheron (1994, 1998), the BCOP assay uses isolated corneas from the eyes of freshly slaughtered cattle. Corneas free of defects are dissected with a 2 to 3 mm rim of sclera remaining to assist in subsequent handling, with care taken to avoid damage to the corneal epithelium and endothelium. Isolated corneas are mounted in specially designed corneal holders that consist of anterior and posterior compartments, which interface with the epithelial and endothelial sides of the cornea, respectively. Both chambers are filled with medium and the device is then incubated at 32 ± 1°C for one hour to allow the corneas to equilibrate with the medium and to resume normal metabolic activity. Following the equilibration period, fresh medium is added to both chambers, and a baseline opacity measurement is performed. Corneal opacity is measured quantitatively as the amount of light transmission through the cornea.

Two treatment protocols are used, one for liquids and surfactants, and one for solids. Test substances are applied to the epithelial surface of the cornea by addition to the anterior chamber of the corneal holder.

Liquids are tested undiluted; surfactants are tested at a concentration of 10% in saline or deionized water. Corneas are incubated horizontally for 10 ± 1 minutes at 32 ± 1°C. The test substance is removed from the anterior compartment and the epithelial surface is washed at least three times. After refilling both chambers with fresh medium, a second opacity measurement is taken and the corneas are incubated again at 32 ± 1°C for two hours prior to taking a final opacity measurement.

Solids are tested as solutions or suspensions at 20% concentration in saline or deionized water. Corneas are incubated horizontally for four hours at 32 ± 1°C. The test substance is removed from the compartment and the epithelial surface is washed at least three times with medium or until the corneal surface is free of visible particles. Fresh medium is added to both chambers and an opacity measurement is taken without further incubation.

Immediately after completing the final opacity measurements, corneal permeability is determined quantitatively by evaluating changes in the barrier properties of the epithelium to sodium fluorescein. To the anterior compartment of the corneal holder, 1 mL of sodium fluorescein (0.4% for liquids and surfactants, 0.5% for solids) is added. The corneas are incubated horizontally for 90 minutes at 32 ± 1°C. The amount of dye that penetrates the cornea is determined by measuring the OD of the medium in the posterior chamber with a microplate reader or UV/VIS spectrophotometer set at 490 nm.

A mean corrected opacity value (± standard deviation [SD]) and a mean corrected permeability value (OD units ± SD) are calculated for each treatment group. Most BCOP studies calculate an *In Vitro* Score for irritancy that combines both values using the following empirically derived formula (Sina et al. 1995): *In Vitro* Score = opacity value + 15 x OD_{490}

value. A substance producing an *In Vitro* Score from 0 to 25 is considered a mild irritant, from 25.1 to 55 a moderate irritant, and from 55.1 and above a severe irritant. A few laboratories do not calculate an *In Vitro* Score, but evaluate the opacity and permeability values independently. Also, some companies, such as S.C. Johnson & Son, Inc., do not use the classification system described above to assign an ocular irritancy classification, but instead compare BCOP data for newly tested substances to benchmark materials, relying on a system of comparative toxicity instead of cutoff scores (Cuellar N and Swanson J, personal communication).

These procedures were initially developed to assess the ocular irritation potential of pharmaceutical manufacturing intermediates and raw materials (Sina and Gautheron 1994; Sina 1994). However, as the BCOP test method gained more widespread use, the protocol has been modified by different investigators interested in using the assay to evaluate the ocular irritancy potential of other types of materials, including surfactant-based personal care cleaning formulations (Gettings et al. 1996), home care products (Casterton et al. 1996), alkaline liquid laundry detergents (Cater et al. 2002), oxidizing/reactive cleaning products (Swanson et al. 2003), and petrochemical products (Bailey et al. 2004). As a result of the different testing needs of different investigators, additional endpoints have been used, such as assessment of corneal hydration (Ubels et al. 1998; Cooper et al. 2001; Jones et al. 2001), and histological assessment of morphological alterations in the cornea (Curren et al. 2000; Swanson and Harbell 2000; Cater et al. 2001; Cooper et al. 2001; Jones et al. 2001; Burdick et al. 2002).

If a histological evaluation of the cornea is performed, the cornea is fixed in an appropriate fixative (e.g., 10% neutral buffered formalin) after completing the corneal permeability steps of the assay. The cornea is fixed at room temperature for at least 24 hours before processing. After embedding the corneas, they are sectioned and stained with an appropriate stain such as hematoxylin and eosin. Corneal sections are examined for lesions in the epithelium, stroma, and endothelium. Sections from treated corneas are compared to those from concurrent negative and positive control corneas (Evans 1998; Curren et al. 2000).

Other common modifications to the basic BCOP protocol include use of variable test substance exposure times and post-exposure periods that are specific to certain types of substances or products. For example, shorter exposure times are used for volatile organic solvents (Harbell J, personal communication), longer exposure times are used for diluted materials or for increased sensitivity in the mild range of irritancy (Gettings et al. 1996; Bruner et al. 1998; Cater et al. 2002, 2003), and longer post-exposure expression periods are used to test substances with a potentially delayed onset of irritancy (Rees et al. 2001; Cuellar et al. 2003, 2004; Gran et al. 2003; Swanson et al. 2003).

2.2 Description and Rationale for the Test Method Components

The publicly available BCOP test method protocols reviewed for this section follow the basic methodology originally developed for the assay as outlined by Gautheron et al. (1994) and Sina and Gautheron (1994). The essential principles of the test method protocol include isolating and culturing the bovine cornea, treating the isolated cornea with a test substance,

collecting opacity and permeability data, and evaluating the data in relation to a prediction model (Curren and Harbell 1998). However, given the various uses and applications of the BCOP test method by different investigators and laboratories, and the evolution of the assay over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method. Variations in the publicly available BCOP protocols include different instrumentation to evaluate opacity, different prediction models or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations. These test method protocol differences are described in detail in **Section 2.2.1**, where variations in specific test method components for the BCOP assay are discussed.

The test method has been evaluated in several interlaboratory studies (Gautheron et al. 1994; Sina et al. 1995; Balls et al. 1995; Southee 1998) that have led to important refinements in the test method protocol. These refinements have been incorporated into two modified BCOP protocols: 1) the protocol used during Phase II of the European Community sponsored prevalidation study of the BCOP assay conducted from 1997 to 1998 (Southee 1998); and 2) the current protocol used by a contract testing laboratory for routine evaluation of the ocular irritancy potential of test substances and materials (Institute for *In Vitro* Sciences [IIVS], Gaithersburg, Maryland). The refinements in these protocols are based partly on experience gained with the assay, and partly on experiments designed to identify specific aspects of the protocol that might contribute to intra- and inter-laboratory variability.

The following sections describe in detail the major components of the BCOP test method protocol. Similarities and differences in the test method components of available BCOP protocols are discussed. For many of these components, no rationale for inclusion in the BCOP was provided in the published literature; in such cases, historical use is considered the rationale. For each test method component, a summary is presented of information obtained from:

- IIVS, a nonprofit foundation that has performed the BCOP assay since 1997 in a GLP compliant testing facility.
- INVITTOX Protocol No.124 (1999). This protocol was used for the European Community sponsored prevalidation study of the BCOP assay conducted in 1997-1998.
- A literature search and review of publicly available BCOP protocols, which are based on the methodology first reported by Gautheron et al. (1992). These protocols are summarized in **Appendix A**.
- Discussion and personal communication with Dr. John Harbell (IIVS) and scientific experts who are members of the ICCVAM Ocular Toxicity Working Group (OTWG).

2.2.1 Materials, Equipment, and Supplies Needed

2.2.1.1 *Bovine Eyes: Source, Collection/Handling and Quality*

Source: Several BCOP studies noted that bovine eyes were obtained from a local slaughterhouse that was close enough to the testing laboratory to allow for transport of the eyes to the laboratory within two to four hours after the animals were killed (Gautheron et al. 1994; Rachui et al. 1994; Sina et al. 1995; Casterton et al. 1996; INVITTOX 1996; INVITTOX 1999). Other BCOP studies noted that the bovine eyes were likewise obtained

from a local slaughterhouse, but reported different periods of time until use of the eyes. For example, Bruner et al. (1998) reported that eyes were used within 12 hours after receipt at the laboratory, and Cerven and Moreno (1998) reported that the eyes were examined within one hour after receipt at the laboratory without noting the amount of time that had passed postmortem. At IIVS, bovine eyes generally arrive in the testing laboratory within four to five hours of the first eyes being enucleated at the slaughterhouse, and eyes are processed immediately upon arrival at the laboratory (Harbell J, personal communication). Therefore, while a formal study to determine the maximal time not to be exceeded during the transport of eyes to the testing facility was not found in the published scientific literature, a maximum of five hours has been used in most BCOP protocols and appears to produce consistent results.

No detail was provided in the study reports on the specific breed, age, or sex of the cattle used as the source of the bovine eyes. Based on information from other sources, it was found that the cattle sent to slaughterhouses are typically killed either for human consumption (e.g., calves for veal; steers 9 to 30 months old for prime, choice, select, or standard grades of beef), or for other commercial uses (e.g., cattle 42 to 96 months for commercial, utility, or cutter grades of beef). The cattle in the former category tend to be raised specifically for meat production and thus are of cattle breeds (e.g., Hereford) used to optimize the quality and quantity of beef for human consumption. The cattle in the latter category can include dairy cattle (e.g., Holstein) that are no longer useful for milk production (Doughty et al. 1995; North Dakota State University Extension Service 1999).

Although bovine eyes are widely used in ocular irritancy evaluations, only a few studies were found that addressed potential sources of variability in bovine eyes obtained from slaughterhouse operations (Doughty et al. 1995; Doughty 1997, 2004). In one study, central corneal thickness (CCT) values ranged from 750 to 1450 µM (mean and SD of 1015 ± 104 µM) and horizontal corneal dimensions ranged from 27.5 to 34.5 mm (mean and SD of 29.8 ± 1.3 mm) in bovine eyes obtained from 315 Holstein and Hereford cattle killed at a local slaughterhouse over a one-year period (Doughty et al. 1995). These variations in corneal dimensions were proposed to be a result of obtaining the eyes from animals of different ages. Corneas with a horizontal dimension greater than 30.5 mm and CCT values equal to 1100 µM or greater were likely obtained from cattle older than eight years, while those with a horizontal diameter less than 28.5 mm and CCT less than 900 µM were likely from cattle less than five years old (Doughty et al. 1995). For this reason, eyes from mature cattle (i.e., greater than 60 months old) are not typically recommended. Additionally, eyes from cattle less than 12 months of age are believed to be inadequate since the eyes are still developing and the corneal thickness and corneal diameter are considerably smaller than that reported for eyes from adult cattle. However, as discussed below, a recent study suggests that eyes from younger animals may indeed be useful.

It should be noted that these findings may be applicable only to the specific cattle breeds and slaughterhouse operation used in the study. However, they are suggestive of potential variability in corneas sizes and thicknesses of bovine eyes obtained from slaughterhouse operations. Limited information could be found on whether variable cornea sizes from animals of different ages might impact the performance of the BCOP test method. During

the European Community prevalidation study of BCOP, a small study was conducted to evaluate whether cornea size influenced BCOP test method results obtained for ethanol (Southee 1998). The investigators reported that the results suggested no apparent relationship between cornea size, basal opacity, or cornea response to ethanol. In addition, data provided by Johnson & Johnson Pharmaceutical Research and Development for 19 test substances suggests that the performance of the BCOP when using eyes from young (6-8 months) versus adult (> 24 months) animals is comparable (see **Section 9.2.4**). However, because there are limited data on this matter, further evaluation of potential variability among corneas from slaughterhouse animals may be necessary to investigate whether the size or age of the cornea influences the responsiveness of the cornea to irritating substances.

Collection/Handling: Most BCOP studies noted that the bovine eyes were excised by a slaughterhouse employee with care taken to avoid damage to the cornea; however, details on the enucleation procedure and the specific steps taken to avoid corneal damage were not provided in any of the study reports. Depending on the slaughterhouse operation, it may take several hours for a slaughterhouse employee to collect the required number of eyes for use in a BCOP study at the testing facility.

IIVS notes that they use bovine eyes that are collected by slaughterhouse employees at various times following exsanguination and decapitation of the cattle. To minimize mechanical and other types of damage to the eyes, this laboratory prefers the eyes be enucleated as soon as possible postmortem and requests that slaughterhouse employees not use detergent when rinsing the animal head to prevent exposure of the bovine eyes to potentially irritating substances (Harbell J, personal communication). To the extent possible, IIVS communicates their need for undamaged bovine eyes to the slaughterhouse, while recognizing the constraints of the slaughterhouse environment.

Because the bovine eyes are collected during the process of slaughter, it is recognized that the bovine eyes may have been exposed to blood and other biological substances, including bacteria and other microorganisms (Doughty 1997).

The BCOP studies varied in how the bovine eyes were handled after enucleation at the slaughterhouse and during transit prior to arrival at the testing facility. The two major variables in handling were differences in the solution used to store the eyes, and differences in the temperature of the eye storage container. Most studies noted that the eyes were immersed completely in Hanks' Balanced Salt Solution (HBSS) in a suitably sized container. Of the 18 studies reviewed, four reported addition of the antibiotics penicillin and streptomycin to the HBSS (Bruner et al. 1998; INVITTOX 1999; Cooper et al. 2001; Jones et al. 2001), while the other studies appear not to have used antibiotics. With regard to the temperature of the collection vessel, some studies maintained the storage container at ambient temperature (Gautheron et al. 1994; Rachui et al. 1994; Casterton et al. 1996; INVITTOX 1996; INVITTOX 1999), while others maintained it on ice to keep the eyes cool and to minimize ambient temperature variation that would result due to seasonal changes (Cooper et al. 2001; Jones et al. 2001). The matter of temperature maintenance of the eye collection vessel was not addressed in the other reviewed studies.

Quality of Eyes: Currently, it appears that there are no standardized criteria for the selection of bovine eyes for the BCOP assay. Most BCOP studies reported that the eyes were carefully examined visually for defects, including opacity, scratches, and neovascularization, once they had arrived at the laboratory. A few studies also noted use of microscopes to assist in identifying damaged corneas. Rachui et al. (1994) commented that the eyes were carefully examined visually, or with the aid of a stereomicroscope. Swanson et al. (1995) stated that the corneas were examined microscopically after they were dissected, and only corneas free of defects were used in the BCOP assay.

The quality of the corneas is evaluated at later steps in the assay, as well. For example, corneas that have a high baseline opacity reading (e.g., opacity greater than 10) after the initial one-hour equilibration period are discarded, a practice that is consistent among the reviewed BCOP protocols. Opacity that develops in the cornea prior to application of a test substance sometimes results from fine scratches not noticeable upon visual inspection (Harbell J, personal communication).

2.2.1.2 *Instrument to Measure Light Transmission Through the Cornea*

Changes in light passage through the cornea have been most commonly assessed with a white light, dual-beam opacitometer (e.g., Spectro Designs OP-KIT, STAG BIO, Electro-Design). This type of opacitometer provides a center-weighted reading of light transmission through the cornea. There are 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. The BCOP assay was developed with the center-weighted opacitometer, and a majority of BCOP studies in the peer-reviewed literature report using this type of opacitometer. However, the center-weighted readings may underestimate opacity that develops as spots on the periphery of the isolated cornea (Southee 1998; van Goethem et al. 2002), and therefore some BCOP users have modified the method of reading opacity. Casterton et al. (1996) first reported the use of a UV/VIS spectrophotometer to evaluate corneal opacity. Corneal holders were modified to fit into the spectrophotometer and light absorbance (570 nm) readings performed through the center of the cornea. Absorbance values use a different scale than values obtained from the white light opacitometer; thus, BCOP data from the two instruments cannot be directly compared. This method of measuring opacity requires the use of a different classification procedure or prediction model to identify ocular irritants when compared to the traditional BCOP assay.

Recognizing the limitations of the conventional opacitometer with its center-weighted readings, Janssen Pharmaceutica/Johnson & Johnson Pharmaceutical Research recently developed a new laser-based opacitometer that uses an adjustable laser beam in combination with a calibrated photocell (van Goethem et al. 2002). This opacitometer was designed to provide a more even distribution of light across the corneal surface and, thus, may provide an improved method of opacity assessment. However, the database of BCOP studies using this type of opacitometer is still relatively small, and thus additional studies are required to determine if such instruments provide a definitive advantage over the conventional opacitometer (i.e., center-weighted readings)

2.2.1.3 *Instrument to Evaluate Permeability*

Over half of the BCOP studies used a UV/VIS spectrophotometer set at 490 nm to measure the amount of sodium fluorescein (based on optical density) that permeated through the cornea into the posterior chamber of the corneal holder. The remaining studies used a microtiter plate (microplate) reader (e.g., Dynatech MR 5000 and Molecular Devices V_{max} kinetic microplate readers) to measure the amount of sodium fluorescein. The basic design of the two instruments is the same in that a selected wavelength of light passes through the samples and a photosensitive tube detects the amount of light transmitted through the sample. For this reason, either instrument would appear adequate. However, a standard spectrophotometer measures one sample at a time, while a microplate reader is capable of measuring the absorbance of 96 samples in about eight seconds. Thus, the microplate reader offers the advantage of processing large numbers of samples in a short amount of time.

2.2.1.4 *Organ Culture Media*

A few variations in organ culture media were found in the publicly available BCOP study reports. All protocols used some form of complete Minimum Essential Medium (complete MEM), supplemented with 1% fetal bovine serum (FBS). One of the major differences, however, is that the earlier protocols used complete MEM containing phenol red (now considered an outdated practice), while the more recent protocols used complete MEM without phenol red. As part of the European Community prevalidation study of the BCOP assay, investigators evaluated the effect of phenol red in the BCOP incubation medium (Southee 1998). Results from a series of separate assays indicated that complete MEM without phenol red produced lower background opacity readings than phenol red containing MEM. The study report also noted that fluctuation in background values was less for medium without phenol red, attributed in part to the low background values. However, phenol red is useful in the medium during the rinsing procedure, when the test substance must be removed completely from the cornea; residual test substance can sometimes be identified by a shift in color of the phenol red (Harbell J, personal communication).

A second notable variation is that some protocols prewarmed the complete MEM to 32°C, the temperature at which the corneal equilibration step and all incubations are performed. Prewarming the organ culture medium eliminates the time needed for the media temperature to equilibrate with the incubator system or the water bath. A few protocols also reported adjusting the pH of the complete MEM from 7.2 to 7.4 prior to use in the assay, although most did not. Adjustment of pH to a physiological level was likely performed in situations when sodium bicarbonate was added to the MEM by the testing facility to provide buffering capacity to the media. However, MEM with appropriate buffering capacity can be purchased, obviating the need for pH adjustment. Other slight differences appear to be related to the level of detail provided in the study reports. For example, some protocols reported use of the standard complete MEM supplements, such as L-glutamine, Ca^{++}, Mg^{++}, and sodium bicarbonate, while others did not, making it unclear whether the same supplements were used in different BCOP studies.

2.2.1.5 *Solvents*

Differences in the use of solvents have been noted. Some reports noted that solid compounds were prepared as a 20% solution or suspension in 0.9% NaCl (Vanparys et al. 1993; INVITTOX 1996; INVITTOX 1999). In comparison, some solid and surfactant test substances were prepared in MEM (Gautheron et al. 1994; Rachui et al. 1994; Sina et al. 1995; Chamberlain et al. 1997; Cerven and Moreno 1998). IIVS uses sterile, deionized water or saline to dissolve or suspend solid test substances (Harbell J, personal communication). The European Community prevalidation study report noted that use of saline is preferred for dilutions, since it may prevent possible buffering effects and enhanced penetration of the test substance that could result from the use of organic solvents (Southee 1998).

2.2.1.6 *Incubation Apparatus*

A majority of BCOP studies reported using a water bath for incubations (Rachui et al. 1994; INVITTOX 1996; Bruner et al. 1998; Cerven and Moreno 1998; INVITTOX 1999). A few studies reported carrying out incubations at room temperature (Sina et al. 1995; Casterton et al. 1996), while still others reported using a forced air incubator (Cassidy and Stanton 1997; Cooper et al. 2001); IIVS also currently uses a forced air incubator in its studies.

An experiment was conducted during the European Community prevalidation study of the BCOP assay to evaluate whether similar results are obtained for the same test substance, when the assay is conducted using a water bath or a forced air incubator. This experiment evaluated one test substance identified as "CTAB", which produces a severe response in the isolated cornea. Half of the exposed corneas were incubated for 30 minutes in a water bath, while half were incubated for 30 minutes in a forced air incubator; all other procedures were the same. The study authors concluded that there was a "distinct" difference in opacity and permeability values, and consequently, the mean *in vitro* score obtained for CTAB, depending on the incubation system used. The authors, however, did not state that the results were statistically significant. The study report notes "the water bath provides a more stable temperature than the air incubator which fluctuates when the door is opened. Water also provides greater heat conductivity, and hence the holders will reach 32º C quicker" (Southee 1998).

Others have noted that the water bath allows for better heat transfer, but is technically more difficult to use. Sometimes there are cross-contamination problems, when water from the water bath seeps into the corneal holder or when the test substance seeps into water bath (Harbell J, personal communication).

Both types of incubators have advantages and disadvantages. The water bath offers greater temperature control but greater opportunity for cross contamination. Until more information becomes available about the comparative advantages and disadvantages of the forced air incubator and the water bath, it would appear that both would be adequate for performing incubations.

2.2.1.7 *Corneal Holder*

As described by Gautheron et al. (1992) and Sina and Gautheron (1998), the corneal holder for the BCOP assay consists of two chambers, each with a 5 mL volume. The main part of

the chamber is composed of either polypropylene (Sina and Gautheron 1998) or clear Plexiglas (Casterton 1998). The chamber design consists of a glass window on the outside of the chamber, and a 17 mm circular opening on the inner side on which the cornea rests (Gautheron et al. 1992; Ubels et al. 2002). The anterior chamber interfaces with the epithelial side of the cornea, while the posterior chamber interfaces with the endothelium. After the cornea is mounted over an O-ring that is positioned around the opening of the posterior chamber, the chambers are clamped together with three screws (Gautheron et al. 1992). Dosing holes located on the top of each chamber allow the epithelial and endothelial sides of the cornea to be treated independently.

The distributors of the opacitometer (e.g., Spectro Designs OP-KIT, STAG BIO, Electro-Design) also supply the corneal holders. It appears that the laboratories that have used a UV/VIS spectrophotometer to measure opacity had the corneal holders specially made and designed for use with that instrument (Casterton et al. 1996; Casterton 1998; Ubels et al. 1998).

More recently, studies by Ubels et al. (2000, 2002) have suggested potential limitations regarding the conventional corneal holder: 1) it has a circular opening 17 mm in diameter, yet the bovine cornea is oval shaped and has dimensions of about 24 mm vertically and 30 mm horizontally; 2) it has flat inner surfaces, whereas the bovine cornea is convex or curved. These elements of the corneal holder reportedly force the bovine cornea into an unnatural shape when mounted in the holder, causing the cornea to wrinkle. Ubels et al. (2002) also noted damage to all three corneal cell layers (epithelium, stroma, and endothelium) where the cornea comes in contact with the circular edge of the holder opening.

Recognizing some of the potential limitations of the conventional corneal holder, Ubels et al. (2002) designed a new corneal holder with dimensions that better fit the bovine cornea and maintain its natural shape during the BCOP assay. The new holder was designed to contact the 2 to 3 mm rim of sclera left around the bovine cornea during dissection, rather than the corneal tissue. Studies showed that this refined corneal holder does not cause wrinkling of the mounted bovine cornea, nor does it damage the cell layers around the edge of the cornea (Ubels et al. 2002). However, the availability of this new corneal holder for purchase or use by other laboratories is not known. It would seem appropriate that consideration be given to the newly designed corneal holder as a potential refinement of the assay, once it does become commercially available, since it appears that this holder better fits the natural shape and curvature of the bovine cornea.

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

As described below in **Section 2.2.4.4**, test substances are typically applied as neat chemicals (liquids), or diluted to prescribed concentrations (surfactants and solids) with preferred solvents. A few studies also described testing of personal care products, such as shampoos, at proposed end-user concentrations to mimic potential human exposure scenarios (Cooper et al. 2001; Jones et al. 2001).

2.2.3 Endpoints Measured

In the BCOP assay, opacity is determined by the amount of light transmission through the cornea, and permeability is determined by the amount of sodium fluorescein dye that penetrates all corneal cell layers (i.e., the epithelium on the outer cornea surface through the endothelium on the inner cornea surface). In a majority of the BCOP studies reviewed, corneal opacity was measured quantitatively with the aid of a center-weighted opacitometer, resulting in opacity values measured on a continuous scale. The concentration of sodium fluorescein in the posterior corneal chamber, which interfaces with the endothelial side of the cornea, was quantitatively measured with the aid of UV/VIS spectrophotometry. Spectrophotometric measurements evaluated at 490 nm are recorded as optical density or absorbance values, which are measured on a continuous scale.

The measurement of opacity is described in detail in **Section 2.2.1.2**. As previously noted, a few BCOP studies reported using a UV/VIS spectrophotometer instead of an opacitometer to evaluate corneal opacity (Casterton et al. 1996; Ubels et al. 2003).

The measurement of permeability is standard across the reviewed BCOP studies. Typically, 1 mL of 4 mg/mL sodium fluorescein solution in MEM is used when testing liquid and surfactant substances, and a 5 mg/mL solution is used when testing solid substances. No rationale could be found for the use of different concentrations of sodium fluorescein for different types of substances. The sodium fluorescein solution is added to the anterior chamber, and the holder incubated horizontally for 90 minutes (Gautheron et al. 1992, 1994).

The stock solutions of sodium fluorescein used for the BCOP assay are prepared to the specified concentrations, and then verified using a UV/VIS spectrophotometer to ensure the absorbances of the solutions fall within set limits. The UV/VIS spectrophotometer used for permeability measurements is calibrated with dilutions of sodium fluorescein solution to determine the linear portion of the absorbance curve and to define the limits outside of which the test substances require dilution (Southee 1998).

More recent additions/endpoints to this assay include histological assessment of alterations in the cornea, and, less commonly, assessment of corneal hydration (Bruner et al. 1998; Ubels et al. 1998; Cooper et al. 2001; Jones et al. 2001).

Based on the results of a major validation study of BCOP (Balls et al. 1995), it was found that certain severe ocular irritants are underpredicted using only the opacity and permeability endpoints. These findings prompted Curren et al. (2000) to investigate the usefulness of examining histological changes in the cornea in conjunction with the primary BCOP endpoints of opacity and permeability. Curren and colleagues found that three materials underpredicted using only the opacity and permeability endpoints -- parafluoroaniline, quinacrine, and sodium oxalate -- produced notable cellular damage throughout the epithelium and in other tissues that was indicative of severe ocular injury. For example, parafluoroaniline produced death of keratocytes, quinacrine produced microvacuolization throughout the epithelium as well as in keratocytes and the endothelium, and sodium oxalate produced refractile, crystal-like material throughout the epithelium into the basement membrane. Thus, assessment of histopathology in the BCOP assay may be considered

essential for ocular irritants where the mode of action does not result in significant opacity or permeability.

It is widely recognized that histological evidence of corneal damage (or lack thereof) provides additional information for an assessment of ocular irritation. However, the additional expense and time required for such a detailed examination may not be warranted in all cases, such as when severe corneal effects are clearly indicated from the opacity and permeability assessments of the BCOP assay. Instead, histopathological effects could be useful for discriminating borderline moderate/severe cases, identifying alternate mechanisms of severe ocular damage that do not produce significant opacity or permeability, or for evaluating new chemistries where the mode of action is not readily predictable. Also, certain chemical classes, such as oxidizing agents that have a delayed onset of irritation *in vivo*, may require a histological assessment to fully evaluate the extent of injury. Therefore, the decision to perform a histological assessment of the treated cornea should likely be left to the discretion of the investigator. However, it would seem prudent for the corneas from all studies to be fixed in an appropriate fixative (e.g., 10% neutral buffered formalin), so that the tissues are available if histology is necessary or requested at a later time.

At IIVS, the scoring of lesions in a histological evaluation of the isolated cornea is based primarily on the depth of injury, which is predictive of the degree and duration of the injury (Maurer et al. 2002). The three main tissue layers of the cornea (epithelium, stroma, endothelium) are evaluated, and the nature, degree and depth of lesion in each tissue layer are noted. Tissues from the treated corneas are always compared with tissues from the concurrent negative control cornea to distinguish between test substance induced injury and artifacts of handling or processing (Harbell J, personal communication).

2.2.4 Duration of Exposure

2.2.4.1 *Pre-Exposure Preparations*

Pre-exposure preparations are consistent across BCOP protocols. Corneas free of defects are dissected with a 2 to 3 mm rim of sclera remaining to assist in subsequent handling, with care taken to avoid damage to the corneal epithelium and endothelium. Isolated corneas are mounted in specially designed corneal holders that consist of anterior and posterior compartments, which interface with the epithelial and endothelial sides of the cornea, respectively. Both chambers are filled with medium and the device is then equilibrated at 32°C for one hour to allow the corneas to equilibrate with the medium (the approximate temperature of the corneal surface *in vivo* is 32°C). This is intended to allow the corneas to resume normal metabolic activity. Following the equilibration period, fresh medium is added to both chambers and baseline opacity readings are taken for each cornea. Any corneas that show tissue damage or high opacity (e.g., > 10 opacity units) are discarded. The mean opacity of all equilibrated corneas is calculated. A minimum of three corneas with opacity values close to the average value for all corneas are selected as negative (or solvent) control corneas. The remaining corneas are then distributed into treatment groups and positive/other control groups.

2.2.4.2 *Effects of Residual Equilibration Medium in the Test Substance Chamber*
As part of the European Community prevalidation study, the investigators evaluated whether residual medium left in the anterior chamber after the pre-exposure incubation had an effect on the opacity and permeability of the cornea to ethanol. Increasing volumes of complete MEM (ranging from 0 to 150 µL) were added to the anterior chamber with 0.75 mL of ethanol to simulate residual medium in the anterior chamber. After a 10-minute incubation at 32°C, opacity and permeability measurements were performed. The results showed that increasing amounts of residual medium produced a corresponding increase in the final *in vitro* score of ethanol. The *in vitro* score for ethanol with no residual media was 28.7, while the *in vitro* score for ethanol with 150 µL of media was 48.8 (Southee 1998).

Based on these results, the prevalidation study report recommended that an aspiration method be used to remove as much medium as possible from the anterior chamber prior to addition of the test substance. The study report noted that one suitable method for removing all traces of incubation medium is to use a micropipette tip or blunt needle attached to a vacuum pump.

2.2.4.3 *Test Substance Exposure Volume*
A majority of BCOP protocols consistently applied 0.75 mL of test substance to the cornea (Gautheron et al. 1994; Rachui et al. 1994; Balls et al. 1995; Swanson et al. 1995; INVITTOX 1996; Cassidy and Stanton 1997; Bruner et al. 1998; Cerven and Moreno 1998; INVITTOX 1999; Cooper et al. 2001; Jones et al. 2001). Liquids are typically tested neat, while surfactants and solids are solubilized or suspended at prescribed concentrations.

A few protocols reported using 0.5 mL of test substance solution or suspension (Sina et al. 1995; Chamberlain et al. 1997). However, this volume is no longer used because in some cases it failed to cover the corneal surface completely (Harbell J, personal communication). In addition, one report noted a test substance volume of 1.0 mL (Casterton et al. 1996). However, this exception was likely used due to the fact that a unique corneal holder was used in this protocol, one customized for making opacity measurements with a UV/VIS spectrophotometer rather than an opacitometer, which required a larger volume than traditionally used (i.e., 0.75 mL).

2.2.4.4 *Concentration Tested*
For the European Commission (EC) sponsored interlaboratory assessment of the BCOP assay, Gautheron et al. (1994) tested liquids neat (100%), surfactants at a concentration of 10%, and nonsurfactant solids at a concentration of 20% (w/v). The EC/British Home Office (HO) validation study of alternatives to the Draize eye test used the same concentrations in its evaluation of the BCOP assay (Balls et al. 1995), as did the European Community prevalidation study of the BCOP assay (Southee 1998). A majority of the other publicly available protocols used the same test substance concentrations, with a few exceptions. To address specific product development questions, Cooper et al. (2001) and Jones et al. (2001) tested surfactant-based hair-care formulations (shampoos and conditioners) at concentrations of 10% and 100%. Also, Gran et al. (2003) found that a test substance concentration of 50% (in addition to longer exposure/post-exposure times) produced a better correlation to *in vivo* results for certain reactive/oxidative solids, such as sodium percarbonate. Instead of testing

solids at a 20% concentration, Casterton et al. (1996) applied solid test substances undiluted (neat) to the cornea.

Therefore, historical use generally supports testing liquid substances neat, surfactants at 10%, and nonsurfactant solids at 20%. However, it is recognized that these concentrations may require adjustment for certain chemical or product classes.

2.2.4.5 *Application of Test Substance to Bovine Cornea*

A majority of the BCOP studies used two treatment protocols, one for liquids and surfactants, and one for nonsurfactant solids (Gautheron et al. 1992, 1994; Rachui et al. 1994; Balls et al. 1995; Sina et al. 1995; Chamberlain et al. 1997; Cerven and Moreno 1998; INVITTOX 1999). For both treatment protocols, the test substances were applied to the epithelial surface of the cornea using a micropipettor. The test substances were injected into the anterior chamber of the corneal holder through dosing holes on the top of the chamber (closed chamber method).

IIVS uses the closed chamber method for nonviscous to slightly viscous liquids and solubilized solids. However, they have developed a refined procedure for application of semiviscous to viscous test substances, known as the "open chamber method." In this method, the window-locking ring and glass window are removed from all appropriate anterior chambers and the holders are placed into a horizontal position (anterior chamber facing up). Approximately 0.75 mL of the viscous test substance (or enough test substance to completely cover the cornea) is applied directly to the epithelial surface of the cornea using a micropipettor or other appropriate device, such as a spatula. The corneal holder is reassembled prior to incubation of the test substance (Harbell J, personal communication).

Casterton et al. (1996) reported a different procedure for application of solid substances. Solid substances were applied directly onto the cornea by removing the glass window of the corneal holder. Although, a specific weight or volume of solid was not reported, the authors stated that enough test substance was added to cover the cornea thoroughly.

2.2.4.6 *Test Substance Exposure Duration*

Most BCOP protocols incubated liquids and surfactants for 10 minutes at 32 ± 1°C. The test substance was removed from the compartment and the epithelial surface washed at least three times. After replacing the medium, an opacity measurement was taken. The corneas were then returned to the incubator for an additional two hours and another opacity reading taken, which was used for the calculation of corneal opacity. Solutions or suspensions of solids were incubated horizontally for four hours at 32 ± 1°C. The test substance was removed from the compartment and the epithelial surface washed at least three times with medium or until the corneal surface was free of visible particles. Fresh medium was added to both chambers and an opacity measurement was taken without further incubation (Gautheron et al. 1992, 1994; Rachui et al. 1994; Balls et al. 1995; Sina et al. 1995; Chamberlain et al. 1997; Cerven and Moreno 1998; INVITTOX 1999).

Shorter exposure times have been suggested for alcohols and volatile organic solvents, since the irritancy of these substances has been overpredicted with an exposure time of 10 minutes

(Harbell J, personal communication). Some protocol refinements may have to be made if the irritancy of alcohols and volatile organic solvents are consistently overestimated. Longer exposure times (e.g., 60 minutes and 24 hours) have been suggested for better discrimination of mild to moderate ocular irritants, and to differentiate subtle differences between similar formulations (Bruner et al. 1998; Cater et al. 2002, 2003; Harbell J, personal communication).

IIVS reported that they use different exposure times to address certain chemicals/chemical classes (e.g., sodium percarbonate, volatile solvents), expected consumer exposure models (e.g., diluted shampoo), or to enhance comparisons across a chemical class (Gran et al. 2003; Harbell J, personal communication).

For solid test substances, Casterton et al. (1996) used a shorter exposure time of one hour after applying the test substances undiluted (neat) to the cornea. Exposure was followed by a 1-hour post-rinse incubation period. This reduced exposure time has not been widely evaluated. Historical use generally supports an exposure time of 10 minutes for liquids and surfactants, and four hours for nonsurfactant solids. However, it is recognized that these generic exposure times may require adjustment for certain chemical classes, such as alcohols and volatile solvents.

2.2.4.7 *Post-Exposure Incubation*

A majority of BCOP studies in the literature reported incubating the corneas that had been treated with liquids or surfactants for an additional two hours at 32 ± 1ºC after the 10-minute test substance exposure and the post-treatment rinse. Corneas treated with solid test substance were exposed to the test substance for four hours, and were not further incubated. However, Casterton et al. (1996) used a 1-hour post-exposure incubation when testing solids.

Bruner et al. (1998) used longer post-exposure times to better discriminate the irritancy of formulations of a similar composition. IIVS sometimes uses longer post-exposure incubation times for better discrimination of mild to moderate ocular irritants and for substances with a delayed response (Harbell J, personal communication). IIVS also uses different post-exposure incubation times to address certain chemical (e.g., peroxides) and product classes and expected consumer exposure models (Gran et al. 2003).

Historical use generally supports a post-exposure time of two hours for liquids and surfactants. Corneas treated with solids typically do not require further incubation beyond the 4-hour exposure period. However, it is recognized that these generic post-exposure times may require adjustment for certain chemical or product classes.

2.2.5 Known Limits of Use

While a wide range of substances with various physicochemical characteristics can be tested in the BCOP assay, water insoluble solid substances that are less dense than water (i.e., float on top of the solvent) do not adequately contact the cornea during treatment (Sina and Gautheron 1998). Thus, the standard BCOP protocol for solid test substances (Gautheron et al. 1994) cannot be used for low density, water insoluble substances. In addition, Chamberlain et al. (1997) noted some false negative responses for substances tested with the

standard BCOP protocol (Gautheron et al. 1994) that had a delayed onset of irritation *in vivo*. However, test method users are addressing these limitations. For example, the method of applying solid test substances used by Casterton et al (1996), in which solids are sprinkled neat onto the cornea, may be useful to address the limitation of testing low density, insoluble solid substances. Protocols with longer exposure and post-exposure periods are under development to detect substances with a delayed onset of irritancy (Gran et al. 2003). However, the longest exposure/post-exposure period found is 24 hours (Bruner et al. 1998; Gran et al. 2003).

Another potential limitation of the test method is that, although it takes into account some of the ocular effects evaluated in *in vivo* rabbit ocular irritancy tests and to some degree their severity, it does not consider all of the effects assessed *in vivo*. Reversibility of corneal lesions cannot be evaluated *per se* in the BCOP assay, but test method users propose that an assessment of the initial depth of corneal injury can be used to predict irreversible or reversible effects (Maurer et al. 2002). Furthermore, in Europe and Japan, there are concerns about the use of bovine tissue due to the risk of transmitting Bovine Spongiform Encephalopathy (BSE).

2.2.6 Nature of the Response Assessed

2.2.6.1 *Corneal Opacity*

Corneal opacity is measured quantitatively with an opacitometer (e.g., ElectroDesign, Riom, France), which measures differences in light transmission between treated corneas and an air blank. Numerical opacity values with arbitrary units are obtained, with values typically ranging from 0 to 500, with higher opacity values occasionally reported.

2.2.6.2 *Permeability*

The amount of dye that permeates the cornea is determined by measuring the OD/absorbance of the medium in the posterior chamber with a spectrophotometer set at 490 nm.

2.2.6.3 *Histology*

Although a more recent addition to the BCOP assay, a histological evaluation of the type, degree and depth of injury at the tissue level, resulting from exposure of the cornea to a test substance appears to be a very useful addition to the assay (Curren et al. 2000; Cooper et al. 2001).

2.2.7 Appropriate Controls and the Basis for Their Selection

2.2.7.1 *Negative Controls*

Some differences were found in the negative controls used in the BCOP assay. Seven BCOP studies used complete MEM as the negative control (Gautheron et al. 1994; Rachui et al. 1994; Rougier et al. 1994; Sina et al. 1995; Bruner et al. 1998; Cooper et al. 2001; Jones et al. 2001). Two studies used 0.9% saline (INVITTOX 1996; INVITTOX 1999). IIVS uses sterile, deionized water (Harbell J, personal communication). To test the possible differences in the use of complete MEM or saline as the negative control, the European Community prevalidation study compared the BCOP results obtained for saline and complete MEM (without phenol red). When incubated for 10 minutes, there was no apparent difference in the results in the opacity and permeability values of complete MEM and saline (Southee

1998). It appears that the three commonly used negative controls for the BCOP assay offer no distinct advantages or disadvantages.

However, it is clear that a negative control is useful in the BCOP test, so that nonspecific changes in the test system can be detected. This type of control also provides a baseline for the assay endpoints, and ensures that the assay conditions do not inappropriately result in an irritant response. Any of the three commonly used negative controls (i.e., MEM without phenol red, 0.9% saline, or sterile, deionized water) is acceptable as long as the same negative control is used consistently within a laboratory.

2.2.7.2 *Positive Controls*

As discussed by Harbell and Curren (2002), the function of the positive control is to ensure the test system is operating within normal limits and each experiment is properly executed, such that the toxic effects of interest can be properly detected. A concurrent positive control is included in each experiment to develop a historical database. Results from the positive control are compared to the historical control range and used to evaluate whether a particular study is acceptable. Because the positive control should allow for detection of an over- or under-response in the assay, the selected positive control should not produce responses at either the extreme low or the extreme high end of assay response.

In the BCOP assay, different positive controls are used for the testing of liquid and solid test substances because of the different protocols for these two types of substances. Harbell and Curren (1998) recommend positive controls that produce both opacity and permeability (e.g., ethanol for liquid test substances and imidazole for solid test substances) in the BCOP assay. About half of the BCOP studies used one or more positive control substances. The most frequently used positive control for testing liquid test substances was 100% ethanol (Swanson et al. 1995; Cassidy and Stanton 1997; Bruner et al. 1998; Southee 1998; Cooper et al. 2001; Jones et al. 2001). Acetone (Gettings et al. 1996; Chamberlain et al. 1997; Harbell and Curren 1998) and N,N-dimethylformamide (Balls et al. 1995) were used less frequently. For solid test substances, only imidazole was used.

Based on historical use in the BCOP assay, 100% ethanol or 100% acetone are the most commonly used positive controls for liquid test substances, while 20% (w/v) imidazole prepared in saline appears to be the only positive control used for solid test substances. Inclusion of a known severe ocular irritant substance in each experiment as a positive control demonstrates the functional adequacy of the test method and the consistency of laboratory operations in accurately identifying ocular corrosives and severe irritants. A positive control not only ensures the integrity of the test system and its proper execution, but also provides a measure of test method performance over time.

2.2.7.3 *Solvent Control*

The protocol for testing solids requires that the test substance be dissolved or suspended in saline or water, which also are used for the negative control. However, other solvents are generally not used in the BCOP assay, following on the practice of not using solvents to dissolve test substances in the *in vivo* rabbit eye test. However, it would seem prudent that if a special solvent (other than sterile, deionized water or saline) is used to dissolve test

substances, a solvent control be added to the BCOP study. Such a control demonstrates that the solvent does not interfere with the test system.

2.2.7.4 *Benchmark Substances*

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

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

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

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

2.2.8.1 *Negative/Solvent Controls*

A majority of BCOP studies reported using negative controls to correct the opacity and permeability values of the treated corneas. No range of acceptable/unacceptable values for the negative control was found.

Historically, solvent controls have not been used in the BCOP assay.

It would seem appropriate to establish an upper limit of both opacity and permeability for the negative or solvent control. Negative and solvent controls must produce the anticipated response to ensure the test system is functioning properly and that the specific test is valid.

2.2.8.2 *Positive Controls*

In the BCOP studies that used positive controls, the accepted range were typically an *In Vitro* Irritancy Score that fell within two SDs of the historical mean value for the testing facility. The accepted range is updated every 3 months at IIVS.

An example of historical data for positive controls was provided by IIVS (current as of July 22, 2004), as shown in the table below.

Positive Control	Opacity	OD_{490}	*In Vitro* Score
Ethanol (10 min exposure)			
Mean (n = 632)	31.2	1.422	52.7
SD	4.8	0.345	6.4
CV	15.3%	24.3%	12.1%
Upper and lower limits	21.7 – 40.7	0.742 – 2.112	39.9 – 65.4
Imidazole (4 hour exposure)			
Mean (n = 125)	76.4	1.768	103.0
SD	18.4	0.488	16.6
CV	24.1%	27.6%	16.2%
Upper and lower limits	39.7 – 113.2	0.792 – 2.745	69.7 – 136.2

CV = Coefficient of variation; n = Number of tests; SD = Standard deviation.

Positive controls are typically used as one of the criteria for determination of a valid test. If the positive control value falls within the accepted range, the test is considered valid. If the positive control value falls outside of the accepted range, the test may need to be repeated.

2.2.8.3 *Benchmark Substances*

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

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

2.2.9.1 *Corneal Opacity*

Corneal opacity is measured quantitatively with an opacitometer (e.g., ElectroDesign, Riom, France), which measures differences in light transmission between treated corneas and an air blank. Numerical opacity values with arbitrary units are obtained, with values ranging from 0 to 225. Higher opacity values have been reported by Swanson et al. (1995). Raw data are typically recorded in laboratory notebooks and electronically.

2.2.9.2 *Permeability*

The amount of dye that permeates the cornea is determined by measuring the OD/absorbance of the medium in the posterior chamber with a spectrophotometer set at 490 nm. Raw data are typically recorded in laboratory notebooks and electronically.

2.2.9.3 *Histology*

IIVS notes that they typically record histological observations of treated corneas electronically. The data include observations on each corneal tissue layer, in addition to information related to the specific BCOP study, such as test substance concentration, exposure time, and post-exposure time. Additionally, photomicrographs are prepared for illustrative purposes. These images are prepared using a Spot Insight Digital camera and Spot 4.0.8 software (Diagnostic Instruments, Inc., Sterling Heights, Michigan). Each photomicrograph is stored in an appropriate digital image log (Harbell J, personal communication).

Scoring of corneal lesions involves recording the nature, degree, and depth of the lesion observed in each tissue layer. The predominant lesions observed across the individual corneas within a treatment group are noted and serve the basis for the overall evaluation for a treatment group (Harbell J, personal communication).

2.2.10 Type of Media in Which Data Are Stored

It can be inferred that studies performed in compliance with GLP guidelines (e.g., Balls et al. 1995; Swanson et al. 1995; Swanson and Harbell 2000; Southee 1998; Bailey et al. 2004) stored the data in a manner suitable for GLP compliant studies. It would seem appropriate that data from the BCOP be stored and archived in a manner consistent with international GLP guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines are nationally and internationally recognized rules designed to produce high-quality laboratory records. These guidelines provide a standardized approach to report and archive laboratory data and records, and information about the test protocol, to ensure the integrity, reliability, and accountability of a study (EPA 2003a,b; FDA 2003).

2.2.11 Measures of Variability

Variability in the BCOP assay has been traditionally evaluated by calculating the mean (± SD) for the opacity values and the OD_{490} values for each treatment group and control group. Calculation of the mean score and SD provides the user with information on the performance of the test method. These values allow for an assessment of the performance of the test conducted and whether the observed variability between replicates is greater than would be considered acceptable.

2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data

A majority of early BCOP studies used the mean opacity and mean permeability values (OD_{490}) for each treatment group to calculate an *in vitro* score for each treatment group:

***In Vitro* Irritancy Score** = mean opacity value + (15 x mean OD_{490} 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 derived nearly identical equations. However, Casterton et al. (1996) reported evaluating the opacity and permeability values independently.

As experience was gained with the assay and additional chemical and product classes were tested, it was found that some substances can induce significant permeability without an appreciable increase in opacity, and vice versa. For example, the anionic surfactant sodium lauryl sulfate (5%) can destroy the corneal epithelium and produce a high permeability value (OD_{490} = 2.538) without producing significant opacity (value of 7.7) (Cater et al. 2001). Other anionic and nonionic surfactants (Harbell J, personal communication), as well as some surfactant-based product formulations (Gettings et al. 1996), produce similar results in the BCOP assay. Therefore, while the *In Vitro* Irritancy Score has been used historically in the BCOP assay to provide a numerical value for comparison of the relative irritancy of test

substances, this scoring system is not applicable for substances that produce irritation through only one of the two assay endpoints.

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

Once the opacity and OD_{490} values have been corrected for background opacity and the negative control values, they are entered into the formula for an *In Vitro* Irritancy Score. *In vitro* irritancy categories have been historically assigned based on predetermined ranges. The original prediction model was proposed by Gautheron et al. (1994) as follows:

***In Vitro* Score Range**	***In Vitro* Classification**
0 - 25	mild irritant
25.1 - 55	moderate irritant
55.1 - 80	severe irritant

This same prediction model was used for the EC/HO validation study (Balls et al. 1995), with the exception that the investigators added a fourth classification of "very severe" for substances that produced an *in vitro* score greater than 80.1.

This original classification system was based on studies with pharmaceutical intermediates exposed for 10 minutes (liquids) or four hours (solids).

For the European Community prevalidation study, the investigators attempted to relate the prediction model to *in vivo* data (MMAS scores) (Southee 1998):

Draize Scale	**Draize Classification**	***In Vitro* Scale**	***In Vitro* Classification**
0 - 0.9	Minimal	0 - 3	Nonirritant
1 - 25	Minimal/slight	3.1 - 25	Mild irritant
26 - 56	Moderate	25.1 - 55	Moderate irritant
57 - 84	Marked	55.1 - 80	Severe irritant
85 - 110	Extreme	> 80.1	Very severe irritant

Most other BCOP studies used the following *in vitro* classification system for BCOP *In Vitro* Irritancy Scores:

***In Vitro* Score Range**	***In Vitro* Classification**
0 - 25	Mild irritant
25.1 - 55	Moderate irritant
> 55.1	Severe irritant

Casterton et al. (1996) assigned irritation classes based on the endpoint (opacity or permeability) with the highest score for its respective range:

***In Vitro* Opacity or Permeability Ranges**	***In Vitro* Classification**
Opacity < 0.400 or Permeability < 0.175	Mild irritant
0.400 ≤ Opacity < 1.300 or 0.175 ≤ Permeability < 0.600	Moderate irritant
Opacity > 1.300 or Permeability > 0.600	Severe irritant

Some companies, such as S.C. Johnson & Son, Inc., do not use any of the classification schemes described above, but instead compare BCOP data for newly tested substances to benchmark materials, relying on a system of comparative toxicity instead of cutoff scores (Cuellar N and Swanson J, personal communication).

However, based on historical usage, it would seem appropriate that an *In Vitro* Irritancy Score of 55.1 and above be used for identification of ocular corrosives and severe irritants. However, this score is not appropriate for anionic and nonionic surfactants since they can damage the epithelium and produce high permeability values, without inducing opacity. For anionic/nonionic surfactants and other substances that produce significant permeability but minimal opacity, a permeability value > 0.600 may be a more appropriate threshold for a severe response. Benchmark substances are recommended for assaying the responses of test substances of different product or chemical classes. Additionally, histological evaluation of the corneas can be instrumental in identifying occult changes (e.g., peroxide-induced stromal damage) (Harbell and Curren 1998), and may reduce false negative results, especially for substances that do not produce significant opacity and/or permeability in the BCOP assay.

Based on an accuracy assessment (see **Section 6.0**) of seven BCOP studies that evaluated severe *in vivo* eye irritants (GHS Category 1), use of an *In Vitro* Irritancy Score of 55.1 and above, or a permeability value > 0.600 as a threshold identifies a majority (84%, 36/43) of the severely irritating chemicals tested (see **Section 6.0**).

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

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

Test and Control Substances

- chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known
- the CAS Registry Number (RN), if known
- purity and composition of the substance or preparation (in percentage(s) by weight), to the extent this information is available
- physicochemical properties such as physical state, volatility, pH, stability,

chemical class, water solubility relevant to the conduct of the study
- treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- stability, if known

Information Concerning the Sponsor and the Test Facility
- name and address of the sponsor
- name and address of the test facility
- name and address of the Study Director

Justification of the Test Method and Protocol Used

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

Criteria for an Acceptable Test
- acceptable concurrent negative control ranges based on historical data
- acceptable concurrent positive control ranges based on historical data
- if applicable, acceptable concurrent benchmark control ranges based on historical data

Test Conditions
- description of test system used
- calibration information for measuring device used for measuring opacity and permeability (e.g., opacitometer and spectrophotometer)
- supporting information for the bovine corneas used including statements regarding their quality
- details of test procedure used
- test concentration(s) used
- description of any modifications of the test procedure
- reference to historical data of the model (e.g., negative and positive controls, proficiency substances, benchmark substances)
- description of evaluation criteria used

Results
- tabulation of data from individual test samples (e.g., opacity and OD_{490} values and calculated *in vitro* irritancy score for the test substance and the positive, negative, and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means ± SD for each experiment)

Description of Other Effects Observed

Discussion of the Results

Conclusion

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

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

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

2.3 Basis for Selection of the Test Method System

As discussed in **Section 1.1.1**, the assay developers wanted to develop a cornea-based assay, because the cornea is one of the main targets during accidental eye exposures. In addition, corneal effects are weighted heavily in the original *in vivo* ocular irritancy scoring systems (e.g., 80 out of a possible 110 points in the Draize eye test scoring system). Opacity in the isolated cornea was initially investigated since it is the only corneal endpoint graded in many *in vivo* ocular irritancy assays. Studies indicated, however, that some known irritant substances, such as sodium lauryl sulfate and certain medium-length chained alcohols, destroy the corneal epithelium without producing significant opacity. Damage to the epithelium was subsequently quantified for these substances by measuring penetration of the dye sodium fluorescein through the isolated cornea. Gautheron and colleagues refined the assay to measure opacity and permeability, two important components of ocular irritation, and found that the two endpoints predicted the ocular irritancy of a variety of substances (Gautheron et al. 1992, 1994; Sina and Gautheron 1998).

Use of the BCOP test method offers some advantages over the traditional *in vivo* rabbit eye test. Bovine eyes are a relatively inexpensive, abundant by-product of the beef industry. Since the cornea is isolated from animals slaughtered for other purposes, the test method avoids the use of living animals bred specifically for the purpose of toxicity testing. The endpoints of opacity and permeability are measured quantitatively, minimizing the potential variability that could result from subjective evaluations used in the traditional *in vivo* rabbit eye test. The BCOP test method also allows precise control over the test substance volume, concentration and exposure time, as well as the post-exposure period during which irritation is expressed in the isolated cornea. Thus, different exposure and post-exposure conditions can be readily modeled in this system. Finally, when histology is added to the BCOP assay, a permanent record of the tissue is available.

2.4 Proprietary Components

The BCOP assay does not employ any proprietary components.

2.5 Basis for the Number of Replicate and Repeat Experiments

2.5.1 Sample Replicates

The numbers of corneas used to test a substance varied from study to study with three to six corneas used per compound. Early studies using the BCOP test method used six corneas (Gautheron et al. 1992, 1994). In the first interlaboratory study of the BCOP assay, Gautheron et al. (1994) observed that reducing the number of treated corneas to three did not adversely affect the assay results. There appeared to be a close correlation between scores obtained using three and six corneas. The authors concluded that three corneas were likely sufficient to obtain valid results.

2.5.2 Experimental Replicates

None of the published reports indicated that repeating experiments is necessary. However, based on sound scientific judgment, it would seem reasonable to expect that equivocal responses or divergent results among test cornea would mandate repeating the experiment.

2.6 Compliance with Good Laboratory Practice Guidelines

Southee (1998) reported that the BCOP studies were performed in compliance with GLP guidelines for nonclinical laboratory studies. IIVS also conducts GLP-compliant BCOP assays (e.g., Swanson et al. 1995; Gettings et al. 1996; Swanson and Harbell 2000; Bailey et al. 2004). However, other study reports did not note that the studies were conducted consistent with GLP guidelines. Conducting studies under GLP guidelines increases confidence in the quality and reliability of test data. Furthermore, if data using this test method is to be submitted to the EPA or another agency in response to Federal testing requirements, then compliance with appropriate GLP guidelines will be required.

2.7 Study Acceptance Criteria

A test is acceptable if the positive control(s) gives an *In Vitro* Irritancy Score that falls within two SDs of the current historical mean, which is to be updated every three months. The negative/solvent control responses should result in opacity and permeability values that are less than the laboratory's established upper limits of opacity and permeability values for bovine corneas treated with the respective negative or solvent control.

3.0 SUBSTANCES USED FOR VALIDATION OF THE BCOP TEST METHOD

3.1 Rationale for the Substances or Products Selected for Use

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

Of the 23 BCOP reports considered in developing this BRD, only eight contained or provided sufficient *in vitro* and *in vivo* data for an accuracy analysis[1]. These eight reports are: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Casterton et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004).

A total of 161 substances and formulations were evaluated in the eight studies, of which 69 were commercial products or formulations. **Sections 3.1.1** through **3.1.8** address the rationale for the chemicals or products tested in each of these studies.

3.1.1 Gautheron et al. (1994)

In the EC interlaboratory assessment of the BCOP assay, 52 substances were studied, including 22 liquids, 22 solids, and eight surfactants (both solids and liquids). The substances were selected to:

- represent a broad range of chemical classes and structures (e.g., alcohol, polycyclic aromatic hydrocarbon, acid, base, ether, phenol, halogenated hydrocarbon)
- include a wide range of solubilities
- cover the range of ocular irritancy categories *in vivo*, from nonirritant to severe eye irritant (i.e., MAS scores ranging from 1.3 to 103)

One of the test substances, thiourea, was found to be extremely toxic via ocular exposure by Balls et al. (1995), killing the three rabbits on which it was tested. Thiourea was excluded from the accuracy and reliability analyses for the Balls et al. (1995) study. For consistency, thiourea also was excluded from the accuracy and reliability analyses for the Gautheron et al. (1994) study. Therefore, the final list of test substances included a total of 51 substances available for the accuracy and reliability analyses in **Sections 6.0** and **7.0**. However, for the EPA (EPA 1996), EU (EU 2001), and GHS (UN 2003) classification systems, three, three, and two of the *in vivo* studies, respectively, did not provide sufficient data to assign an ocular irritancy classification.

---

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

3.1.2 Balls et al. (1995)

In the EC/HO validation study, the test substances were initially selected from the 1992 ECETOC Eye Irritation Reference Chemicals Data Bank (ECETOC 1992) based on the following criteria:

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

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

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

The final list of test substances included a total of 51 substances, four of which were tested at two different concentrations and two of which were tested at three concentrations, for a total of 59 different tests used for the accuracy and reliability analyses in **Sections 6.0** and **7.0**. For the EPA (EPA 1996), EU (EU 2001), and GHS (UN 2003) classification systems, six, nine, and five of the *in vivo* studies, respectively, did not provide sufficient data to assign an ocular irritancy classification.

3.1.3 Swanson et al. (1995)

Twenty full-strength industrial and household cleaning formulations were evaluated undiluted to determine the utility of the BCOP assay to predict the ocular irritation potential of these types of products. The substances were surfactant-based aqueous product formulations with pH values ranging from 1 to 14. Product types include toilet bowl cleaner, floor cleaner, meat room degreaser, all-purpose cleaner, bathroom cleaner, pot and pan cleaner, floor stripper, glass cleaner, and metal cleaner. However, only a subset of nine of these substances could be included in the accuracy evaluations described in **Section 6.0**, since *in vivo* ocular irritation classifications (i.e., EPA 1996, EU 2001, UN 2003) could not be assigned to 11 substances (see **Section 4.0**) that had been evaluated using a modified

rabbit eye test protocol which used a 30 μL test substance volume instead of the 100 μL volume on which the EPA, EU, and GHS ocular irritancy classification systems are based.

3.1.4 Gettings et al. (1996)

This report described results from Phase III of the CTFA Evaluation of Alternatives Program, a three-phase program that evaluated promising *in vitro* alternative test methods in relation to the *in vivo* rabbit eye test. Each phase of the program evaluated a specific product type; Phases I and II evaluated hydro-alcoholic and oil/water formulations, respectively, while Phase III evaluated surfactant-based personal care cleansing formulations. The 25 products tested in Phase III were representative surfactant-containing cleansing formulations, such as hair shampoos, liquid soap, eye make-up remover, and bubble bath. The selected formulations were chosen to provide a range of ocular irritancy responses in the *in vivo* rabbit eye test (from nonirritating to moderately irritating), which is the highest level of irritancy generally achieved by this class of products. Because it was found that a majority of the formulations produced irritant responses either in the middle (MAS ~ 45) or the nonirritating (MAS ~ 0) end of the Draize ocular irritation range, a decision was made to test dilutions (25% v/v in distilled water) of 10 of the products at the middle of the range to have a more uniform distribution of irritant responses. While there were 25 substances available for the accuracy and reliability analyses in **Sections 6.0** and **7.0**, for the EPA (EPA 1996) and EU (EU 2001) classification systems, two of the *in vivo* studies did not provide sufficient data to assign an ocular irritancy classification.

3.1.5 Casterton et al. (1996)

Ninety-seven test substances were selected primarily based on the availability of historical *in vivo* rabbit eye data. Fifteen of the test substances evaluated in the BCOP test method were selected from the formulations tested in the CTFA Evaluation of Alternatives Program – Phase III, and 48 were selected from the substances included in the ECETOC Eye Irritation Reference Chemicals Data Bank (ECETOC 1992). Twenty-one test substances were Amway products with *in vivo* data, while the remaining substances were surfactant raw materials with *in vivo* data available from the suppliers. A secondary rationale was to evaluate a wide range of chemicals and products, both industrial and consumer. However, detailed *in vivo* reference data were available for only a subset of 56, 54, or 55 of these substances for the EPA (EPA 1996), EU (EU 2001), and GHS (UN 2003) classification systems, respectively, as described in **Section 4.0**.

3.1.6 Southee (1998)

The selection of the 16 test substances in this BCOP study was based on including substances that represented a range of physical forms and irritancy and also had high quality *in vivo* eye irritation data. The test substances were selected from substances included in the ECETOC Eye Irritation Reference Chemicals Data Bank (ECETOC 1992). Fourteen of the substances had sufficient *in vivo* data to assign EPA (EPA 1996) and EU (EU 2001) classifications, while 15 of the substances had sufficient *in vivo* data to assign GHS (UN 2003) classifications.

3.1.7 Swanson and Harbell (2000)

Thirteen test substances were selected to evaluate the effect of increasing concentrations of ethanol and other solvents on the ocular irritancy of insect repellent formulations, while maintaining a constant concentration of the active ingredient. However, detailed *in vivo* reference data were available for only a subset of nine these substances, as described in **Section 4.0**.

3.1.8 Bailey et al. (2004)

The 16 test substances in this study were selected to evaluate whether the BCOP assay was useful for predicting the ocular irritation potential of unique petroleum products (e.g., lubricant additive packages, base stocks, cutting fluids, solvents, monomers). Test substances included solids, nontransparent, transparent, and semiviscous or viscous liquids. Thirteen of the substances had sufficient *in vivo* data to assign EPA (EPA 1996) and EU (EU 2001) classifications, while 14 of the substances had sufficient *in vivo* data to assign GHS (UN 2003) classifications.

3.2 Rationale for the Number of Substances Tested

The rationale for the number of substances tested in the studies is not known.

3.3 Chemicals or Products Evaluated

Descriptive information for each of the substances tested in the BCOP assay was obtained, to the extent possible, from the information provided in the study reports. When provided, the specific information extracted for each substance included its name, source/supplier, purity, CASRN, product class, concentration tested, and the study citation. No attempt was made to identify the source/supplier or the purity of a substance if the information was not included in the study report. 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. **Appendix B** provides the available information on the name, CASRN, and chemical/product class of each substance evaluated in the BCOP test method. Components of the formulations tested in the BCOP assay are also provided in **Appendix B**, to the extent this information was available. **Tables 3-1** and **3-2** provide the chemical and product classes, respectively, of the test substances evaluated with the BCOP assay. Because the purity, source/supplier, and concentration of substances tested in multiple laboratories varied depending on the testing laboratory, this study specific information is provided **Appendix B** with the BCOP test method data.

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

Chemical Class	# of Substances
Acyl halide	3
Alcohol	22
Aldehyde	1
Alkali	3
Aluminum compound	1
Amide	2
Amidine	6
Amine	10
Amino acid	4
Boron compound	1
Carboxylic acid	17
Ester	12
Ether/Polyether	9
Formulation	69
Heterocyclic compound	12
Hydrocarbon	18
Imide	2
Inorganic salt	6

Chemical Class	# of Substances
Ketone	12
Lactone	3
Nitrile compound	1
Nitro compound	2
Oil	1
Onium compound	12
Organic salt	3
Organic sulfur compound	5
Organophosphate	1
Organosilicon compound	1
Phenol	1
Polycyclic compound	3
Terpene	1
Wax	1

As shown in **Table 3-1**, 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. Other chemical classes tested include amines, ethers/polyethers, inorganic and organic salts, and organic sulfur compounds. The formulations tested include hair shampoos, personal care cleansers, detergents, bleaches, insect repellents, petroleum products, and fabric softener.

As shown in **Table 3-2**, 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 detergents, insect repellents, lubricants, personal care cleansers, pesticides, and plasticizers.

3.3.1 Gautheron et al. (1994)

Regarding descriptive information about the test substances, the EC interlaboratory study report includes specific chemical names of the 52 test substances, but not chemical and product classes. The physical form, and the CASRN of the test substances also are provided. Liquids were tested undiluted, while surfactants were tested at 10% in assay medium. Solids were tested either as a solution or suspension at 20% in assay medium. However, chemical characteristics, purity, and stability of the test substance in the test medium were not described.

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

Product Class	# of Substances
Adhesive	1
Agricultural chemical	2
Antifreeze agent	1
Bactericide/Fungicide/ Disinfectant/Germicide	11
Beverage	1
Bleach	3
Chelating agent	2
Chemical/synthetic intermediate	28
Cleaner	15
Cleanser (personal care)	13
Coupling agent	1
Cutting fluid	2
Degreaser	1
Dessicant	1
Detergent	11
Drug/Pharmaceutical/ Therapeutic agent and/or Metabolite	17
Dry cleaning preparation	1
Dye, in manufacture of	3
Emulsifier	1
Etching and/or electroplating	2
Explosive	1
Fabric softener	1
Fertilizer	1

Product Class	# of Substances
Flame retardant	1
Flavor ingredient	3
Food additive	1
Herbicide	3
Insect repellant	8
Lubricant/lubricant additive	6
Paint, lacquer, varnish (component)	1
Pesticide	8
Petroleum product	16
Photographic chemical/ developing agent	2
Plant growth regulator	2
Plasticizer	4
Preservative	2
Reagent	5
Shampoo (hair)	14
Soap	3
Solvent	34
Surfactant	39
Anionic surfactant	3
Cationic surfactant	6
Nonionic surfactant	5
Thermometer fluid	1

3.3.2 Balls et al. (1995)

The 51 substances tested in the EC/HO validation study, included a wide range of chemical and product classes. For each test substance, the authors provided a CASRN, chemical class, source/supplier, catalog number, purity, form tested, and concentration tested in the study report.

3.3.3 Swanson et al. (1995)

Twenty full-strength industrial and household cleaning formulations were evaluated undiluted in this study. The materials were surfactant-based aqueous product formulations with pH values ranging from 1 to 14. Product types include toilet bowl cleaner, floor cleaner, meat room degreaser, all-purpose cleaner, bathroom cleaner, pot and pan cleaner, floor stripper, glass cleaner, and metal cleaner. The authors of this study provided the components (percent composition) and pH of each formulation. The ingredients that contribute to irritancy were provided in the study publication. The formulas are from S.C. Johnson & Son, Inc. and JohnsonDiversey, Inc.

3.3.4 Gettings et al. (1996)

In this study, 25 surfactant-based cleaning formulations were evaluated in the BCOP assay, with each product tested as a 10% (w/v) solution of the formulation that had been tested *in vivo* at either 100% or a dilution of 25%. Generic names of the formulations were provided in the study report, such as Baby Shampoo No. 1, Mild Shampoo, Liquid Soap No. 1, Gel Cleaner, Skin Cleaner, Bubble Bath, and Eye Make-Up Remover. The components of each formulation were provided in the study report, including percent concentration (w/w). However, the sources/suppliers of the formulations were not provided.

3.3.5 Casterton et al. (1996)

The only descriptive information about the test substances provided in the study report is the name of each chemical or formulation tested in the BCOP assay. However, some descriptive information, such as CASRNs and chemical/product classes, could be readily obtained from other sources for a majority of the test substances.

3.3.6 Southee (1998)

In the study report for the European Community Prevalidation Study of the BCOP assay, the authors provided the chemical name, CASRN, source/supplier, catalog number, purity, form tested, and concentration tested for each test substance.

3.3.7 Swanson and Harbell (2000)

Ethanol and 12 ethanol-containing insect repellent formulations were evaluated in this study. The concentration of the active ingredient was the same in all the formulations, but the concentration of ethanol and other organic solvents varied among the formulations. The authors of this study provided the components (percent composition) of each formulation. The test substances were obtained from S.C. Johnson & Son, Inc.

3.3.8 Bailey et al. (2004)

The study report provided the name and a physical description of each test substance, and the pH for liquid materials. Information about product classes also was provided.

3.4 Coding Procedures Used in the Validation Studies

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

3.4.1 Gautheron et al. (1994)

Coding of test substances was used during the EC study. Chemicals were sampled, coded, and shipped by an independent company (MCS-Pharma, Erstein, France). The study participants were aware of the identities of the substances to be tested, and, for safety reasons, received substance codes and safety sheets to be used in case of an emergency.

3.4.2 Balls et al. (1995)

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

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

3.4.3 Swanson et al. (1995)

The formulations were coded when tested in the BCOP assay. Test substances were assigned a numeric code by S.C. Johnson & Son, Inc., and testing was performed by laboratory personnel without knowledge of the identities of the formulations.

3.4.4 Gettings et al. (1996)

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

3.4.5 Casterton et al. (1996)

Coding procedures were not discussed in this study report.

3.4.6 Southee (1998)

Test substances were each assigned a numeric code, specific for each of the three testing laboratories. BIBRA performed the coding and distribution of the test substances, which were tested blind by each of the testing laboratories. Each laboratory tested 10 coded substances on two separate occasions. The results from each laboratory were sent to BIBRA for data analysis (i.e., comparison with the proposed prediction model). The codes were broken in October 1997 for subsequent statistical analysis of the data.

3.4.7 Swanson and Harbell (2000)

The formulations were coded when tested in the BCOP assay. Test substances were assigned a numeric code by S.C. Johnson & Son, Inc., and testing was performed by laboratory personnel with knowledge that the formulations were solvent-based (primarily ethanol) insect repellents.

3.4.8 Bailey et al. (2004)

The substances were coded when tested in the BCOP assay. Testing was performed by laboratory personnel without knowledge of the identities of the formulations.

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

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

4.1.1 Draize Rabbit Eye Test

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

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

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

4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols

Since the original description of the *in vivo* rabbit eye test method, regulatory agencies in the U.S., as well as in other countries, have modified the test method protocol to suit their specific needs and goals in protecting human health (**Table 4-2**). Regulatory agencies generally recommend using healthy adult albino rabbits (e.g., White New Zealand). The

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

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

[1]From Draize et al. (1944)

[2]Scores of 0 are assigned for each parameter if the cornea, iris, or conjunctiva are normal.

eyes of each test rabbit are examined within 24 hours prior to test initiation. A quantity of 0.1 mL (for liquids) or 0.1 g (for pulverized solid, granular, or particulate test substances) is placed into the conjunctival sac of one eye of each rabbit, after pulling the lower lid away from the eyeball. The other eye remains untreated. The lids are held together for about one second to decrease loss of test substance from the eye. Although the observation period varies, the eyes are typically examined at 24-hour intervals for at least 72 hours after application of the test substance for adverse effects to the cornea, conjunctiva, and iris. The length of the observation period should be sufficient to evaluate reversibility of any of the observed effects, but generally does not exceed 21 days. The ocular effects observed are usually those described by Draize et al. (1944) in **Table 4-1**. For current uses, other lesions,

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

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

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

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

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

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

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

such as pannus[1] and herniation of the cornea, also are noted. Corneal, iris, and conjunctival lesions are scored using the individual numerical grades described in **Table 4-1**, but weighted scores and an overall score for irritation are not typically calculated or used for U.S. or European regulatory purposes.

Depending on the regulatory agency, the number of rabbits required for a study of ocular irritation can vary. To minimize pain and suffering of rabbits exposed to potentially corrosive agents, the EPA and European regulatory agencies suggest that, if a test substance is anticipated to produce a severe effect (e.g., corrosive effect), a test in a single rabbit may be conducted. If a severe effect is observed in this rabbit, further testing does not need to be conducted and classification and labeling of a test substance can proceed on the effects observed in a single rabbit. In cases where more than one rabbit is tested, at least three should be examined to classify the ocular effects produced by the test substance (EU 2004; EPA 1998). In contrast, regulations for other U.S. agencies (e.g., CPSC, FDA) require at least six rabbits be examined to classify the effects produced by a test substance (CPSC

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

2003). The differences in current *in vivo* test protocols in the U.S. appear to reflect each agency's objectives for eye irritation testing; EPA regulates industrial chemicals while the CPSC and FDA regulate household consumer products, pharmaceuticals, cosmetics, and toiletries.

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

4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems

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

The FHSA guideline states that a test substance is considered an eye irritant if four or more of six rabbits have positive ocular scores in nonirrigated eyes within 72 hours after instillation of the test substance (CPSC 2003). A positive score is defined by corneal opacity or iritis scores of ≥1, or conjunctival redness or chemosis scores of ≥2. In addition, if only one of the six rabbits shows ocular effects within 72 hours, the test substance in considered nonirritating to the eye. If two or three rabbits have positive ocular scores, the test is repeated in a second group of six rabbits. Then, if the criteria for an ocular irritant for the

second test (three or more positive rabbits) or a nonirritant (0 positive rabbits) are met, a classification is made. However, if only one or two rabbits have positive scores in the second test, the test is repeated a third and final time. If one or more rabbits have positive ocular scores in the third test, the test substance is classified as an ocular irritant. If none of the rabbits have positive ocular scores in the third test, the test substance is classified as a nonirritant (CPSC 2003).

The EPA classification guideline considers the kinds of ocular effects produced in the *in vivo* rabbit eye test, as well as the reversibility and the severity of the effects (EPA 1996). However, unlike the FSHA system, incidence is not considered, as classification is based on the rabbit that exhibits the most severe response in a group of three or more rabbits. Data from all observation times are used for EPA classification. Corneal opacity or iritis scores of ≥1, or conjunctival redness or chemosis scores of ≥ 2 define a positive score. EPA labeling regulations also require an assessment of the reversibility of positive scores. If a positive score persists for > 21 days, the substance is classified as a Category I eye irritant, which is defined as "corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for > 21 days." Substances that cause positive corneal opacity, iritis, or conjunctival scores that clear in 8 to 21 days are designated as Category II eye irritants. If positive scores induced by a substance clear within 7 days, the substance is labeled Category

III. A minimal effect (i.e., inconsequential or complete lack of irritation), or an effect that clears within 24 hours of application is designated as Category IV.

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

The GHS for the classification and labeling of hazardous chemicals (UN 2003) is an initiative developed through the cooperative efforts of the International Labour Office, the OECD, and the UN to promote an internationally-harmonized approach for classifying chemicals according to their health hazards. For the purpose of harmonizing classification of ocular irritants, the UN adopted an approach put forth by the OECD in its *Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods* (OECD 1996). A tiered testing and evaluation strategy using available data from dermal irritation studies, data from validated alternative toxicological methods, knowledge of structure activity relationships, and screening for pH extremes (≤ 2 or ≥ 11.5; considering acid or alkaline reserve) has been proposed (UN 2003). In addition, a single harmonized hazard category is proposed for irreversible effects on the eye/serious damage to eye (Category 1). Irreversible effects according to the GHS system include grade 4 corneal lesions at any time during the *in vivo* test, positive responses on day 21 (e.g., score >0 for any endpoint evaluated), and cases where two or more of three rabbits exhibit a mean score (24-, 48-, 72-hours) for corneal opacity ≥ 3 and/or iritis > 1.5. A single harmonized hazard category, Category 2, is proposed for reversible effects on the eye; however, for regulatory authorities that prefer to distinguish irritants in this group, subcategories have been developed based on whether effects reverse within 7 or 21 days. Category 2A is defined as an eye irritant with effects that fully reverse within 21 days. Category 2B is considered mildly irritating to the eyes, and is designated for substances whose effects reverse fully within 7 days. Reversible effects include positive responses in two or more of three rabbits, where the mean score (24-, 48-, 72-hours) for corneal opacity or iritis ≥ 1 (but < 3 or < 1.5, respectively), or conjunctival redness or chemosis ≥ 2. Additional details on the GHS classification system are provided in **Section 4.3**.

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

The BCOP studies evaluated in this document include *in vivo* reference data generated using the basic procedures described above for the *in vivo* rabbit eye test method. For the Gautheron et al. (1994) study, 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 D**).

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.

The MMAS were developed from Draize scores calculated 24 hours or more after instillation of the test substance. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for each of these substances are available in the ECETOC Reference Chemicals data bank (ECETOC 1998). All 59 of these substances were classified by NICEATM according to the EU (2001) classification system; only 55 and 57 substances, respectively, were classified according to the EPA (1996) and the GHS (UN 2003) ocular irritancy classification systems, due to lack of sufficient *in vivo* data (**Appendix D**).

In 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 D**). 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 CTFA Phase III study, data were obtained from a modified Draize eye test. Details of the protocol are provided in Gettings et al (1996). Six rabbits (three male, three female) were used for each test substance. The right eye of each rabbit was anesthetized prior to instillation of 0.1 mL of test substance into the conjunctival sac. Ocular irritation was evaluated at 1 hour, and at 1, 2, 3, 4 and 7 days. If irritation persisted, ocular responses were observed at 7-day intervals up to a maximum of 21 days. MAS were determined according to Williams et al. (1982). Data were classified according to the scheme proposed by Kay and Calandra (1962) and the FHSA (1947). MAS, maximum average total scores for each endpoint (cornea, iris, conjunctiva), number of positive responses, maximum day to clear, and FHSA and Kay/Calandra irritancy categories are reported in the paper for the 25 test substances. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for each of these substances were provided by the CTFA. The 25 substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For the 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, while 14 of the substances have been classified according to the EPA (1996) and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For the Casterton et al. (1996) study, the authors noted that they used *in vivo* reference data from existing sources. Fifteen of the test substances evaluated in the BCOP test method were selected from the formulations tested in the CTFA Evaluation of Alternatives Program – Phase III, and 48 were selected from the substances included in the ECETOC Eye Irritation Reference Chemicals Data Bank (ECETOC 1992). Twenty-one test substances were Amway products with historical *in vivo* data, while the remaining substances were surfactant raw materials with *in vivo* data available from the suppliers. Only a subset of these data were available to NICEATM. The Access Business Group provided copies of original study reports containing *in vivo* reference data for 13 of the Amway product formulations evaluated in Casterton et al. (1996). Detailed *in vivo* data for the 15 surfactant-based formulations tested in Gettings et al. (1996) were available from the CTFA. *In vivo* data for 32 other substances were available in ECETOC (1998).

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, utilizing six animals per substance.

ExxonMobil Biomedical Sciences, Inc. provided detailed *in vivo* reference data for the 16 petrochemical products evaluated by Bailey et al. (2004). All substances had been tested previously using the standard Draize eye irritancy test protocol, which consisted of instilling 0.1 mL of undiluted test substance into the conjunctival sac of three or six rabbits.

4.3 *In Vivo* Classification Criteria Used for BRD Analysis

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

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

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

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

4.3.1 GHS Classification Rules Used for BRD Analysis

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

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

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

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

After each rabbit was categorized, the ocular irritancy potential of the substance was determined. As shown in **Table 4-4**, substance classification depended on the proportion of rabbits that produced the same response. As noted above, if a substance was tested in more than three rabbits, decision criteria were expanded. Generally, the proportionality needed for classification was maintained (e.g., 1 out of 3 or 2 out 6 rabbits were required for classification for most categories). However, in some cases, additional classification rules were necessary to include the available data. These additional rules are distinguished by italicized text in **Table 4-4**.

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

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

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

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

4.3.2 EPA Classification Rules Used for BRD Analysis

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

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

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

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

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

4.3.3 EU Classification Rules Used for BRD Analysis

Substance classification using the EU classification system was conducted sequentially (EU 2001). While average Draize scores are used for classification, the calculation of average

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

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

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

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

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

Much of the published data on the prediction of ocular irritancy potential for test substances using the *in vivo* rabbit eye test method was limited to average score data (e.g., MAS, MMAS) or irritancy classification (e.g., mild, moderate, severe, or EU classification). An attempt was made to obtain the original records and/or compiled reports for the *in vivo* reference data. Although the original study records were not obtained for any of the studies, compiled *in vivo* data reports were obtained from the following organizations: 1) S.C. Johnson & Son, Inc. for the Swanson et al. (1995) and Swanson and Harbell (2000) studies; 2) the CTFA for the Gettings et al. (1996); 3) Access Business Group for the Casterton et al. (1996) study; and 4) ExxonMobil Biosciences, Inc. for the Bailey et al. (2004) study. Additionally, individual animal data were available from the ECETOC eye irritation data bank (ECETOC 1998).

4.5 *In Vivo* Data Quality

Ideally, all data supporting the validity of a test method should be obtained and reported from studies conducted in accordance with GLP guidelines, which are nationally and

internationally recognized rules designed to produce high-quality laboratory records (OECD 1998; EPA 2003a, 2003b; FDA 2003). These guidelines provide an internationally standardized approach for the conduct of studies, reporting requirements, archival of study data and records, and information about the test protocol, in order to ensure the integrity, reliability, and accountability of a study.

The extent to which the *in vivo* rabbit eye studies, 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 published reports. Although an attempt was made to obtain the original study records, such records could not be obtained. 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, about half of the *in vivo* studies were conducted according to GLP guidelines; for the other half, GLP compliance was not explicitly stated. 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; Gettings et al. 1996; Casterton et al. 1996; Swanson and Harbell 2000), the extent of GLP compliance was not provided, so the extent of GLP compliance is not known.

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

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

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

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

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

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

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

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

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

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

Accuracy of the *in vivo* test method would ideally be assessed by comparison of ocular effects observed in the rabbit to those effects produced in humans. A review of the literature indicates that there are few studies in which rabbit and human responses have been carefully compared under controlled conditions to assess the accuracy of the *in vivo* test method. Therefore, most studies conduct retrospective evaluations and comparisons of responses between humans and rabbits. A review indicates that a number of studies show that responses to mild to moderate irritants were generally similar between rabbits and humans (Lewin and Guillery 1913; Suker 1913; Leopold 1945; Carpenter and Smyth 1946;

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

McLaughlin 1946; Nakano 1958; Barkman 1969; Grant 1974). A review of these studies can be found in McDonald et al. (1987). For a severe irritant, Grant (1974) and Butscher (1953) showed that accidental exposure to neat thioglycolic acid produced similar responses in humans and rabbits.

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

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

Based largely on the protocol of Draize et al. (1944), the original regulatory requirements for eye irritation testing mandated the use of at least six rabbits. In recognition of animal welfare concerns, several evaluations were conducted to assess the reliability of the test method and the consequences of reducing the number of rabbits per test from six to as few as two (DeSousa et al. 1984; Solti and Freeman 1988; Talsma et al. 1988; Springer et al. 1993; Dalbey et al. 1993; Berdasco et al. 1996). With the exception of Dalbey et al. (1993), each study concluded that reducing the number of rabbits from six to three would not have an unacceptable reduction on the predictivity of ocular irritancy classification/categorization. Analyses were performed using MAS, internal irritancy classification schemes, and/or regulatory classification schemes as endpoints for comparison. Several of these studies (DeSousa et al. 1984; Talsma et al. 1988; Dalbey et al. 1993) revealed that correlations between three-rabbit and six-rabbit classifications were the highest among substances classified on the extreme ends of the irritancy range (i.e., nonirritants and severe irritants). These studies noted that the majority of variability among rabbit responses was observed among substances classified in the middle range of irritation (i.e., mild and moderate irritants). Accordingly, Dalbey et al. (1993) concluded that the observed variability in the middle range of irritation justified the continued routine use of six rabbits. However, based primarily on the results of these evaluations, the EPA (EPA 1998), EU (EU 2001), and the OECD (in revised TG 405), recommended the use of a maximum of three rabbits, although additional rabbits could be tested under certain circumstances (e.g., to confirm weak or moderate responses).

To further address the reliability of the rabbit eye test, ICCVAM and NICEATM used the available *in vivo* data to estimate the likelihood of underclassifying a positive substance or

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

5.0 BCOP TEST METHOD DATA AND RESULTS

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

As noted in **Section 3.1**, only a subset of the BCOP data obtained for this evaluation was useful for an accuracy analysis. These data were extracted from eight publications, data submissions, or study reports: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Casterton et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey (2004). The scientific methods described in these eight BCOP study reports provided various levels of detail. To the extent possible, information about the test method components discussed in **Section 2.0** was extracted from each publication and summarized in **Appendix A4**, so that any differences among the protocols are evident. Details about the following test method components are included in the appendix to the extent this information was available:

- collection of bovine eyes (e.g., transport conditions, temperature)
- cornea preparation
- pretreatment incubation/equilibration in corneal holder (e.g., duration and temperature)
- treatment groups used (i.e., number of corneas used per test substance)
- controls used
- treatment procedures for corneas
- endpoints assessed
- evaluation of test results
- calculation of *in vitro* score
- *in vitro* classification of ocular irritancy
- criteria for an acceptable test
- compliance with GLP

As is evident in **Appendix A4,** there are differences in various aspects of the test method protocols. These differences are summarized below:

- Four of the studies (Swanson et al. 1995; Gettings et al. 1996; Swanson and Harbell 2000; and Bailey et al. 2004) noted transporting the bovine eyes from the slaughterhouse to the testing facility over ice, as recommended in the proposed standardized protocol. Four other studies noted that isolated bovine eyes were transported at ambient temperature (Gautheron et al. 1994; Balls et al. 1995; Casterton et al. 1996; Southee 1998).
- Only four of the studies (Swanson et al. 1995; Southee 1998; Swanson and Harbell 2000; Bailey et al. 2004) noted transporting the bovine eyes in HBSS containing antibiotics.
- Although all studies reportedly used complete MEM for maintaining the isolated bovine corneas during incubations, only the more recent studies (Swanson et al. 1995; Casterton et al. 1996; Southee 1998; Swanson and Harbell 2000; Bailey et al. 2004) specified using MEM *without* phenol red for incubations.
- All studies used an opacitometer to measure opacity, except for Casterton et al. (1996), which used a UV/VIS spectrophotometer.

- The number of corneas used per test substance in each study varied from three to six per treatment group.
- All of the studies tested solid test substances as a 20% solution or suspension with an incubation period of four hours, with the exception of Casterton et al. (1996). In this study, solids were applied directly to the corneal surface and incubated for one hour.
- Casterton et al. (1996) independently evaluated both opacity and permeability for classifying the potential ocular irritancy of test substances. Gettings et al. (1996) used permeability values only to classify the *in vitro* ocular irritancy of the surfactant-based personal care formulation evaluated, because these materials are known to damage the epithelial layer of the cornea without producing significant opacity. In contrast, the remaining BCOP studies calculated an *in vitro* score equal to the mean opacity value plus 15 times the mean permeability value. This *in vitro* score was used to classify the ocular irritancy of test substances.
- The *in vitro* classification of severe ocular irritants was similar for Gautheron et al. (1994), Balls et al. (1995), and Southee (1998). Gautheron et al. (1994) defined a test substance as a severe irritant if it produced an *in vitro* score of 55.1 or greater. Balls et al. (1995) and Southee (1998) defined a severe irritant as one that produced an *in vitro* score between 55.1 and 80; an *in vitro* score greater than 80 was considered a very severe irritant. In contrast, Casterton et al. (1996) defined a severe irritant as a substance that produced either an opacity value greater than 1.300 or a permeability value greater than 0.600. For the surfactant-based personal care formulations evaluated by Gettings et al. (1996), it was recommended that a severe irritant be defined as a substance that produces a permeability value greater than 0.600 (Harbell J, personal communication), since these materials do not produce appreciable opacity in the isolated bovine cornea, but can damage the epithelium and increase permeability.
- Gautheron et al. (1994) evaluated the use of preserved corneas, in addition to using freshly isolated bovine corneas, in the BCOP assay.

The impact of how differences among test method protocols could impact the data and results is unknown.

5.2 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. Two *Federal Register* (*FR)* Notices (Vol. 69, No. 57, pp. 13859-13861, March 24, 2004, and Vol. 70, No. 38, pp. 9661-9662, February 28, 2005; both available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), were published requesting original BCOP test method data and *in vivo* reference data. In addition, authors of published BCOP studies were contacted to request original BCOP data and *in vivo* reference data from the 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. John Harbell submitted between-experiment (intralaboratory) permeability data for the Gettings et al. (1996) 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.

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.

5.3 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 OD_{490} values for each treatment group and control group. The mean opacity and mean permeability (OD_{490}) 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 OD_{490} 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 nonirritant 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 prediction model first proposed by Gautheron et al. (1994), for which *in vitro* irritancy categories were based on predetermined ranges of *in vitro* scores:

***In Vitro* Score Range**	***In Vitro* Classification**
0 - 25	Mild irritant
25.1 - 55	Moderate irritant
≥ 55.1	Severe irritant

This original classification system was based on studies with pharmaceutical intermediates in which bovine corneas were exposed for 10 minutes (liquids) or four hours (solids). The correlation of these categories to accepted regulatory categories for ocular irritation (i.e., GHS, EPA, EU) is unknown.

This same prediction model was used for the EC/HO validation study (Balls et al. 1995), with the exception that the investigators added a fourth classification of "very severe" for substances that produced an *in vitro* score greater than 80.1.

For the European Community prevalidation study, the investigators attempted to relate the prediction model to *in vivo* data (MMAS scores) (Southee 1998):

Draize Scale	**Draize Classification**	***In Vitro* Scale**	***In Vitro* Classification**
0 - 0.9	Minimal	0 - 3	Nonirritant
1 - 25	Minimal/slight	3.1 - 25	Mild irritant
26 - 56	Moderate	25.1 - 55	Moderate irritant
57 - 84	Marked	55.1 - 80	Severe irritant
85 - 110	Extreme	> 80.1	Very severe irritant

Gettings et al. (1996) did not report a classification scheme to assign irritancy potential to tested substances based on *in vitro* scores.

Casterton et al. (1996) assigned irritation classes based on the endpoint (opacity or permeability) with the highest score for its respective range:

***In Vitro* Opacity or Permeability Ranges**	***In Vitro* Classification**
Opacity < 0.400 or Permeability < 0.175	Mild irritant
0.400 ≤ Opacity < 1.300 or 0.175 ≤ Permeability < 0.600	Moderate irritant
Opacity > 1.300 or Permeability ≥ 0.600	Severe irritant

The rationale for the cutoffs used in this classification scheme was not provided and the correlation of these categories to accepted regulatory categories is unknown.

As described above, the surfactant-based personal care formulations evaluated by Gettings et al. (1996) do not produce appreciable opacity in the isolated bovine cornea, but can increase permeability by damaging the epithelium. Thus, it was recommended that a severe irritant be defined as a substance that produces a permeability value ≥ 0.600 (Harbell J, personal communication). Also, some companies, such as S.C. Johnson & Son, Inc., note that they do not use any of the classification systems described above to assign an ocular irritancy classification, but instead compare BCOP data for newly tested substances to benchmark materials, relying on a system of comparative toxicity instead of cutoff scores (Cuellar N and Swanson J, personal communication).

5.4 Summary of Results

BCOP data were collected for a total of 161 test substances among the eight studies evaluated. A summary of results used to evaluate test method accuracy is shown in **Appendix D**. **Appendix D1** provides a table, sorted first by reference then alphabetically by substance, with the name of the substance tested, the CASRN, 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. **Appendix D2** provides the same information, but is sorted alphabetically by test substance to indicate which substances were tested in multiple studies. Other supporting information, such as the source, purity and physicochemical characteristics of the test substances, was included in the tables to the extent this information was available. No attempt was made to identify the source, purity, and physicochemical characteristics of a test substance, if the authors did not provide such information. If not provided, the CASRN was obtained from various sources, including the National Library of Medicine's ChemID database. Chemical and product classes were assigned based on the MeSH classification system (available at http//www.nlm.nih.gov/mesh). Each of the eight studies evaluated varied with respect to the level of detail of data that was provided, as described below.

5.4.1 Gautheron et al. (1994)

In this interlaboratory evaluation of the BCOP test method, BCOP data were extracted for 52 test substances, which were evaluated in 11 or 12 laboratories. Four of these laboratories (numbers 2, 3, 8, and 10) used a modified protocol, in which preserved corneas were used in place of freshly collected corneas. Laboratory 10 completed studies on only 23 of the test substances. NICEATM classified each test substance based on the *in vitro* classification system described by the authors of the study.

The *in vivo* reference data were provided by Dr. Philippe Vanparys, allowing for an accuracy analysis of up to 50 substances in relation to the *in vivo* ocular irritancy classifications assigned by NICEATM for the substances according to the EPA (EPA 1996), EU (EU 2001), and GHS (UN 2003) classification systems. Not all of the 52 substances tested could be evaluated because some ($n \leq 3$) of the *in vivo* studies did not provide sufficient data to assign an ocular irritancy classification for each classification scheme. However, because the 52 test substances were tested *in vitro* using a standardized protocol in eight laboratories, an interlaboratory reliability analysis also could be conducted for this study.

5.4.2 Balls et al. (1995)

In this evaluation of the BCOP test method, 51 chemicals were evaluated in five laboratories. Four of these chemicals were tested at two different concentrations and two were tested at three concentrations, for a total of 59 different test substances. BCOP test method data on the 59 tests were not included in the published report. Rather, the study report included scatter plots showing the relationship between the BCOP data (i.e., mean opacity value, mean permeability value, and mean *in vitro* score) obtained in the lead laboratory and the MMAS for the entire set of test substances. However, the mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance in each laboratory were provided by ECVAM for all 59 test substances. Detailed *in vivo* data are available for the 59 test substances (including the different concentrations tested) in ECETOC (1998), allowing for an accuracy analysis of the 59 substances in relation to the *in vivo* ocular irritancy classifications assigned by NICEATM for the substances according to the EPA (1996), GHS (UN 2003) and EU (2001) classification systems. Because the 59 test substances were tested using a standardized protocol in five laboratories, an interlaboratory reliability analysis could be conducted for this study. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by the authors of the study to classify each test substance.

5.4.3 Swanson et al. (1995)

In this study of 20 full-strength cleaners and floor strippers, *in vitro* data were extracted for 13 formulations with sufficient *in vivo* reference data to allow for an accuracy analysis. The mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance (in one laboratory) were provided by S.C. Johnson & Son, Inc. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by Gautheron et al. (1994) to classify each test substance.

5.4.4 Gettings et al. (1996)

In the CTFA Evaluation of Alternatives Program – Phase III, 25 surfactant-based personal care cleansing formulations were evaluated in one laboratory. The mean permeability (OD_{490}) and mean *in vitro* score were provided for each formulation in the study report. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM assigned a classification to each test substance based on the mean permeability value obtained for each substance. A substance that produced a permeability value ≥ 0.600 was classified as a severe ocular irritant.

5.4.5 Casterton et al. (1996)

For this study, *in vitro* data were extracted for 15 personal care product formulations, 13 household cleaning product formulations, and 32 chemicals with available *in vivo* reference data. The mean opacity value and the mean permeability value were provided in the study report, as well as the laboratory specific *in vitro* classification for each test substance. These data were obtained from one laboratory.

Although BCOP data were reported in this publication for an additional 37 chemicals and consumer product formulations, detailed *in vivo* reference data were not available for these

substances, precluding an accuracy analysis for this set of substances. Therefore, the BCOP data for these 37 substances are not included in this document.

5.4.6 Southee (1998)

In this study, 16 test substances were evaluated in three laboratories. The mean opacity value, mean permeability value (OD_{490}), number of replicates, mean *in vitro* score, SD for all mean values, and *in vitro* classification were provided for each test substance. Each laboratory tested each substance on at least two separate occasions. Imidazole, ethanol, and benzalkonium chloride were each tested in at least seven different experiments by each laboratory.

5.4.7 Swanson and Harbell (2000)

In this study of 13 ethanol containing insect repellent formulations, *in vitro* data were extracted for ethanol and eight formulations with sufficient *in vivo* reference data to allow for an accuracy analysis. The mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance (in one laboratory) were provided by S.C. Johnson & Son, Inc. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by Gautheron et al. (1994) to classify each test substance.

5.4.8 Bailey et al. (2004)

In this study of 16 petrochemical products, *in vitro* data were extracted for all of the test substances, which had sufficient *in vivo* reference data to allow for an accuracy analysis. The mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance (in one laboratory) were provided by ExxonMobil Biomedical Sciences, Inc. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by Gautheron et al. (1994) to classify each test substance.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines and with the use of coded chemicals (OECD 1998; EPA 2003a, 2003b; FDA 2003). The data quality was evaluated by a review of the methods section in literature references and the submitted reports. The data quality presented in the reviewed literature references can be evaluated to the extent this information was provided in the published reports. Based on the available information, the reports that identified following GLP guidelines or used data obtained according to GLP guidelines were Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Swanson and Harbell (2000), and Bailey et al. (2004). Likewise, the reports that identified using coded chemicals were Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004).

5.6 Lot-to-lot Consistency of Test Substances

Ideally, the lot-to-lot consistency of test substances is evaluated to ensure that the same substance, with the same physicochemical properties, is being evaluated over the duration of the study. A description of the procedures used to evaluate the lot-to-lot consistency was provided in the published reports. No attempt was made to review original records to assess the procedures used to evaluate different batches of tested substances.

Gettings et al. (1996) noted that all samples were dispensed from a single source to ensure test substance consistency. The samples were placed into a secondary container, labeled with appropriate chemical code information and then forwarded to the participating testing laboratories. There is no information about the time frame in which the studies were conducted or whether additional aliquots of the samples were forwarded to specific testing laboratories.

Balls et al. (1995) noted that substances with the same source and specification as those tested *in vivo* were obtained, whenever possible, to test *in vitro*. When such a situation was not possible, a specification as close as possible to what was evaluated *in vivo* was selected. Aliquots of each test substance were prepared at one time and forwarded to the participating testing laboratories. There is no information about the time frame in which the studies were conducted or whether additional aliquots of the samples were forwarded to specific testing laboratories.

There was no information about maintaining lot-to-lot consistency in any of the other reports reviewed.

5.7 Availability of Data for External Audit

Study notebooks and other supporting records are known to be available, upon request, for an external audit, for the following studies: Swanson et al. (1995), Gettings et al. (1996), Swanson and Harbell (2000), and Bailey et al. (2004). The availability of data for an external audit for the other reports described in this section has not been determined.

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 correctly identify ocular corrosives and severe irritants, as defined by the EPA (1996), the EU (2001), and the GHS (UN 2003)[1], was evaluated using two approaches. In the first approach, the performance of the BCOP assay was assessed separately for each *in vitro-in vivo* comparative study (i.e., publication or data submission) reviewed in **Sections 4.0** and **5.0**. In the second approach, the performance of the BCOP was assessed after pooling data across comparative studies that used the same method of data collection and analysis. The three ocular hazard classification systems considered during this analysis use different classification schemes and decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results (see **Section 1.0**). All three regulatory classification systems are based on individual animal data in terms of the magnitude of the response and, for the EPA and GHS, the amount of time it takes for the ocular lesions to clear. Thus, to evaluate the accuracy of the BCOP test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at the different observation times were needed for each substance. However, these data were not consistently available in the studies considered, which limited the number of test results that could be used to assess test method accuracy. Furthermore, most of the *in vivo* classifications used for the analyses presented in this section are based on the results of a single study. Unless otherwise indicated, variability in the *in vivo* classification is unknown.

[1] For the purposes of this analysis, an ocular corrosive or severe irritant was defined as a substance that would be classified as Category 1 according to the GHS classification system (UN 2003), as Category I according to the EPA classification system (EPA 1996), or as R41 according to the EU classification system (EU 2001) (see **Section 1.0**).

In addition, the accuracy assessments conducted were based on BCOP data that were evaluated differently. As discussed in **Section 2.2.12**, a majority of BCOP studies used the mean opacity and mean permeability values (OD_{490}) for each treatment group to calculate an *In Vitro* Irritancy Score for each test substance. However, Casterton et al. (1996) assigned irritation classes based on the endpoint (opacity or permeability) with the highest score for its respective range. Conversion of the BCOP data in Casterton et al. (1996) to an *In Vitro* Irritancy Score was not attempted since opacity was measured with a UV/VIS spectrophotometer instead of an opacitometer; the author's classifications were used for this analysis. Gettings et al. (1996) used the *In Vitro* Irritancy Score and permeability score alone to classify the 25 surfactant-based formulations evaluated in the CTFA Phase III study, and it was found that the permeability score alone better predicted the *in vivo* ocular classification according to the FHSA classification system. Thus, for this accuracy analysis, only permeability scores are considered for Gettings et al. (1996).

Accuracy of BCOP for Individual Studies: For the "per study" accuracy analysis, two different types of analyses were performed. In the first analysis, the BCOP ocular irritancy potential of each test substance in each study was determined (**Appendix C**). For the three studies where the same test substance was evaluated in multiple laboratories within the same study (i.e., Gautheron et al. 1994; Balls et al. 1995; Southee 1998), the BCOP ocular irritancy classification for each independent test result was determined. Subsequently, an overall BCOP ocular irritancy classification was assigned for each chemical in the study based on the majority of ocular irritancy classification calls (e.g., if two tests classified a substance as a moderate irritant and three tests classified a substance as a severe irritant, the overall *in vitro* irritancy classification for the substance would be severe irritant). When there was an even number of different irritancy classifications for test substances (e.g., two tests classified a substance as a moderate irritant and two tests classified a substance as a severe irritant), the more severe irritancy classification was used for the overall classification for the substance (severe irritant, in this case). Once the ocular irritancy potential classification was determined for each substance in each of the studies, the ability of the BCOP test method to identify ocular corrosives and severe irritants, based on the three different classification systems, was determined for each study (**Appendix D**).

The second analysis conducted in the "per study" evaluation used each independent test result for each substance that had been tested in multiple laboratories (Gautheron et al. 1994, Balls et al. 1995, and Southee 1998). Each *in vitro* classification obtained when a test substance was evaluated in multiple laboratories was used separately to assess test method accuracy (i.e., results were not combined across multiple testing laboratories to develop an overall BCOP ocular irritancy classification). The ability of the BCOP test method to identify ocular corrosives and severe irritants, based on the three different regulatory classification systems, was then determined.

Accuracy of BCOP for Pooled Studies: For an overall analysis of accuracy for BCOP, results from the six different comparative studies that used the same BCOP analysis approach (i.e., calculation of an *In Vitro* Irritancy Score = opacity + (15 x OD_{490}) or use of permeability value only for substances that produce permeability without appreciable opacity) were combined and an overall ocular classification was determined for each

substance. When the same test substance was evaluated in multiple studies, the overall BCOP ocular irritancy potential was based on the majority of calls among all of the studies (see **Appendix C**). Once the overall *in vitro* ocular irritancy classification was determined for each test substance, the classification was compared to the *in vivo* ocular irritancy classification (**Appendix D**).

6.1.1 GHS Classification System: BCOP Test Method Accuracy

Accuracy analyses for ocular corrosives and severe irritants, as defined by the GHS classification system[2] (UN 2003), were performed for the following eight studies: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Casterton et al. (1996), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004). The GHS classification assigned to each test substance is presented in **Appendix D**. The performance characteristics (i.e., accuracy, sensitivity, specificity, positive predictivity, negative predictivity, false positive rate, and false negative rate) were determined for each of the eight studies based on the available *in vivo* reference data for the substances tested in these studies (**Table 6-1**). Of the eight studies, Gautheron et al. (1994), Balls et al. (1995), and Southee (1998) provided BCOP data for substances tested in multiple laboratories; the first set of accuracy calculations for these studies in **Table 6-1** represents the results obtained using the consensus call for each test substance, while the second set of accuracy calculations for each study represents the results obtained when each independent test result from each laboratory was considered separately.

Based on the data provided in the eight studies, when a single call was used per test substance per study, the BCOP test method has an accuracy of 67% to 100%, a sensitivity of 48% to 100%, a specificity of 66% to 100%, a false positive rate of 0% to 34%, and a false negative rate of 0% to 52% (**Table 6-1)**.

Using the first accuracy analysis approach (single call per test substance), the three BCOP studies that evaluated test substances in multiple laboratories (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) have an accuracy of 70% to 74%, a sensitivity of 57% to 77%, a specificity of 66% to 88%, a false positive rate of 12% to 34%, and a false negative rate of 23% to 43%. In contrast, when BCOP study results from multiple laboratories are considered separately rather than being combined to provide an overall classification for each substance, the BCOP test method has an accuracy of 70% to 79%, a sensitivity of 69% to 77%, a specificity of 66% to 83%, a false positive rate of 17% to 34%, and a false negative rate of 24% to 31%. These performance characteristics are provided in **Table 6-1**. The values obtained for the second analysis approach changed little in comparison to the first accuracy analysis approach for the Balls et al. (1995) study, but changed more substantially for the Gautheron et al. (1994) and the Southee (1998) studies.

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

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

Data Source	Anal.[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gautheron et al. 1994	IVIS	47/52	74[4]	35/47	71	5/7	75	30/40	33	5/15	94	30/32	25	10/40	29	2/7
			77[e]	432/558	69	62/90	79	370/468	39	62/160	93	370/398	21	98/468	31	28/90
Balls et al. 1995[6]	IVIS	54/59	70[4]	38/54	77	17/22	66	21/32	61	17/28	81	21/26	34	11/32	23	5/22
			70[5]	190/270	77	85/110	66	105/160	61	85/140	81	105/130	34	55/160	24	26/110
Swanson et al. 1995	IVIS	8/20	100	8/8	100	6/6	100	2/2	100	6/6	100	2/2	0	0/2	0	0/6
Gettings et al. 1996	Perm	23/25	87	20/23	75	6/8	93	14/15	86	6/7	88	14/16	7	1/15	25	2/8
Casterton et al. 1996	O/P	55/97	67	37/55	48	13/27	86	24/28	76	13/17	63	24/38	14	4/28	52	14/27
Southee 1998	IVIS	15/16	73[4]	11/15	57	4/7	88	7/8	80	4/5	70	7/10	12	1/8	43	3/7
			79[5]	110/139	76	57/75	83	53/64	84	57/68	75	53/71	17	11/64	24	18/75
Swanson & Harbell 2000	IVIS	9/13	78	7/9	100	1/1	75	6/8	33	1/3	100	6/6	25	2/8	0	0/1
Bailey et al. 2004	IVIS	14/16	93	13/14	67	2/3	100	11/11	100	2/2	92	11/12	0	0/11	33	1/3
Pooled Studies[7]		147/203	81	119/147	84	36/43	80	83/104	63	36/57	92	83/90	20	21/104	16	7/43

[1]Anal. = Analytical method used to transform the sample data into BCOP classification; IVIS = *In Vitro* Irritancy Score developed by Gautheron et al. (1994); Perm = Permeability value only used to classify *in vitro* ocular irritancy in the BCOP assay (an OD_{490} value > 0.600 was considered a severe irritant); O/P = Irritation class based on the endpoint (opacity or permeability) with the highest score for its respective range (Casterton et al. 1996).

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

[3]The data on which the percentage calculation is based.

[4]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the multiple testing laboratories and tests (for substances tested multiple times in a laboratory).

[5]Performance calculated using each individual *in vitro* classification from each testing laboratory and test.

[6]The test substance 1% benzalkonium chloride was tested in two different *in vivo* studies, producing discordant results with respect to GHS classification (study 1 = Category 2B and study 2 = Category 1). The analysis was performed using the Category 1 classification.

[7]Data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004) were pooled together and an overall *in vitro* classification was assigned for each test substance based on the majority and/or most severe classification obtained across tests and testing laboratories. Data from Casterton et al. (1996) were not included in this analysis since the protocol used to generate BCOP data differed considerably from the other studies (e.g., A spectrophotometer was used to measure opacity instead of an opacitometer, and solids were applied neat instead of as a 20% solution or suspension).

In terms of an overall accuracy analysis, using data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000) and Bailey et al. (2004), the BCOP test method has an accuracy of 81%, a sensitivity of 84%, a specificity of 80%, a false positive rate of 20%, and a false negative rate of 16%. The performance characteristics for the pooled studies are provided in **Table 6-1**.

As described in **Sections 3.0** and **4.0**, appropriate *in vivo* data were not available for all of the substances evaluated in some of the studies. For example, in the Swanson et al. (1995) study, only eight of the 20 substances had appropriate *in vivo* data to assign a GHS classification.

6.1.1.1 *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 with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., pesticides, surfactants, pH, physical form).

As indicated in **Table 6-2**, there were some notable trends in the performance of the BCOP test method among these subgroups of substances. The chemical class of substances that was most consistently overpredicted according the GHS classification system (i.e., were false positives) by the BCOP test method is alcohols. Eight out the 21 overpredicted substances were alcohols. Additional chemical classes represented among the overpredicted substances were ketones (4), carboxylic acids (3), heterocyclic compounds (2), esters (1), and hydrocarbons (1). Among the 35 substances labeled as surfactants only 5% (1/21, a surfactant-containing formulation) were overpredicted by the BCOP test method.

With regard to physical form of the substances overpredicted by the BCOP test method, 18 were liquids and two were solids. Considering the proportion of the total available database, liquids (92/124; 74%) appear more likely than solids (32/124; 26%) to be overpredicted by the BCOP test method.

Although there were a relatively small number (7) of substances represented, alcohols (2) were most often underpredicted (i.e., were false negatives[3]) by the BCOP test method according to the GHS classification system (see **Appendix D**). As can be seen in **Table 6-2**, the 35 substances labeled as surfactants were rarely underpredicted by the BCOP test method (7% [1/14] false negative rate).

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 (32/124; 26%) appear more likely than liquids (92/124; 74%) to be underpredicted by the BCOP test method.

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

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

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	147	20	21/104	16	7/43
Chemical Class[6]					
Alcohol	18	53	8/15	67	2/3
Amine/Amidine	8	0	0/4	0	0/4
Carboxylic acid	15	38	3/8	14	1/7
Ester	12	12	1/8	0	0/4
Ether/Polyether	6	0	0/5	0	0/1
Heterocycle	12	33	2/6	17	1/6
Hydrocarbon	12	8	1/12	-	0/0
Inorganic salt	5	0	0/3	0	0/2
Ketone	10	40	4/10	-	0/0
Onium compound	11	0	0/3	0	0/8
Properties of Interest					
Liquids[7]	92	26	18/68	4	1/24
Solids[7]	32	10	2/20	42	5/12
Pesticide	8	33	1/3	40	2/5
Surfactant – Total[8]	35	5	1/21	7	1/14
-nonionic	5	0	0/4	0	0/1
-anionic	3	0	0/2	100	1/1
-cationic	6	0	0/1	0	0/5
pH – Total[9]	28	-	-	21	5/24
- acidic (pH < 7.0)	11	-	-	18	2/11
- basic (pH > 7.0)	15	-	-	23	3/13
- equals 7	2	-	-	-	-
Category 1 Subgroup[10] - Total	38[11]	-	-	18	7/38
- 4 (CO=4 at any time)	20	-	-	15	3/20
- 3 (severity/persistence)	1	-	-	0	0/1
- 2 (severity)	4	-	-	25	1/4
- 2-4 combined[12]	25	-	-	16	4/25
- 1 (persistence)	13	-	-	23	3/13

[1]GHS = Globally Harmonized System (UN 2003).
[2]N = Number of substances.
[3]False Positive Rate = The proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = The proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Physical form (i.e., solid or liquid) not known for some substances, and therefore the overall number does not equal the sum of the solid and liquid substances.
[8]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.
[9]Total number of GHS Category 1 substances for which pH information was obtained.
[10]NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including Corneal Opacity [CO]=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO = 4 at any time.
[11]The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated, since some substances could not be classified into the subgroups used in the evaluation.
[12]Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

There was no definitive difference among the underpredicted substances for which pH information was available, as two were acidic (pH < 7.0) and three were basic (pH > 7.0), and considering the comparable proportion of acidic and basic underpredicted substances

(2/11; 18% vs. 3/13; 23%). Finally, the seven underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions (3/13; 23%) rather than on severe lesions (4/25; 16%), as evidenced by an analysis of NICEATM-defined GHS Category 1 sub-groupings (**Table 6-2**).

Table 6.3 shows the effects on the BCOP test method performance statistics of excluding from the data set problematic classes (i.e., that gave the most discordant results, according to the GHS classification system). In general, exclusion of alcohols, ketones or solids individually resulted in small changes in the performance statistics, with the exception that the exclusion of solids from the data set caused a four-fold decrease in the false negative rate from 16% (7/43) to 4% (1/29). When both alcohols and ketones were excluded from the data set, changes in the performance statistics were noted, with accuracy increasing from 81% (119/147) to 88% (103/117), and the false positive rate decreasing from 20% (21/104) to 12% (9/77). The largest changes were observed when all three discordant classes were excluded from the data set; accuracy increased from 81% (119/147) to 92% (78/85), the false positive rate decreased from 20% (21/104) to 12% (7/58), and the false negative rate decreased from 16% (7/43) to 0% (0/27).

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

Data Set	*Accuracy*		*False Positive Rate[2]*		*False Negative Rate[3]*	
	%	*No.[4]*	*%*	*No.*	*%*	*No.*
Overall	*81*	*119/147*	*20*	*21/104*	*16*	*7/43*
w/o Alcohols	*86*	*109/126*	*14*	*12/86*	*13*	*5/40*
w/o Ketones	*81*	*113/138*	*19*	*18/95*	*16*	*7/43*
w/o Solids	*82*	*93/113*	*23*	*19/84*	*4*	*1/29*
w/o Alcohols & Ketones	*88*	*103/117*	*12*	*9/77*	*13*	*5/40*
w/o Alcohols & Ketones & Solids	*92*	*78/85*	*12*	*7/58*	*0*	*0/27*

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

6.1.2 EPA Classification System: BCOP Test Method Accuracy

Accuracy analyses for ocular corrosives and severe irritants, as defined by the EPA classification system[4] (EPA 1996), were performed for the following eight studies: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Casterton et al. (1996), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004). The EPA classification assigned to each test substance is presented in **Appendix D.** The performance characteristics of the eight studies are shown in **Table 6-4** and are based on the available *in vivo* reference data for each study. Of the eight studies, Gautheron et al. (1994), Balls et al. (1995), and Southee (1998) provided BCOP data for substances tested in multiple laboratories; the first set of accuracy calculations for these studies in **Table 6-4** represents the results obtained using the consensus call for each test substance, while the second set of accuracy calculations for each study represents the results obtained when each independent test result from each laboratory was considered separately.

Based on the data provided in these eight studies, when a single call was used per test substance per study, the BCOP test method has an accuracy of 62% to 92%, a sensitivity of 40% to 100%, a specificity of 50% to 100%, a false positive rate of 0% to 50%, and a false negative rate of 0% to 100% (**Table 6-4)**.

Using the first accuracy analysis approach (single call per test substance), the three BCOP studies that evaluated test substances in multiple laboratories (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) have an accuracy of 64% to 73%, a sensitivity of 40% to 72%, a specificity of 63% to 78%, a false positive rate of 22% to 37%, and a false negative rate of 28% to 60%. In contrast, when BCOP study results from multiple laboratories are considered separately rather than being combined to provide an overall classification for each substance, the BCOP test method has an accuracy of 66% to 75%, a sensitivity of 60% to 72%, a specificity of 63% to 76%, a false positive rate of 24% to 37%, and a false negative rate of 28% to 40% (**Table 6-4)**. The values obtained for the second analysis approach changed little in comparison to the first accuracy analysis approach for the Balls et al. (1995) study, but changed more substantially for the Gautheron et al. (1994) and the Southee (1998) studies.

In terms of an overall accuracy analysis, using data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000) and Bailey et al. (2004), the BCOP test method has an accuracy of 79%, a sensitivity of 75%, a specificity of 81%, a false positive rate of 19%, and a false negative rate of 25%. The performance characteristics for the pooled studies are provided in **Table 6-4**.

As described in **Section 4.0**, *in vivo* data were not available for all of the substances evaluated in some of the studies. For example, for the Swanson et al. (1995) study, only eight of the 20 substances had sufficient *in vivo* data to assign an EPA classification.

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

6.1.2.1 *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).

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

Data Source	Anal.[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gautheron et al. 1994[4]	IVIS	48/52	73[5]	35/48	71	5/7	73	30/41	31	5/16	94	30/32	27	11/41	29	2/7
			75[6]	427/571	65	51/79	76	376/492	30	51/167	93	376/404	24	116/492	35	28/79
Balls et al. 1995[4]	IVIS	53/59	66[5]	35/53	72	13/18	63	22/35	50	13/26	82	22/27	37	13/35	28	5/18
			66[6]	175/265	72	65/90	63	110/175	50	65/130	82	110/135	37	65/175	28	25/90
Swanson et al. 1995	IVIS	8/20	88	7/8	100	6/6	50	1/2	86	6/7	100	1/1	50	1/2	0	0/6
Gettings et al. 1996	Perm	25/25	80	20/25	60	6/10	93	14/15	86	6/7	78	14/18	7	1/15	40	4/10
Casterton et al. 1996	O/P	56/97	62	35/56	41	11/27	83	24/29	69	11/16	60	24/40	17	5/29	59	16/27
Southee 1998[4]	IVIS	14/16	64[5]	9/14	40	2/5	78	7/9	50	2/4	70	7/10	22	2/9	60	3/5
			70[6]	80/115	60	27/45	76	53/70	61	27/44	75	53/71	24	17/70	40	18/45
Swanson & Harbell 2000[4]	IVIS	9/13	89	8/9	75	3/4	100	5/5	100	3/3	83	5/6	0	0/5	25	1/4
Bailey et al. 2004	IVIS	13/16	92	12/13	0	0/1	100	12/12	-	0/0	92	12/13	0	0/12	100	1/1
Pooled Studies[7]		143/203	79	113/143	75	30/40	81	83/103	60	30/50	89	83/93	19	20/103	25	10/40

[1]Anal. = Analytical method used to transform the sample data into BCOP classification; IVIS = *In Vitro* Irritancy Score developed by Gautheron et al. (1994); Perm = Permeability value only used to classify *in vitro* ocular irritancy in the BCOP assay (an OD_{490} value > 0.600 was considered a severe irritant); O/P = Irritation class based on the endpoint (opacity or permeability) with the highest score for its respective range (Casterton et al. 1996).

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

[3]The data on which the percentage calculation is based.

[4]The test substance ethanol was evaluated in two different *in vivo* studies (ECETOC 1998; Swanson and Harbell 2000), producing discordant results with respect to EPA classification (study 1 = Category III and study 2 = Category I). The analysis was performed using the Category I classification.

[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the multiple testing laboratories and tests (for substances tested multiple times in a laboratory).

[6]Performance calculated using each individual *in vitro* classification from each testing laboratory and test.

[7]Data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004) were pooled together and an overall *in vitro* classification was assigned for each test substance based on the majority and/or most severe classification obtained across tests and testing laboratories. Data from Casterton et al. (1996) were not included in this analysis, since the protocol used to generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer, and solids were applied neat instead of as a 20% solution or suspension).

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 class of substances that was most consistently overpredicted according the EPA classification system (i.e., were false positives) by the BCOP test method is alcohols. Nine out the 20 overpredicted substances were alcohols. Additional chemical classes represented among the overpredicted substances were ketones (4), carboxylic acids (3), heterocyclic compounds (2), esters (2), hydrocarbons (1), inorganic salts (1), and onium compounds (1). Among the 35 substances labeled as surfactants only 9% (2/22) were overpredicted by the BCOP test method (10% Triton X-100 and a surfactant-containing formulation).

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

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	143	19	20/103	25	10/40
Chemical Class[6]					
Alcohol	18	56	9/16	100	2/2
Amine/Amidine	8	0	0/6	0	0/2
Carboxylic acid	14	38	3/8	17	1/6
Ester	9	22	2/9	-	0/0
Ether/Polyether	6	0	0/5	100	1/1
Heterocycle	11	33	2/6	20	1/5
Hydrocarbon	12	8	1/12	-	0/0
Inorganic salt	5	25	1/4	0	0/1
Ketone	10	40	4/10	-	0/0
Onium compound	9	25	1/4	0	0/5
Properties of Interest					
Liquids[7]	90	29	18/70	5	1/20
Solids[7]	31	10	2/21	50	5/10
Pesticide	9	25	1/4	40	2/5
Surfactant – Total[8]	35	9	2/22	23	3/13
-nonionic	5	20	1/5	-	0/0
-anionic	3	0	0/2	0	0/1
-cationic	4	0	0/1	0	0/3
pH – Total[9]	25	-	-	32	6/19
- acidic (pH < 7.0)	9	-	-	30	3/10
- basic (pH > 7.0)	14	-	-	33	3/9
- equals 7	2	-	-	-	-

[1]EPA = U.S. Environmental Protection Agency (EPA 1996).
[2]N = Number of substances.
[3]False Positive Rate = The proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = The proportion of all positive substances that are falsely identified as negative *in vitro.*
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Physical form (i.e., solid or liquid) not known for some substances, and therefore the overall number does not equal the sum of the solid and liquid substances.
[8]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.
[9]Total number of EPA Category I substances for which pH information was obtained.

With regard to physical form of the substances overpredicted by the BCOP test method, 18 were liquids and two were solids. Considering the proportion of the total available database, liquids (90/121; 74%) appear more likely than solids (31/121; 26%) to be overpredicted by the BCOP test method

Although there were a relatively small number (10) of substances represented, alcohols (2) were most often underpredicted (i.e., were false negatives) by the BCOP test method according to the EPA classification system (see **Appendix D**). As can be seen in **Table 6-5**, some of the 35 substances labeled as surfactants were underpredicted by the BCOP test method (23% [3/13] false negative rate).

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 (31/121; 26%) appear more likely than liquids (90/121; 74%) to be underpredicted by the BCOP test method.

There was no definitive difference among the underpredicted substances for which pH information was available, as three were acidic (pH < 7.0) and three were basic (pH > 7.0), and considering the comparable proportion of acidic and basic underpredicted substances (3/10; 30% vs. 3/9; 33%).

6.1.3 EU Classification System: BCOP Test Method Accuracy

Accuracy analyses for ocular corrosives and severe irritants, as defined by the EU (2001) classification system[5], were performed for the following eight studies: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Casterton et al. (1996), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000) and Bailey et al. (2004). Of these reports, Gautheron et al. (1994), Balls et al. (1995), and Southee (1998) provided BCOP data for substances tested in multiple laboratories. The EU classification assigned to each test substance is presented in **Appendix D.**

Based on the data provided in these eight studies, when a single call was used per test substance per study, the BCOP test method has an accuracy of 68% to 92%, a sensitivity of 52% to 100%, a specificity of 64% to 100%, a false positive rate of 0% to 36%, and a false negative rate of 0% to 48% (**Table 6-6)**.

Using the first accuracy analysis approach (single call per test substance), the three BCOP studies that evaluated test substances in multiple laboratories (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) have an accuracy of 68% to 79%, a sensitivity of 67% to 74%, a specificity of 64% to 88%, a false positive rate of 12% to 36%, and a false negative rate of 26% to 33%. In contrast, when BCOP study results from multiple laboratories are considered separately rather than being combined to provide an overall classification for each substance, the BCOP test method has an accuracy of 69% to 83%, a sensitivity of 69% to 83%, a specificity of 65% to 83%, a false positive rate of 17% to 35%, and a false negative rate of 17% to 31% (**Table 6-6)**.

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

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

Data Source	Anal.[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gautheron et al. 1994[4]	IVIS	48/52	73[5]	35/48	71	5/7	73	30/41	31	5/16	94	30/32	27	11/41	29	2/7
			77[6]	437/570	69	62/90	78	375/480	37	62/167	93	375/403	22	105/480	31	28/90
Balls et al. 1995	IVIS	50/59	68[5]	34/50	74	14/19	64	20/31	56	14/25	80	20/25	36	11/31	26	5/19
			69[6]	171/248	75	71/95	65	100/153	57	71/124	81	100/124	35	53/153	25	24/95
Swanson et al. 1995	IVIS	9/20	89	8/9	100	6/6	67	2/3	86	6/7	100	2/2	33	1/3	0	0/6
Gettings et al. 1996	Perm	23/25	87	20/23	75	6/8	93	14/15	86	6/7	88	14/16	7	1/15	25	2/8
Casterton et al. 1996	O/P	54/97	70	38/54	52	13/25	86	25/29	76	13/17	68	25/37	14	4/29	48	12/25
Southee 1998	IVIS	14/16	79[5]	11/14	67	4/6	88	7/8	80	4/5	78	7/9	12	1/8	33	2/6
			83[6]	110/133	83	57/69	83	53/64	84	57/68	82	53/65	17	11/64	17	12/69
Swanson & Harbell 2000	IVIS	9/13	78	7/9	100	1/1	75	6/8	33	1/3	100	6/6	25	2/8	0	0/1
Bailey et al. 2004	IVIS	13/16	92	12/13	67	2/3	100	10/10	100	2/2	91	10/11	0	0/10	33	1/3
Pooled Studies[7]		143/203	80	114/143	82	33/40	79	81/103	60	33/55	92	81/88	21	22/103	18	7/40

[1]Anal. = Analytical method used to transform the sample data into BCOP classification; IVIS = *In Vitro* Irritancy Score developed by Gautheron et al. (1994); Perm = Permeability value only used to classify *in vitro* ocular irritancy in the BCOP assay (an OD_{490} value > 0.600 was considered a severe irritant); O/P = Irritation class based on the endpoint (opacity or permeability) with the highest score for its respective range (Casterton et al. 1996).

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

[3]The data on which the percentage calculation is based.

[4]Accuracy analysis based on EEC (1984) classifications in Gautheron et al. (1994).

[5]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the multiple testing laboratories and tests (for substances tested multiple times in a laboratory).

[6]Performance calculated using each individual *in vitro* classification from each testing laboratory and test.

[7]Data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al (2004) were pooled together and an overall *in vitro* classification was assigned for each test substance based on the majority and/or most severe classification obtained across tests and testing laboratories. Data from Casterton et al. (1996) were not included in this analysis, since the protocol used to generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer, and solids were applied neat instead of as a 20% solution or suspension).

The values obtained for the second analysis approach changed slightly in comparison to the first accuracy analysis approach for the Balls et al. (1995) and the Gautheron et al. (1994) studies, but changed more substantially for the Southee (1998) study.

In terms of an overall accuracy analysis, using data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000) and Bailey et al. (2004), the BCOP test method has an accuracy of 80%, a sensitivity of 82%, a specificity of 79%, a false positive rate of 21%, and a false negative rate of 18%. The performance characteristics for the pooled studies are provided also in **Table 6-6**.

As described in **Section 4.0**, appropriate *in vivo* data were not available for all of the substances evaluated in some of the studies. For example, in Swanson et al. (1995), only nine of the 20 substances evaluated in this study had sufficient *in vivo* data to assign an EU classification.

6.1.3.1 *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., pesticides, surfactants, pH, physical form).

As indicated in **Table 6-7**, there were some notable trends in the performance of the BCOP test method among these subgroups of substances. The chemical class of substances that was most consistently overpredicted according the EU classification system (i.e., were false positives) by the BCOP test method is alcohols. Seven out the 22 overpredicted substances were alcohols. Additional chemical classes represented among the overpredicted substances were carboxylic acids (4), ketones (4), heterocyclic compounds (2), esters (1), and hydrocarbons (1). Among the 35 substances labeled as surfactants only 9% (2/22) were overpredicted by the BCOP test method (15% sodium lauryl sulfate and a surfactant-containing formulation).

With regard to physical form of the substances overpredicted by the BCOP test method, 19 were liquids and two were solids. Considering the proportion of the total available database, liquids (90/120; 75%) appear more likely than solids (30/120; 25%) to be overpredicted by the BCOP test method

Although there were a relatively small number (7) of substances represented, alcohols (2) were most often underpredicted (i.e., were false negatives) by the BCOP test method according to the EU classification system (see **Appendix D**). As can be seen in **Table 6-7**, the 35 substances labeled as surfactants were rarely underpredicted by the BCOP test method (8% [1/13] false negative rate).

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,

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

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	143	21	22/103	18	7/40
Chemical Class[6]					
Alcohol	17	50	7/14	67	2/3
Amine/Amidine	7	0	0/4	0	0/3
Carboxylic acid	14	44	4/9	20	1/5
Ester	12	12	1/8	0	0/4
Ether/Polyether	6	0	0/5	0	0/1
Heterocycle	12	33	2/6	17	1/6
Hydrocarbon	12	8	1/12	-	0/0
Inorganic salt	5	0	0/3	0	0/2
Ketone	10	40	4/10	-	0/0
Onium compound	11	0	0/3	0	0/8
Organic salt	7	0	0/3	0	0/4
Properties of Interest					
Liquids[7]	90	28	19/67	4	1/23
Solids[7]	30	10	2/20	50	5/10
Pesticide	7	33	1/3	50	2/4
Surfactant – Total[8]	35	9	2/22	8	1/13
-nonionic	5	0	0/4	0	0/1
-anionic	3	33	1/3	-	0/0
-cationic	6	0	0/1	0	0/5
pH – Total[9]	26	-	-	27	6/22
- acidic (pH < 7.0)	14	-	-	25	3/12
- basic (pH > 7.0)	10	-	-	30	3/10
- equals 7	2	-	-	-	-

[1]EU = European Union (EU 2001).
[2]N = Number of substances.
[3]False Positive Rate = The proportion of all negative substances that are falsely identified as positive *in vitro*.
[4]False Negative Rate = The proportion of all positive substances that are falsely identified as negative *in vitro*.
[5]Data used to calculate the percentage.
[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh)
[7]Physical form (i.e., solid or liquid) not known for some substances, and therefore the overall number does not equal the sum of the solid and liquid substances.
[8]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.
[9]Total number of EU Category R41 substances for which pH information was obtained.

solids (30/120; 25%) appear more likely than liquids (90/120; 75%) to be underpredicted by the BCOP test method.

There was no definitive difference among the underpredicted substances for which pH information was available, as three were acidic (pH < 7.0) and three were basic (pH > 7.0), and considering the comparable proportion of acidic and basic underpredicted substances (3/12; 25% vs. 3/10; 30%).

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

While there were some differences in results among the three hazard classification systems evaluated (i.e, EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that BCOP test method performance was comparable among the three systems. As can be seen in **Tables 6-1**, **6-4**, and **6-6**, the overall accuracy of the BCOP test method ranged from 79% to 81%, depending on the classification system used. Sensitivity and specificity ranged from 75% to 84% and 79% to 81%, respectively. The false positive rate ranged from 19% to 21%, while the false negative rate ranged from 16% to 25%. Given the relatively homogeneous performance of the BCOP test method among the three classification systems, the discussion below encompasses all three hazard classification systems, unless otherwise indicated.

6.2.1 Discordance Among Chemical Classes

The accuracy analysis indicated that alcohols are often overpredicted (50% to 56% [7/14 to 9/16] false positive rate, depending on the classification system used) in the BCOP test method. Ketones (40% [4/10]), carboxylic acids (38% to 44% [3/8 to 4/9]), and heterocyclic compounds (33% [2/6]) also had high false positive rates. The numbers of substances among the remaining chemical classes were too few to resolve any definitive trends in overprediction by the BCOP test method. For the purposes of these analyses, NICEATM considered five substances to be the threshold number per chemical class for consideration, and thus chemical classes represented by fewer than five substances were not considered.

Although there were a small number of underpredicted substances (4 to 5), alcohols (2) were most often underpredicted by the BCOP test method. The other chemical classes represented were carboxylic acids (1), ethers/polyethers (1), and heterocyclic compounds (1).

6.2.2 Discordance Among Physical or Chemical Properties of Interest

With regard to physical form of the substances overpredicted by the BCOP test method, 18 to 20 were liquids and two were solids. Considering the proportion of the total available database, liquids (90/120 to 92/124) appear more likely than solids (30/120 to 32/124) to be overpredicted by the BCOP test method.

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 indicated above, solids (42% to 50% false negative rate) appear more likely than liquids (4% to 5% false negative rate) to be underpredicted by the BCOP test method.

Exclusion of three discordant classes (i.e., alcohols, ketones and solids) from the data set resulted in an increased accuracy (from 81% to 92%), a decreased false positive rate (from 20% to 12%) and a decreased false negative rate (from 16% to 0%).

The 35 substances labeled as surfactants were rarely underpredicted by the BCOP test method for substances classified as severe by the EU (EU 2001) and GHS (UN 2003) classification systems (i.e., R41 or Category 1) as evidence by the false negative rates

ranging from 7% to 8%. Substances classified as severe (i.e., Category I) by the EPA classification system (EPA 1996) were more often underpredicted (false negative rate of 23%). However, although the available database was smaller (n = 7 to 9), substances labeled as pesticides were more often underpredicted by the BCOP test method (false negative rates ranging from 40% to 50%).

Considering the comparable proportion of acidic and basic underpredicted substances (18% to 30% [2/11 to 3/10] vs. 23% to 33% [3/13 to 3/9]), there was little difference among the underpredicted substances for which pH information was available. However, it is noted that pH information was available for only a portion of the 40 to 43 severe irritant substances (i.e., Category 1, Category I, or R41) in the database for each classification system.

Finally, with respect to the GHS classification system only, the seven underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions (false negative rate of 23% [3/13]), rather than on severe lesions (false negative rate of 17% [4/24]), as evidenced by an analysis of NICEATM-defined GHS Category 1 sub-groupings (**Table 6-2**).

[This Page Intentionally Left Blank]

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). Repeatability refers to the closeness of agreement between test results obtained within a single laboratory, when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM 1997, 2003). Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Interlaboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test chemicals, and indicates the extent to which a test method can be transferred successfully among laboratories. A reliability assessment includes reviewing the rationale for selecting the substances used to evaluate test method reliability, a discussion of the extent to which the substances tested represent the range of possible test outcomes and the properties of the various substances for which the test method is proposed for use, and a quantitative and/or qualitative analysis of repeatability and intra- and inter-laboratory reproducibility. In addition, measures of central tendency and variation are summarized for historical control data (negative, vehicle, positive), where applicable.

Quantitative BCOP test method data were available for replicate corneas within individual experiments or for replicate experiments within an individual laboratory for four studies (Gettings et al. 1996; Southee 1998; data submission from Dr. Joseph Sina; data submission from Dr. Freddy Van Goethem). Therefore, an evaluation of the repeatability and/or intralaboratory reproducibility of the BCOP test method could be conducted. Additionally, comparable BCOP data were available for multiple laboratories within each of three comparative validation studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998), which allowed for an evaluation of the interlaboratory reproducibility of the BCOP test method.

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

The quality of a reliability evaluation depends on the extent to which the substances tested adequately represent the range of physicochemical characteristics and response levels that the test method must be capable of evaluating.

The rationale for substance selection used in the various intralaboratory and multilaboratory studies was previously discussed in **Section 3.0**. In brief, substances were selected for inclusion based on available *in vivo* rabbit eye data for comparison, to cover the range of ocular irritation potential, and to include substances with different physicochemical properties (e.g., solids, liquids).

As noted previously, the EC/HO validation study reported by Balls et al. (1995) evaluated the performance and reproducibility of the BCOP test method using 60 “substances” (i.e., there were 52 different substances with four substances tested at two different concentrations and

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

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Assessment of Intralaboratory Repeatability and Reproducibility

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

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

Three of the studies discussed in **Section 6.0** included intralaboratory data (Gautheron et al. 1994, Gettings et al. 1996, and Southee 1998). For the Southee (1998) study, quantitative BCOP test method data were available for replicate corneas within individual experiments repeated two to five times for each test substance in three different laboratories. CV analyses were performed on within-experiment and between-experiment BCOP data, using the *In Vitro* Irritancy Score obtained for each test substance within each of the three testing laboratories. For the Gettings et al. (1996) study, Dr. John Harbell provided the mean permeability data obtained from three different experiments on the 25 surfactant-based formulations evaluated the CTFA Phase III study, as well as the mean permeability value for the three experiments, the standard deviation and the corresponding %CV values. In addition, Dr. Joseph Sina submitted a study of 43 substances, which included detailed BCOP data for replicate corneas. A CV analysis was conducted on the subset of substances provided by Dr. Sina that were tested using an incubation temperature of 32˚C, the temperature most commonly used in the BCOP for incubations as indicated in **Appendix A**; substances incubated at room temperature were not included in this analysis. For the Gautheron et al. (1994) study, Dr. Freddy Van Goethem provided individual cornea data collected in one of the participating laboratories (Janssen Pharmaceutica), which used six corneas per test substance. A %CV value was calculated for the opacity and permeability values and the *In Vitro* Irritancy Score for each test substance.

7.2.1.1 *Southee (1998)*

Intralaboratory Repeatability: In this study, 16 substances were evaluated in three laboratories multiple times (two to five experiments) for a total of 122 tests. Each test used three corneas. A %CV value was calculated for the opacity value, the permeability value, and the *In Vitro* Irritancy Score for each test (**Appendix E1**). **Tables 7-1**, **7-2**, and **7-3** summarize the mean and the %CV values of the *In Vitro* Irritancy Score for each test conducted in Laboratory 1, Laboratory 2, and Laboratory 3, respectively. The results for each laboratory are sorted by %CV values from lowest to highest value.

Table 7-1 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 1, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score (n = 3 corneas)	%CV	*In Vitro* Prediction
Benzalkonium chloride	138.0	0.1	Severe
NaOH (10%)	227.1	1.5	Severe
Benzalkonium chloride	137.9	1.6	Severe
Imidazole	142.0	2.1	Severe
Benzalkonium chloride	135.0	3.8	Severe
Imidazole	137.4	4.8	Severe
Imidazole	131.0	5.1	Severe
Benzalkonium chloride	195.0	5.8	Severe
4-Carboxybenzaldehyde	47.3	6.1	Moderate
Hexadecyltrimethylammonium bromide (10%)	20.0	6.3	Mild
4-Carboxybenzaldehyde	47.1	6.5	Moderate
Imidazole	145.7	8.3	Severe
Sodium lauryl sulfate (15%)	17.3	9.9	Mild
Glycerol	1.1	10.2	Mild
Butyl cellosolve	99.2	10.7	Severe
Methyl ethyl ketone	108.7	10.9	Severe
NaOH (10%)	245.0	11.7	Severe
Benzalkonium chloride	156.5	11.9	Severe
Ethanol	41.7	13.8	Moderate
Butyl cellosolve	92.8	14.0	Severe
Ethanol	31.5	14.2	Moderate
Ethanol	36.6	16.3	Moderate
Parafluoroaniline	38.3	19.6	Moderate
Methyl ethyl ketone	101.7	20.8	Severe
Ethanol	29.6	21.6	Moderate
Imidazole	112.0	22.0	Severe
Ammonium nitrate	5.9	23.4	Mild
Hexadecyltrimethylammonium bromide (10%)	23.1	25.3	Mild
Ethanol	37.6	28.6	Moderate
Triton X-100 (5%)	3.4	30.3	Mild
Parafluoroaniline	37.5	32.7	Moderate
Propyl-4-hydroxybenzoate	5.2	36.6	Mild
Triton X-100 (5%)	5.8	40.9	Mild
Propyl-4-hydroxybenzoate	3.6	44.1	Mild
Sodium lauryl sulfate (15%)	15.9	47.8	Mild
Ammonium nitrate	4.9	50.2	Mild
Glycerol	0.8	70.3	Mild
Tween 20	0.37	134.0	Mild
Tween 20	0.37	157.0	Mild
Sodium oxalate	-0.23	> 500	Mild
Sodium oxalate	-0.13	> 500	Mild
Mean %CV	**48.3**		
Median %CV	**14.2**		

[1]Substances organized by increasing %CV.

Table 7-2 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 2, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score (n = 3 corneas)	%CV	*In Vitro* Prediction
Benzalkonium chloride	157.9	2.1	Severe
NaOH (10%)	235.5	3.1	Severe
Benzalkonium chloride	150.8	4.7	Severe
Imidazole	137.6	4.9	Severe
Butyl cellosolve	111.8	4.9	Severe
NaOH (10%)	241.3	4.9	Severe
Sodium lauryl sulfate (15%)	5.4	5.9	Mild
Imidazole	134.9	7.0	Severe
4-Carboxybenzaldehyde	47.7	7.1	Moderate
Benzalkonium chloride	154.4	7.2	Severe
Imidazole	157.2	8.0	Severe
Ethanol	60.2	8.1	Severe
Propyl-4-hydroxybenzoate	7.4	8.3	Mild
Imidazole	140.1	8.5	Severe
Methyl ethyl ketone	67.8	8.5	Severe
4-Carboxybenzaldehyde	53.8	8.6	Moderate
Ethanol	54.2	9.1	Moderate
Imidazole	138.1	9.4	Severe
Benzalkonium chloride	157.2	11.5	Severe
Benzalkonium chloride	156.9	11.8	Severe
Butyl cellosolve	108.3	11.9	Severe
Ethanol	61.7	12.6	Severe
Sodium oxalate	10.3	13.5	Mild
Ethanol	54.5	15.1	Moderate
Parafluoroaniline	34.9	17.8	Moderate
Sodium oxalate	4.4	2.0	Mild
Methyl ethyl ketone	73.2	21.7	Severe
Parafluoroaniline	31.0	23.2	Moderate
Ethanol	52.7	24.3	Moderate
Ammonium nitrate	3.7	27.5	Mild
Triton X-100 (5%)	3.7	28.7	Mild
Propyl-4-hydroxybenzoate	11.2	28.7	Mild
Hexadecyltrimethylammonium bromide (10%)	34.7	35.0	Moderate
Hexadecyltrimethylammonium bromide (10%)	39.2	41.8	Moderate
Tween 20	0.3	45.8	Mild
Ammonium nitrate	3.9	46.4	Mild
Sodium lauryl sulfate (15%)	5.2	52.3	Mild
Triton X-100 (5%)	1.8	53.0	Mild
Glycerol	0.5	108.0	Mild
Glycerol	0.27	356.0	Mild
Tween 20	0.1	> 500	Mild
Mean %CV	**39.2**		
Median %CV	**11.8**		

[1]Substances organized by increasing %CV.

Table 7-3 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 3, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score (n = 3 corneas)	%CV	*In Vitro* Prediction
Ethanol	45.4	4.3	Moderate
Methyl ethyl ketone	70.3	5.1	Severe
Benzalkonium chloride	151.6	5.1	Severe
Imidazole	124.0	5.5	Severe
Benzalkonium chloride	169.7	6.0	Severe
Imidazole	128.7	6.3	Severe
Ethanol	44.4	6.7	Severe
Benzalkonium chloride	162.8	7.0	Severe
NaOH (10%)	214.8	7.2	Severe
Hexadecyltrimethylammonium bromide (10%)	31.7	7.3	Moderate
Ethanol	54.6	8.2	Moderate
Methyl ethyl ketone	73.5	8.7	Severe
4-Carboxybenzaldehyde	41.8	9.4	Moderate
NaOH (10%)	193.1	9.9	Severe
Benzalkonium chloride	163.4	9.9	Severe
4-Carboxybenzaldehyde	42.2	10.2	Moderate
Benzalkonium chloride	156.9	10.9	Severe
Propyl-4-hydroxybenzoate	6.2	11.8	Mild
Imidazole	123.4	12.0	Severe
Parafluoroaniline	22.1	12.0	Moderate
Ammonium nitrate	5.2	12.4	Mild
Parafluoroaniline	25.9	13.0	Moderate
Imidazole	140.2	13.5	Severe
Butyl cellosolve	94.9	14.5	Severe
Sodium lauryl sulfate (15%)	8.4	16.1	Mild
Ethanol	45.7	18.6	Moderate
Imidazole	139.6	18.6	Severe
Ammonium nitrate	6.7	21.6	Mild
Glycerol	0.8	21.7	Mild
Butyl cellosolve	98.2	22.0	Severe
Sodium lauryl sulfate (15%)	5.6	26.7	Mild
Sodium oxalate	4.6	28.5	Moderate
Ethanol	47.0	30.3	Severe
Sodium oxalate	2.7	33.0	Moderate
Triton X-100 (5%)	1.9	34.4	Mild
Hexadecyltrimethylammonium bromide (10%)	29.9	37.3	Moderate
Triton X-100 (5%)	3.0	37.9	Mild
Propyl-4-hydroxybenzoate	7.7	53.7	Mild
Glycerol	1.0	57.0	Mild
Tween 20	0.3	75.5	Mild
Tween 20	0.0	> 500	Mild
Mean %CV	**30.5**		
Median %CV	**12.4**		

[1]Substances organized by increasing %CV.

The ranges of %CV values for substances classified as severe irritants *in vitro* are 0.1 to 22.0 for Laboratory 1, 2.1 to 21.7 for Laboratory 2, and 5.1 to 30.3 for Laboratory 3. The within experiment mean and median %CV values for the three laboratories for all substances ranged from 30.5 to 48.3 and 11.8 to 14.2, respectively (%CV values listed as >500 were set at 500). Substances classified *in vitro* as mild irritants (i.e., *In Vitro* Irritancy Score >25) tended to have greater %CV values. The three laboratories all had at least one, but not more than two, %CV values greater than 500, which resulted from substances that had *In Vitro* irritancy Scores at or below the accepted background score of 3 to 5.

Intralaboratory Reproducibility: The between experiment %CV values of *In Vitro* Irritancy Scores for substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 are presented in **Tables 7-4**, **7-5**, and **7-6**, respectively. The mean %CV values ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

Table 7-4 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 1, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Exp.	%CV	*In Vitro* Prediction
Tween 20	0.37	2	0	Mild
4-Carboxybenzaldehyde	47.2	2	0.3	Moderate
Parafluoroaniline	37.9	2	1.6	Moderate
Butyl cellosolve	96	2	4.7	Severe
Methyl ethyl ketone	105	2	4.7	Severe
Ethanol	35.4	5	4.9	Moderate
NaOH (10%)	236	2	5.3	Severe
Sodium lauryl sulfate (15%)	16.6	2	6.1	Mild
Imidazole	133.7	5	9.9	Severe
Hexadecyltrimethylammonium bromide (10%)	21.6	2	10.1	Mild
Ammonium nitrate	5.4	2	13.45	Mild
Benzalkonium chloride	141.9	5	17.83	Severe
Glycerol	0.98	2	21.8	Mild
Propyl-4-hydroxybenzoate	4.4	2	25.7	Mild
Triton X-100 (5%)	4.6	2	36.7	Mild
Sodium oxalate	-0.07	2	39.3	Mild
Mean %CV	**12.6**			
Median %CV	**8.0**			

[1]Substances organized by increasing %CV.

Table 7-5 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 2, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Exp.	%CV	*In Vitro* Prediction
NaOH (10%)	238.4	2	1.7	Severe
Benzalkonium chloride	155	5	1.9	Severe
Butyl cellosolve	110	2	2.2	Severe
Sodium lauryl sulfate (15%)	5.3	2	3.1	Mild
Ammonium nitrate	3.8	2	4.3	Mild
Glycerol	0.52	2	4.5	Mild
Methyl ethyl ketone	70.5	2	5.5	Severe
Imidazole	141.6	5	6.3	Severe
Ethanol	56.7	5	7.1	Severe
4-Carboxybenzaldehyde	50.8	2	8.5	Moderate
Parafluoroaniline	32.9	2	8.5	Moderate
Hexadecyltrimethylammonium bromide (10%)	36.9	2	8.6	Moderate
Propyl-4-hydroxybenzoate	9.3	2	29.3	Mild
Tween 20	0.47	2	40.5	Mild
Triton X-100 (5%)	2.7	2	48	Mild
Sodium oxalate	7.4	2	56.4	Mild
Mean %CV			**14.8**	
Median %CV			**6.7**	

[1]Substances organized by increasing %CV.

Table 7-6 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 3, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Exp.	%CV	*In Vitro* Prediction
4-Carboxybenzaldehyde	42	2	0.57	Moderate
Butyl cellosolve	96.5	2	2.4	Severe
Methyl ethyl ketone	71.9	2	3.2	Severe
Benzalkonium chloride	161	5	4.2	Severe
Imidazole	131.2	5	6.3	Severe
NaOH (10%)	203.9	2	7.5	Severe
Ethanol	47.4	5	8.6	Moderate
Parafluoroaniline	24	2	11.3	Mild
Glycerol	0.88	2	13.4	Mild
Hexadecyltrimethylammonium bromide (10%)	33.3	2	14.2	Moderate
Propyl-4-hydroxybenzoate	7	2	15.2	Mild
Ammonium nitrate	5.9	2	18.6	Mild
Tween 20	0.4	2	23.7	Mild
Sodium lauryl sulfate (15%)	7	2	28.5	Mild
Triton X-100 (5%)	2.5	2	31.4	Mild
Sodium oxalate	3.65	2	35.6	Mild
Mean %CV			**14.0**	
Median %CV			**12.4**	

[1]Substances organized by increasing %CV.

7.2.1.2 □ *Data from Dr. Joseph Sina (Merck)*

Intralaboratory Repeatability: In this study, 43 substances were tested in one laboratory using four corneas per test substance. A %CV value was calculated for the opacity and permeability values and the *In Vitro* Irritancy Score for each test substance (**Appendix E2**). However, only 29 of the test substances were evaluated using a protocol that incubated the corneas at 32˚C. The %CVs for the *In Vitro* Irritancy Scores of these 29 substances are shown in **Table 7-7**. The results are sorted by %CV from lowest to highest value. The ranges of %CV values for substances classified as severe irritants *in vitro* are 1.1 to 13 (n = 5). The within experiment mean and median %CV values for this study were 71 and 35%, respectively. Substances classified *in vitro* as mild irritants tended to have greater %CV values. A majority (21 of 29; 72%) of the test substances in this study were classified as mild irritants *in vitro* and, of these, 10 had *In Vitro* Irritancy Scores at or below the accepted the background score of 3 to 5, contributing to higher within experiment mean and median %CV values for this study in comparison with the Southee (1998) study, which included test substances with a greater range of irritancy.

7.2.1.3 □ *Data from Dr. John Harbell (IIVS) for Gettings et al. (1996)*

Intralaboratory Reproducibility: Dr. John Harbell provided permeability values (OD_{490}) for three replicate experiments performed in an individual laboratory for the 25 surfactant-based personal care cleaning formulations evaluated in Gettings et al. (1996). The mean permeability value of these three experiments, as well as the mean and %CV of these data also were provided. All of these data and statistics are shown in **Table 7-8**. The results are sorted by %CV from lowest to highest value. The between experiment mean and median %CV values for this study were 33.4 and 29, respectively, with a %CV range of 5% to 100%.

7.2.1.4 □ *Data from Dr. Freddy Van Goethem for Gautheron et al. 1994)*

Intralaboratory Repeatability: In this study, 52 substances were tested in 11-12 different laboratories. Dr. Freddy Van Goethem provided individual cornea data collected in one of the participating laboratories (Janssen Pharmaceutica), which used six corneas per test substance. A %CV value was calculated for the opacity and permeability values and the *In Vitro* Irritancy Score for each test substance (**Appendix E3**). The %CVs for the *In Vitro* Irritancy Scores of the 52 substances tested are shown in **Table 7-9**. The results are sorted by %CV from lowest to highest value. The ranges of %CV values for substances classified as severe irritants *in vitro* are 1.4 to 24.3 (n = 20). The within experiment mean and median %CV values for this study were 47% and 18%, respectively. Substances classified *in vitro* as mild irritants tended to have greater %CV values (ranging from 11.3% to 312.6% [n = 27]). These results were comparable to those obtained in the intralaboratory repeatability analysis of the BCOP data from Southee (1998) (see **Section 7.2.1.1**).

Table 7-7 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 4 (Dr. Sina, Merck)[1]

Substance	Mean *In Vitro* Irritancy Score (n = 4 corneas)	%CV	*In Vitro* Prediction
3-Trichlorovinylaniline HCL	404	1.1	Severe
2-Amino-3,6-dimethylphenol, hydrobromide salt	150	7.1	Severe
Carbic anhydride	202	8.1	Severe
1,3-Benzenedicarboxaldehyde	29.8	10.9	Moderate
4-Bromo-2,5-dimethylphenol	131	11.7	Severe
Methyl 3-oxo-6-methoxyhexanoate	57.8	13	Severe
R-Hydroxy ester of benzoic acid compound	-12	14	Mild
Quinaldine (2-methylquinoline)	25.5	19	Moderate
Mixture of 2-chloromethyl-4,7-dimethylbenzoxazole and 2-bromomethyl dimethylbenzoxazole	18.1	19.1	Mild
Carbonitrile	21.8	21.5	Mild
Methyl boronic acid	25.1	26.6	Moderate
alpha-Pyranol, 7,7-dioxide	31.5	27.7	Mild
7-Chloroquinaldine	10.6	28.4	Mild
+-Butyl-3R-hydroxy-6-methoxyhepanoate	22.8	28.8	Mild
Cyano methylpyridine	15.5	34	Mild
Cyclic peptide	7.9	36.9	Mild
Substituted cephalosporanic acid	-4.4	40	Mild
S-Hydroxy ester of benzoic acid compound	20.8	42	Mild
t-Butyl-3-oxo-6-methoxyhexanoate	15.3	49	Mild
Aglycone; natural product	11.3	52.4	Mild
N-Acetyl-*p*-anisidine	8.38	58.7	Mild
Cyanopyridinone	-4.3	64	Mild
N-Sulfonamido hydroxyacetophenone	-5.8	117	Mild
Nitropyridinone	-3.7	124	Mild
3-Bromo-7-methyl-9-flurenone	-2.6	140	Mild
Cyclic peptide	2.7	175	Mild
Dimethyl ethylimidazo pyridine	3.36	200	Mild
tert-Butyl-6-methoxy-3-S-(2-thiophenethio) hexanoate	1.5	221	Mild
4-(2-Quinolylmethoxy)aniline	2.8	479	Mild
Mean %CV	**71**		
Median %CV	**35**		

[1]Substances organized by increasing %CV.

Table 7-8 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 5, Microbiological Associates[1]

Formulation	Permeability – O.D. units					
	Exp. 1	Exp. 2	Exp. 3	Mean	SD	%CV
Skin Cleaner - HZI	0.782	0.728	0.796	0.77	0.04	5
Shower Gel - HZS	1.488	1.501	1.655	1.55	0.09	6
Facial Cl Foam - HZR	0.215	0.244	0.259	0.24	0.02	9
Liquid Soap 1 - HZB	0.198	0.176	0.223	0.20	0.02	12
Shampoo 4 - HZV	0.306	0.219	0.279	0.27	0.04	17
Baby Shampoo 2 - HZF	0.505	0.342	0.427	0.42	0.08	19
Baby Shampoo 1 - HZP	0.285	0.202	0.296	0.26	0.05	20
Shampoo 3 - HZM	0.229	0.254	0.16	0.21	0.05	23
Shampoo AntiD - HZY	0.756	0.709	1.075	0.85	0.20	24
Gel Cleaner - HZE	0.186	0.15	0.246	0.19	0.05	25
Shampoo 6 - HZN	0.283	0.184	0.333	0.27	0.08	28
Liquid Soap 2 - HZW	0.356	0.21	0.417	0.35	0.10	28
Shampoo 8 - HZG	0.22	0.131	0.24	0.20	0.06	29
Foam Bath - HZL	0.625	0.976	1.136	0.91	0.26	29
Cleaning Gel – HZQ	0.214	0.114	0.165	0.16	0.05	30
Hand Soap - HZU	0.348	0.187	0.344	0.29	0.09	31
Shampoo 1 - HZC	1.193	0.612	1.067	0.96	0.31	32
Bubble bath - HZK	1.33	0.753	0.785	0.96	0.32	34
Shampoo 5 - HZD	0.318	0.15	0.225	0.23	0.08	36
Shampoo 7 - HZA	0.562	0.406	0.251	0.41	0.16	38
Shampoo 2 - HZX	0.582	0.498	1.036	0.71	0.29	41
Mild Shampoo - HZJ	0.064	0.021	0.066	0.05	0.03	51
Eye Makeup Remover - HZH	0.029	0.001	0.029	0.02	0.02	82
Polishing Scrub - HZT	0.002	0	0.002	0.001	0.00	87
Facial Cleaner - HZZ	0.008	0.004	0	0.004	0.00	100
Mean %CV	**33.4**					
Median %CV	**29.0**					

[1]Substances organized by increasing %CV.

Table 7-9 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 9 (Gautheron et al. 1994)[1]

Substance	Mean *In Vitro* Irritancy Score (n = 6 corneas)	%CV	*In Vitro* Prediction
2-Ethoxyethanol	84.4	1.4%	Severe
Cyclohexanone	141.7	5.8%	Severe
Gluconolactone	87.5	6.0%	Severe
2,4-Pentanedione	50.3	6.8%	Moderate
Promethazine hydrochloride	139.2	7.3%	Severe
Furan	50.2	7.9%	Moderate
Deoxycholic acid, sodium salt	99.6	8.0%	Severe
Benzethonium chloride	165.9	8.8%	Severe
Hexadecyltrimethylammonium bromide	69.9	9.9%	Severe
Quinacrine	57.9	10.0%	Severe
Octanol	60.9	11.2%	Severe
1-Nitropropane	16.6	11.3%	Mild
N-Lauroylsarcosine, sodium salt	62.6	11.6%	Severe
Allyl alcohol	123.3	11.7%	Severe
Butyrolactone	41.6	12.0%	Moderate
1-Phenyl-3-pyrazolidone	13.2	12.4%	Mild
Methanol	99.2	12.9%	Severe
Thiourea	151.4	13.7%	Severe
Ethanol	45.7	14.3%	Moderate
Dimethyl sulfoxide	9.4	14.4%	Mild
Ethyl acetoacetate	25.7	14.8%	Moderate
Pyridine	104.7	15.0%	Severe
2-Methoxyethanol	57.1	15.1%	Severe
Methylisobutyl ketone	19.4	15.9%	Mild
Dibenzoyl-L-tartaric acid	81.5	16.8%	Severe
Imidazole	64.3	17.3%	Severe
2-Aminophenol	13.0	19.0%	Mild
1,2,4-Trimethylbenzene	21.2	21.2%	Mild
1,2,3-Trichloropropane	91.1	22.0%	Severe
Aluminum hydroxide	9.9	23.2%	Mild
Diacetone alcohol	92.9	23.7%	Severe
Propyl-4-hydroxybenzoate	6.2	24.0%	Mild
Laurylsulfobetaine	102.4	24.3%	Severe
2,4-Dichloro-5-sulfamoylbenzoic acid	19.2	24.7%	Mild
3-Glycidoxypropyltrimethoxysilane	17.6	26.7%	Mild
Triethanolamine	3.0	34.5%	Mild
Sodium oxalate	3.2	40.9%	Mild
Triton X-155	3.1	53.3%	Mild
Tetraaminopyrimidine sulfate	2.5	54.7%	Mild
BRIJ-35	1.0	61.7%	Mild
EDTA, dipotassium salt	0.9	63.1%	Mild
Betaine monohydrate	3.5	63.7%	Mild
Magnesium carbonate	0.7	71.4%	Mild
Phenylbutazone	0.5	80.1%	Mild
Anthracene	1.4	87.4%	Mild
Petroleum ether	2.1	91.4%	Mild
Dimethylbiguanide	2.1	124.6%	Mild
Hexane	1.4	128.3%	Mild

Substance	Mean *In Vitro* Irritancy Score (n = 6 corneas)	%CV	*In Vitro* Prediction
2-Mercaptopyrimidine	-0.2	167.3%	Mild
DL-Glutamic acid	-0.2	221.3%	Mild
Iminodibenzyl	0.2	278.9%	Mild
MYRJ-45	0.5	312.6%	Mild
Mean %CV	**46.8%**		
Median %CV	**18.1%**		

[1]Substances organized by increasing %CV

7.2.2 Evaluation of Interlaboratory Reproducibility

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

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

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

In the first approach, a qualitative assessment of reproducibility was conducted. In this evaluation, the individual laboratory *in vitro* ocular irritation classification for each substance was used to evaluate the extent of agreement among the participating laboratories in their ability to identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants. The reliability of BCOP was assessed separately for each study (i.e., publication) reviewed in **Sections 4.0** and **5.0**. In an alternative approach, the reliability of BCOP was assessed after combining test results across comparative studies that used the same data analysis method (i.e., use of *In Vitro* Irritancy Score). **Section 6.0** provides a further description of how data were treated for each type of analysis. Substances classified, based on BCOP data, as corrosive/severe irritants or nonsevere irritants/nonirritants were further classified by their *in vivo* rabbit eye test results, as determined within the GHS, EPA, and EU classification schemes. Because the focus of this reliability assessment is on the interlaboratory reproducibility of BCOP in identifying corrosives/severe irritants versus nonsevere irritants/nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants (e.g., three laboratories could classify a chemical as a nonirritant and one laboratory could classify the same chemical as an moderate irritant; for this analysis this would be considered 100% agreement between laboratories) that would not be apparent from this analysis.

In the second approach, a quantitative assessment of reproducibility was determined by calculating the CV for test substance data for which *In Vitro* Irritancy Scores were available from multiple laboratories. The reproducibility of BCOP was assessed for the studies (i.e.,

publication) reviewed in **Sections 4.0** and **5.0** where individual testing laboratory data were available.

7.2.2.1 *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). The agreement of classification calls among participating laboratories and the relationship to the *in vivo* classification (GHS; UN 2003) for the substances tested in each validation in each study is provided in **Table 7-10**.

For the study by Balls et al. (1995), the five participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 41 (68%) of the 60 substances tested. 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). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (i.e., positive *in vitro* but negative *in vivo*). For instance, 63% (36% + 27%) of the false positives exhibited less than 100% agreement in the irritancy classifications among laboratories.

For the study by Gautheron et al. (1994), there was 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 substances, which were tested in either 11 or 12 laboratories. Discordance in the classification results was present for substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants.

For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

7.2.2.2 *Interlaboratory Reproducibility of Hazard Classification Category Using the EPA 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). The agreement of classification calls among participating laboratories and its relationship to the *in vivo* classification (EPA 1996) for the substances tested in each validation in each study is provided in **Table 7-11**.

Table 7-10 Evaluation of the Reliability of the BCOP Test Method in Predicting Ocular Corrosives and Severe 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% 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	17	13 (76%)			3 (18%)			1 (6%)	
	+/-	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+	5	11	4 (36%)			4 (36%)			3 (27%)	
	-/-	5	21	16 (76%)			2 (10%)			3 (14%)	
	?/-	5	4	3 (75%)						1 (25%)	
	?/+	5	2	2 (100%)							
	Total		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+	11 12	5 1	3 (60%) 1 (100%)		1 (20%)					1 (20%)
	+/-	11 12	1 1	1 (100%)		1 (100%)					
	-/+	11 12	4 5	2 (50%) 2 (40%)	1 (20%)	1 (25%)		1 (25%)			2 (40%)
	-/-	11 12	15 15	12 (80%) 13 (86%)	1 (7%)	2 (13%) 1(7%)			1 (7%)		
	?/-	11 12	1 1	1(100%)				1 (100%)			
	?/+	11	2		1 (50%)				1 (50%)		
	Total		51	35 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)		3 (6%)
Southee (1998)	+/+	3	4	4 (100%)							
	+/-	3	3	3 (100%)							
	-/+	3	1	1 (100%)							
	-/-	3	7	6 (86%)					1 (14%)		
	?/-	3	1	1 (100%)							
	?/+	-	0								
	Total		16	15 (94%)					1 (6%)		

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category 1); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category 2A, 2B) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; 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.

Table 7-11 Evaluation of the Reliability of the BCOP Test Method In Predicting Ocular Corrosives and Severe 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	13	10 (77%)			2 (15%)			1 (8%)	
	+/-	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+	5	13	5 (38%)			5 (38%)			3 (23%)	
	-/-	5	22	15 (68%)			4 (18%)			3 (14%)	
	?/-	5	3	3 (100%)							
	?/+	5	4	4 (100%)							
	Total		60	40 (67%)			12 (20%)			8 (13%)	
Gautheron et al. (1994)	+/+	11 12	4 1	2 (50%) 1 (100%)		1 (25%)					1 (25%)
	+/-	11 12	1 1	 1 (100%)		1 (100%)					
	-/+	11 12	6 5	3 (50%) 2 (40%)	 1 (20%)	1 (17%)		1 (17%)	1 (17%)	 1 (20%)	 1 (20%)
	-/-	11 12	15 15	12 (80%) 13 (86%)	 1 (7%)	2 (13%) 1 (7%)			1 (7%)		
	?/-	11 12	1 1	 1 (100%)				1 (100%)			
	?/+	11	1		1 (100%)						
	Total		51	35 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)	1 (2%)	2 (4%)
Southee (1998)	+/+	3	2	2 (100%)							
	+/-	3	3	3 (100%)							
	-/+	3	2	2 (100%)							
	-/-	3	7	6 (86%)					1 (14%)		
	?/-	3	1	1 (100%)							
	?/+	3	1	1 (100%)							
	Total		16	15 (94%)					1 (6%)		

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

[2]n indicates number of substances.

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

The participating laboratories of Balls et al. (1995) were in 100% agreement in regard to the ocular irritancy classification for 40 (67%) of the 60 substances tested. The agreement among laboratories was greatest for accurately identified corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (77% of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives. For instance, 61% (38% + 23%) of the false positives exhibited less than 100% agreement among laboratories in the irritancy classifications.

The participating laboratories of Gautheron et al. (1994) were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 35 (69%) of the 51 tested substances. Discordant results were observed for substances that were correctly identified as corrosive/severe irritant or nonsevere/irritant/nonirritant, as well as for false negatives and false positives.

For the report by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

7.2.2.3 *Interlaboratory Reproducibility of Hazard Classification Category Using the EU 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). The agreement of classification calls among participating laboratories and its relationship to the *in vivo* classification (EU 2001) for the substances tested in each validation in each study is provided in **Table 7-12**.

The participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 41 (68%) of the 60 substances tested by Balls et al. (1995). The extent of agreement among laboratories was greatest for accurately identified corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (86% of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives, false negatives, and those substances accurately classified as nonsevere irritants/nonirritants. For instance, 63% (36% + 27%) of the false positives, 60% (20% + 40%) of the false negatives and 25% (10% + 15%) of the correctly identified nonsevere irritants/nonirritants exhibited less than 100% agreement among laboratories in irritancy classifications.

The participating laboratories in Gautheron et al. (1994) were in 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 tested substances. Substances that were classified as false positives exhibited the most discordant results, with 60% (20% + 20% + 20%) of false positives exhibiting less than 100% classification agreement among laboratories.

Table 7-12 Evaluation of the Reliability of the BCOP Test Method In Predicting Ocular Corrosives and Severe 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	14	12 (86%)			2 (14%)				
	+/-	5	5	2 (40%)			1 (20%)			2 (40%)	
	-/+	5	11	4 (36%)			4 (36%)			3 (27%)	
	-/-	5	20	15 (75%)			2 (10%)			3 (15%)	
	?/-	5	5	5 (100%)							
	?/+	5	5	3 (60%)			1 (20%)			1 (20%)	
	Total		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+	11 12	5 1	3 (60%) 1 (100%)		1 (20%)					1 (20%)
	+/-	11 12	1 1	 1 (100%)		1 (100%)					
	-/+	11 12	5 5	2 (40%) 2 (40%)	 1 (20%)	1 (20%)		1 (20%)	1 (20%)	 1 (20%)	 1 (20%)
	-/-	11 12	15 15	12 (80%) 13 (86%)	 1 (7%)	2 (13%) 1 (7%)			1 (7%)		
	?/-	11 12	1 1	 1 (100%)				1 (100%)			
	?/+	11	1		1 (100%)						
	Total		51	35 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)	1 (2%)	2 (4%)
Southee (1998)	+/+	3	4	4 (100%)							
	+/-	3	2	2 (100%)							
	-/+	3	1	1 (100%)							
	-/-	3	7	6 (86%)					1 (14%)		
	?/-	3	2	2 (100%)							
	?/+	-	0								
	Total		16	15 (94%)					1 (6%)		

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

[2]n indicates number of substances.

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

For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

7.2.2.4 *Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results*

For the Gautheron et al. (1994) study, 16 substances showed interlaboratory differences in *in vitro* classification (**Table 7-13**). Of these, nine (56%) are organic solvents, including five alcohols, one lactone, one ketone, one heterocyclic compound, and one chlorinated hydrocarbon. Four surfactants, four heterocyclic compounds (two solids and two liquids), and one acid (a solid) also showed interlaboratory differences in *in vitro* classification. Of the 10 liquid substances that produced discordant interlaboratory results in this study, nine are organic solvents.

For the Balls et al. (1995) study, 19 substances showed interlaboratory differences in *in vitro* classification (**Table 7-14**). Of these, 10 (53%) are organic solvents, including seven alcohols, one lactone, one ketone, and one ester. Two ethers, two carboxylic acids, two imides (solid), and one amidine also showed interlaboratory differences in *in vitro* classification. The ten liquid substances that produced discordant interlaboratory results in this study are all organic solvents.

7.2.2.5 *Interlaboratory Reproducibility Based on Coefficient of Variation Analysis of In Vitro Scores*

To provide a quantitative assessment of interlaboratory variability, individual laboratory BCOP test results were used to calculate a mean and CV for the *In Vitro* Irritancy Score for each substance tested in Gautheron et al. (1994), Balls et al. (1995) and Southee (1998) (**Tables 7-15, 7-16, 7-17**).

For the Gautheron et al. (1994) study, a wide range of %CV values for individual substances is evident for the *In Vitro* Irritancy Score (**Table 7-15**). The mean and median %CV values were 168% and 47%, respectively, ranging from 16.5% to 1325% for the entire set of 52 test substances. The 17 substances predicted as severe in the BCOP assay had mean and median %CV values of 36% and 17%, respectively, with a %CV range from 16.5% to 55.7%. Substances classified *in vitro* as mild irritants (i.e., *In Vitro* Irritancy Score < 25) tended to have much greater %CV values. About half (25 of 52; 48%) of the substances tested in this study were classified as mild irritants *in vitro* and, of these, 18 had *In Vitro* Irritancy Scores at or below the accepted the background score of 3 to 5, contributing to a high mean and median %CV for this study. All of the %CV values for individual substances greater than 75% (n = 17) resulted from substances that had *In Vitro* Irritancy Scores at or below the accepted background score of 3 to 5.

Table 7-13 Chemical and Product Classes of Test Substances with Discordant Interlaboratory Results in the Gautheron et al. (1994) Study

Substance	Chemical Class	Product Class	Physical Form Tested	*In Vitro* Classification (% of Labs with Classification)
Butyrolactone	Lactone; Heterocyclic	Solvent; Synthetic intermediate	Liquid	Moderate (10/12; 83%) Severe (2/12; 17%)
Deoxycholic acid, sodium salt	Alcohol; Carboxylic acid	Surfactant	10% Solution	Severe (11/12; 92%) Moderate (1/12; 8%)
Diacetone alcohol	Alcohol; Ketone	Solvent	Liquid	Moderate (8/11; 73%) Severe (3/11; 27%)
2,4-Dichloro-5-sulfamoylbenzoic acid	Amide; Organic sulfur compound	Chemical intermediate	Solid	Mild (8/12; 67%) Moderate (3/12; 25%) Severe (1/12; 8%)
Ethanol	Alcohol	Solvent	Liquid	Severe (7/11; 64%) Moderate (4/11; 36%)
Furan	Heterocyclic compound	Solvent; Chemical intermediate	Liquid	Severe (6/12; 50%) Moderate (6/12; 50%)
Hexadecyltrimethyl-ammonium bromide	Organic salt; Onium compound	Surfactant; Agricultural chemical; Germicide	Liquid	Severe (6/11; 55%) Moderate (5/11; 45%)
N-Lauroylsarcosine, sodium salt	Amide; Amino acid	Surfactant	10% Solution	Moderate (9/11; 82%) Severe (2/11; 18%)
Laurylsulfobetaine	Amine; Onium compound	Surfactant	10% Solution	Severe (10/11; 91%) Moderate (1/11; 9%)
Methanol	Alcohol	Solvent, Chemical intermediate	Liquid	Severe (8/11; 73%) Moderate (2/11; 12%) Mild (1/11; 9%)
2-Methoxyethanol	Alcohol	Solvent	Liquid	Severe (9/11; 82%) Moderate (2/11; 18%)
Octanol	Alcohol	Solvent	Liquid	Moderate (6/11; 55%) Severe (4/11; 36%) Mild (1/11; 9%)
2,4-Pentanedione	Alcohol	Solvent	Liquid	Severe (7/12; 58%) Moderate (5/12; 42%)
Promethazine hydrochloride	Amidine; Heterocyclic compound; Organic sulfur compound	Drug/therapeutic agent	Solid	Severe (9/11; 82%) Moderate (1/11; 9%) Mild (1/11; 9%)
Quinacrine	Heterocyclic compound	Drug/therapeutic agent	Solid	Moderate (5/11; 45%) Mild (4/11; 36%) Severe (2/11; 18%)
1,2,3-Trichloropropane	Hydrocarbon	Solvent	Liquid	Moderate (8/11; 73%) Severe (2/11; 18%) Mild (1/11; 9%)

Table 7-14 Chemical and Product Classes of Test Substances with Discordant Interlaboratory Results in the Balls et al. (1995) Study

Substance	Chemical Class	Product Class	Physical Form	*In Vitro* Classification (No. of Laboratories)
Butyrolactone	Lactone; Heterocycle	Solvent; Synthetic intermediate	Liquid	Severe (3/5; 60%) Moderate (2/5; 40%)
Captan 90 concentrate	Imide; Organic sulfur compound	Pesticide	Solid	Moderate (4/5; 80%) Severe (1/5; 20%)
Cetylpyridinium bromide (10%)	Heterocyclic compound; Onium compound	Surfactant, Germicide	10% Solution	Severe (4/5; 80%) Moderate (1/5; 20%)
Cyclohexanol	Alcohol	Solvent; Chemical intermediate	Liquid	Moderate (3/5; 60%) Severe (2/5; 40%)
Ethanol	Alcohol	Solvent	Liquid	Severe (4/5; 80%) Moderate (1/5; 20%)
2-Ethyl-1-hexanol	Alcohol	Solvent	Liquid	Severe (2/5; 40%) Moderate (2/5; 40%) Mild (1/5; 20%)
Fomesafen	Imide; Ether; Nitro compound	Pesticide	Solid	Severe (2/5; 40%) Mild (2/5; 40%) Moderate (1/5; 20%)
n-Hexanol	Alcohol	Solvent	Liquid	Severe (3/5; 60%) Moderate (2/5; 40%)
Isobutanol	Alcohol	Solvent	Liquid	Moderate (3/5; 60%) Severe (2/5; 40%)
Isopropanol	Alcohol	Solvent	Liquid	Severe (3/5; 60%) Moderate (2/5; 40%)
Maneb	Amine/Amidine; Organic salt	Pesticide	Solid	Severe (2/5; 40%) Mild (2/5; 40%) Moderate (1/5; 20%)
Methyl acetate	Ester	Solvent	Liquid	Moderate (4/5; 80%) Severe (1/5; 20%)
Methyl ethyl ketone	Ketone	Solvent	Liquid	Severe (4/5; 80%) Moderate (1/5; 20%)
1-Napthalene acetic acid	Carboxylic acid; Polycyclic compound;	Pesticide	Solid	Severe (4/5; 80%) Moderate (1/5; 20%)
n-Octanol	Alcohol	Solvent	Liquid	Moderate (3/5; 60%) Severe (1/5; 20%) Mild (1/5; 20%)
Sodium lauryl sulfate (15%)	Carboxylic acid (salt)	Surfactant	10% Solution	Severe (3/5; 60%) Moderate (2/5; 40%)
Trichloroacetic acid (3%)	Carboxylic acid	Herbicide; chemical intermediate	Solution	Severe (4/5; 80%) Moderate (1/5; 20%)
Triton X-100 (5%)	Ether	Surfactant	10% Solution	Severe (4/5; 80%) Moderate (1/5; 20%)
Triton X-100 (10%)	Ether	Surfactant	10% Solution	Severe (4/5; 80%) Moderate (1/5; 20%)

Table 7-15 Coefficient of Variation Analysis of the Interlaboratory Variability of the BCOP Test Method for Gautheron et al. (1994)[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
2-Ethoxyethanol	91.3	12	16.5	Severe
2,4-Pentanedione	59.8	12	24	Severe
Allyl alcohol	156	12	27	Severe
Imidazole	87.9	12	28.5	Severe
Furan	56	12	29.4	Severe
Benzethonium chloride	133.9	11	31.7	Severe
Butyrolactone	45.6	12	32.2	Moderate
Cyclohexanone	105.6	11	33.3	Severe
2-Methoxyethanol	63.5	11	33.6	Severe
Laurylsulfobetaine	80.6	11	34	Severe
Ethyl acetoacetate	31.8	11	34.9	Moderate
Gluconolactone	76.6	11	35	Severe
Methylisobutyl ketone	19.9	11	36	Mild
Pyridine	112.8	11	38.4	Severe
Ethanol	60.7	11	39.1	Severe
3-Glycidoxypropyltrimethoxysilane	16.6	12	40	Moderate
N-Lauroylsarcosine, sodium salt	50	11	41.7	Moderate
Octanol	47.4	11	41.7	Moderate
Deoxycholic acid, sodium salt	93.5	12	43	Severe
2-Aminophenol	7	12	43.5	Mild
Hexadecyltrimethylammonium bromide	66.4	11	45.2	Severe
1-Phenyl-3-pyrazolidone	12.9	12	46.5	Mild
Dibenzoyl-L-tartaric acid	120.5	11	46.8	Severe
Dimethyl sulfoxide	11.4	11	46.9	Mild
1-Nitropropane	7.6	12	46.9	Mild
1,2,4-Trimethylbenzene	16.1	12	47	Mild
Propyl-4-hydroxybenzoate	7.9	11	48	Mild
Promethazine hydrochloride	112.4	11	49.3	Severe
1,2,3-Trichloropropane	47.5	11	50.3	Moderate
Diacetone alcohol	53.5	11	50.8	Moderate
Methanol	84.2	11	55.7	Severe
2,4-Dichloro-5-sulfamoylbenzoic acid	26.3	12	58.5	Moderate
Sodium oxalate	4.8	12	66	Mild
Quinacrine	31.1	11	74.8	Moderate
Petroleum ether	5.5	12	75.4	Mild
Dimethylbiguanide	2.9	11	82	Mild
Magnesium carbonate	3	11	83	Mild
Triethanolamine	2.2	11	101.5	Mild
Aluminum hydroxide	6.8	12	107	Mild
Tetraaminopyrimidine sulfate	6	11	107	Mild
Hexane	1.4	12	143	Mild
Iminodibenzyl	2.4	11	177.5	Mild
2-Mercaptopyrimidine	-1.25	12	208	Mild
Triton X-155	0.55	11	276	Mild
DL-Glutamic acid	0.58	12	330.6	Mild

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
Anthracene	-0.33	12	430	Mild
Betaine monohydrate	0.92	12	432	Mild
MYRJ-45	-0.18	11	962	Mild
EDTA, di-potassium salt	-0.33	12	1009	Mild
BRIJ-35	-0.09	11	1280	Mild
Phenylbutazone	-0.17	12	1325	Mild
Mean %CV	**167.6 (all substances) 84 (excluding MYRJ-45, EDTA, BRIJ-35, phenylbutazone)**			
Median %CV	**46.9**			

[1] Substances organized by increasing %CV.

For the Balls et al. (1995) study, a wide range of %CV values for individual substances is evident for the *In Vitro* Irritancy Score (**Table 7-16**). The mean and median %CV values were 125% and 30.6%, respectively, ranging from 7.6% to 4511% for the entire set of 59 test substances. The 32 substances predicted as severe in the BCOP assay had mean and median %CV values of 25% and 22%, respectively, with a %CV range from 7.6% to 89.4%.

Table 7-17 presents the %CV values for the *In Vitro* Irritancy Score of individual substances tested in the Southee (1998) study. The mean and median %CV values were 32.4% and 22.8%, respectively, with a range of 7.5% to 108.8% for the entire set of test substances.

7.2.3 Additional Analyses of Interlaboratory Reproducibility

The EC Interlaboratory Study (Gautheron et al. 1994): This study found that 82.7% of the substances tested were classified the same by all laboratories when using a three-category system. In this system, substances were classified into one of the following categories: mild irritant (BCOP score [0-25], moderate irritant [25.1-55], and severe irritant [≥55.1]).

The EC/HO Validation Study (Balls et al. 1995): The study authors determined the interlaboratory correlation of BCOP results (permeability value, opacity value and *In Vitro* Irritancy Score) generated from the five laboratories that participated in the EC/HO study (**Table 7-18**). In this analysis, each laboratory was compared to each other laboratory in a pair-wise fashion for all 60 substances tested, as well as for subsets of test substances (water-soluble, water-insoluble, surfactants solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances as shown in **Table 7-18** (see **Appendix F** for all correlation coefficients derived from comparing each laboratory with every other laboratory). Interlaboratory correlation coefficients for the *In Vitro* Irritancy Score generally spanned a range of 0.867 to 0.958 depending on the specific subsets of substances being evaluated. However, the correlation coefficients for the permeability value were lower (e.g., correlation coefficients BCOP – Permeability Value ranged from 0.683 to 0.906 for the full set of test substances). The correlation coefficients for the Opacity Value were slightly higher (0.898 to 0.978) than the correlation for the for the *In Vitro* Irritancy Score.

Table 7-16 Coefficient of Variation Analysis of the Interlaboratory Variability of the BCOP Test Method for Balls et al. (1995)[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
1-Naphthalene acetic acid, Na salt	149.2	5	7.6	Severe
Benzalkonium chloride (10%)	136.5	5	10.9	Severe
Sodium hydroxide (1%)	150	5	12.3	Severe
Cetylpyridinium bromide (6%)	71.2	5	12.7	Severe
Acetone	123	5	14	Severe
Imidazole	112.7	5	14.5	Severe
Benzalkonium chloride (5%)	128.5	5	15.6	Severe
Methyl acetate	54.9	5	17.4	Moderate
Sodium hydroxide (10%)	271.9	5	17.6	Severe
Toluene	35.6	5	18.1	Moderate
Chlorhexidine	114	5	18.3	Severe
Trichloroacetic acid (30%)	264	5	18.7	Severe
Dibenzyl phosphate	378	5	18.8	Severe
2,2-Dimethylbutanoic acid	111.9	5	19.5	Severe
Pyridine	148	5	20.1	Severe
Promethazine hydrochloride	121.4	5	20.4	Severe
Trichloroacetic acid (3%)	75.9	5	21.1	Severe
Benzalkonium chloride (1 %)	88.8	5	21.7	Severe
Parafluoraniline	30.4	5	21.7	Moderate
Methyl ethyl ketone	70.4	5	22.6	Severe
4-Carboxybenzaldehyde	78.3	5	24	Severe
Ethanol	70.6	5	24.1	Severe
Cetylpyridinium bromide (10%)	72	5	24.2	Severe
Triton X-100 (5 %)	78.3	5	24.2	Severe
Triton X-100 (10 %)	70.3	5	25.3	Severe
Isobutanol	56	5	26.1	Severe
n-Hexanol	61.9	5	27	Severe
Sodium lauryl sulfate (15 %)	63.3	5	28	Severe
Cyclohexanol	60.1	5	28.5	Severe
2,6-Dichlorobenzoyl chloride	10.4	5	30.6	Mild
Sodium lauryl sulfate (3 %)	25.8	5	30.9	Mild
Isopropanol	57.9	5	31.3	Severe
Sodium perborate	97	5	35.8	Severe
Methyl isobutyl ketone	12.6	5	36	Mild
1-Naphthalene acetic acid	78.1	5	37.4	Severe
Butyl acetate	34.6	5	38.4	Moderate
Methyl cyanoacetate	12.2	5	39.2	Mild
Ethyl acetate	32	5	40.5	Moderate
Potassium cyanate	15	5	40.9	Mild
2,5-Dimethylhexanediol	20.8	5	41.6	Mild
Benzoyl-L-tartaric acid	169.6	5	43	Severe
gamma-Butyrolactone	60.7	5	45	Severe
Tetraaminopyrimidine sulfate	15.1	5	46.3	Mild
Methylcyclopentane	2.8	5	47.8	Mild
2-Ethyl-1-hexanol	39.8	5	48.2	Moderate
Cetylpyridinium bromide (0.1%)	9.2	5	51.4	Mild

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
Maneb	40.5	5	58.3	Moderate
n-Octanol	40.9	5	58.8	Moderate
Ethyl-2-methylacetoacetate	14.4	5	65.3	Mild
Ethyl trimethyl acetate	17.8	5	66.3	Mild
Ammonium nitrate	9.8	5	69.7	Mild
L-Aspartic acid	1.3	5	73.6	Mild
Captan 90 concentrate	43.8	5	75.8	Moderate
Quinacrine	1.6	5	76.9	Mild
Fomesafen	60.7	5	89.4	Severe
Sodium oxalate	14	5	143	Mild
Polyethylene glycol 400	1.1	5	145	Mild
Glycerol	0.26	5	712	Mild
Tween 20	-0.04	5	4511	Mild
Mean %CV	**125 (all test substances) 50 (excluding Tween 20)**			
Median %CV	**30.6**			

[1]Substances organized by increasing %CV.

Table 7-17 Coefficient of Variation Analysis of the Interlaboratory Variability of the BCOP Test Method for Southee (1998)[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
Butyl cellosolve	100.9	3	7.5	Severe
Benzalkonium chloride	160	3	8.5	Severe
NaOH (10%)	226	3	8.6	Severe
Imidazole	136.9	3	9.1	Severe
4-Carboxybenzaldehyde	46.7	3	9.5	Moderate
Parafluoroaniline	32.1	3	19.1	Moderate
Methyl ethyl ketone	82.5	3	21.6	Severe
Ethanol	48.7	3	22.1	Moderate
Ammonium nitrate	5.03	3	23.4	Mild
Hexadecyltrimethylammonium bromide (10%)	29.3	3	27.1	Moderate
Glycerol	0.72	3	33.5	Mild
Propyl-4-hydroxybenzoate	6.9	3	37.7	Mild
Triton X-100 (5%)	3.3	3	44.8	Mild
Sodium lauryl sulfate (15%)	9.7	3	57.1	Mild
Tween 20	0.23	3	79.8	Mild
Sodium oxalate	3.6	3	108.8	Mild
Mean %CV	**32.4**			
Median %CV	**22.8**			

[1]Substances organized by increasing %CV.

Table 7-18 Interlaboratory Correlation Ranges Determined for Various Subsets of Tested Substances in Balls et al. (1995)

BCOP Test Method Value	Interlaboratory Pearson's Correlation Coefficient (r) of the *In Vitro* Data
Full set of test substances[1] (60)	
BCOP - Permeability Value	0.683-0.906
BCOP - Opacity Value	0.898-0.978
BCOP - *In Vitro* Irritancy Score	0.867-0.958
Chemicals soluble in water (30)	
BCOP - Permeability Value	0.521-0.880
BCOP - Opacity Value	0.927-0.971
BCOP - *In Vitro* Irritancy Score	0.855-0.952
Chemicals insoluble in water (18)	
BCOP - Permeability Value	0.688-0.963
BCOP - Opacity Value	0.896-0.991
BCOP - *In Vitro* Irritancy Score	0.898-0.981
Surfactants (12)	
BCOP - Permeability Value	0.766-0.966
BCOP - Opacity Value	0.947-0.995
BCOP - *In Vitro* Irritancy Score	0.914-0.989
Solids (20)	
BCOP - Permeability Value	0.563-0.934
BCOP - Opacity Value	0.903-0.977
BCOP - *In Vitro* Irritancy Score	0.852-0.960
Solutions (14)	
BCOP - Permeability Value	0.731-0.933
BCOP - Opacity Value	0.955-0.989
BCOP - *In Vitro* Irritancy Score	0.914-0.980
Liquids (26)	
BCOP - Permeability Value	0.612-0.893
BCOP - Opacity Value	0.913-0.967
BCOP - *In Vitro* Irritancy Score	0.851-0.956

[1]As noted in **Section 3.0**, one substance (thiourea) was tested *in vitro* in the BCOP assay but, due to its excessive toxicity *in vivo*, it was excluded from the comparison of *in vitro* and *in vivo* test results, and thus excluded from the evaluation in **Section 7.2.1**. However, *in vitro* data for this substance was included in the original Balls et al. (1995) analysis.

7.3 Historical Positive and Negative Control Data

An example of historical data for positive controls was provided by IIVS (current as of July 22, 2004), as shown in **Table 7-19**.

Table 7-19 Historical Positive Control Data for the BCOP Assay

Positive Control	Opacity	OD_{490}	*In Vitro* Score
Ethanol (10 min exposure)			
Mean (n = 632)	31.2	1.422	52.7
SD	4.8	0.345	6.4
CV	15.3%	24.3%	12.1%
Upper and lower limits[1]	21.7 – 40.7	0.742 – 2.112	39.9 – 65.4
Imidazole (4 hour exposure)			
Mean (n = 125)	76.4	1.768	103.0
SD	18.4	0.488	16.6
CV	24.1%	27.6%	16.2%
Upper and lower limits*	39.7 – 113.2	0.792 – 2.745	69.7 – 136.2

Abbreviations: CV = Coefficient of variation; n = Number of tests; SD = Standard deviation.
[1]The upper and lower limits are the upper and lower 95% confidence limits (+/- 2 SDs) around the mean.

7.4 Summary

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from three studies (Southee 1998; Dr. Sina's submission; Dr. Van Goethem's submission) indicates the extent of intralaboratory repeatability of the BCOP test method for substances predicted as severe eye irritants. For the 16 substances evaluated in the Southee (1998) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas ranged from 11.8 to 14.2 for the three laboratories. For the 29 substances evaluated by Dr. Sina, the within experiment mean and median %CV values for *In Vitro* Irritancy Scores were 71 and 35, respectively. The dataset provided by Dr. Sina included 10 substances with low *In Vitro* Irritancy Scores around the background range of the assay (< 3.5), contributing to the increased variability of this dataset. However, the range of %CV values for the five substances predicted as severe irritants (*In Vitro* Scores >55.1) in this study is 1.1 to 13. For the 52 substances evaluated by Dr. Van Goethem in the Gautheron et al. (1994) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas was 18.1%, comparable to the results obtained with the data from Southee (1998).

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from two studies (Gettings et al. 1996; Southee 1998) indicates the extent of intralaboratory reproducibility of the BCOP test method for substances predicted as severe eye irritants. For the Gettings et al. (1996) study, the between experiment (n = 3) mean and median %CV values for permeability values were 33.4 and 29.0, respectively, for 25 surfactant-based personal care cleaning formulations. For the Southee (1998) study, the mean %CV values for *In Vitro* Irritancy Scores for the 16 substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

A qualitative assessment of the data provided for multiple laboratories in three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) indicates the extent of interlaboratory reproducibility. In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the five participating laboratories for the Balls et al.

(1995) study were in 100% agreement in regard to the ocular irritancy classification for 40 to 41 (67% to 68%) of the 60 substances tested *in vitro* in the study, depending on the classification system used. 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% to 86% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). For the study by Gautheron et al. (1994), regardless of the classification system used, there was 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 substances, which were tested in either 11 or 12 laboratories. For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances, regardless of the classification system used. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as surfactants, organic solvents, chemical intermediates, detergents, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) by performing a %CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. For the Gautheron et al. (1994) study, the 17 substances predicted as severe in the BCOP assay had mean and median %CV values of 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories. For the Balls et al. (1995) study, the 32 substances predicted as severe in the BCOP assay had mean and median %CV values of 25% and 22%, respectively, for results obtained in five laboratories. For the Southee (1998) study, the mean and median %CV values for the *In Vitro* Irritancy Scores of the 16 substances were 32.4% and 22.8%, respectively, for three laboratories.

Balls et al. (1995) also determined the interlaboratory correlation between BCOP test method endpoint data generated by each laboratory for all 60 substances tested[1], as well as for various subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances. Interlaboratory correlation coefficients for the *In Vitro* Irritancy Score generally spanned a range of 0.867 to 0.958 depending on the specific subsets of substances being evaluated. However, the correlation coefficients for the permeability value were lower (e.g., correlation coefficients BCOP – Permeability Value ranged from 0.683 to 0.906 for the full set of test substances). The correlation coefficients for the Opacity Value were higher (0.898 to 0.978) than the correlation for the *In Vitro* Irritancy Score.

[1] In some analyses of the Balls et al. (1995) validation results, 59 substances were considered. In other analyses, 60 substances were considered. The difference in the total number of substances is due to the exclusion of one substance, thiourea, in some analyses due to its excessive *in vivo* toxicity.

[This Page Intentionally Left Blank]

8.0 BCOP TEST METHOD DATA QUALITY

8.1 Adherence to National and International GLP Guidelines

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

Based on the available information, it appears that Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004) conducted the BCOP studies according to GLP guidelines.

The *in vivo* reference studies used for Gautheron et al. (1994), Balls et al. (1995), Southee (1998), and Bailey et al. (2004) appear to have adhered to GLP guidelines. Two of these studies (Balls et al. 1995; Southee1998) used *in vivo* reference data from the ECETOC Eye Irritation Reference Data Bank (ECETOC 1992). These *in vivo* data were generated in GLP-compliant studies conducted according to OECD TG 405 (OECD 1987). In Gautheron et al. (1994), the *in vivo* studies were performed according to European Economic Community (EEC) (1984 and 1991) guidelines, which presumably required adherence to GLP guidelines. Additionally, 48 of the test substances evaluated by Casterton et al. (1996) were included in the ECETOC (1992) publication; thus, the *in vivo* data for these substances were generated according to GLP guidelines. For Bailey et al. (2004), the *in vivo* study reports contained signed statements attesting that the studies were conducting according to GLP guidelines.

8.2 Data Quality Audits

Formal assessments of data quality, such as a quality assurance (QA) audit, generally involve a systematic and critical comparison of the data provided in a study report to the laboratory records generated for a study. No attempt was made to formally assess the quality of the *in vitro* BCOP data included in this BRD, or to obtain information about data quality audits from the authors of the BCOP study reports. The published data on the BCOP assay were limited to calculated *In Vitro* Irritancy Scores and, to a lesser extent, opacity and OD_{490} values. Auditing these reported values would require obtaining the original data for each BCOP experiment, which was not possible within the timeframe of this review.

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

- *Incomplete substance information*: Some BCOP study reports provided limited information about the substances tested. The CASRN, purity, and supplier of the test substances were not consistently reported. Thus, comparisons of data from different studies that evaluated test substances of the same chemical name must be interpreted with caution because of possible differences in test substance purity and suppliers.

- *Data reporting*: A majority of the BCOP studies reported only the mean *In Vitro* Irritancy score with no accompanying standard deviation to indicate the variability of the data.
- *Criteria for an acceptable test*: Acceptance criteria were reported in Balls et al. (1995) and Southee (1998). These reports stated that a test was accepted if the positive control produced an *In Vitro* Irritancy score within two standard deviations of the current historical mean. Although not reported, these same criteria were used in Gettings et al. (1996), Swanson et al. (1995), Swanson and Harbell (2000), and Bailey et al. (2004). However, acceptance criteria were not found for Gautheron et al. (1994) and Casterton et al. (1996).
- *Methodology*: The methods were presented in varying levels of detail and completeness in the study reports. The space limitation of many scientific journals is likely a contributing factor to some of the shorter methodology sections.

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

8.3 Impact of Deviations from GLP Guidelines

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

8.4 Availability of Laboratory Notebooks or Other Records

Study notebooks and other supporting records are known to be available, upon request, for an external audit, for the following studies: Swanson et al. (1995), Gettings et al. (1996), Swanson and Harbell (2000), and Bailey et al. (2004). The availability of laboratory notebooks or other records for the other studies considered for the accuracy (**Section 6.0**) and reliability (**Section 7.0**) analyses was not determined.

8.5 Need for Data Quality

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

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Reports in the Peer Reviewed Literature

A search of MEDLINE, TOXLINE, and Web of Science showed 14 additional scientific publications with BCOP test method results (Gautheron et al. 1992; Vanparys et al. 1993; Rachui et al. 1994; Rougier et al. 1994; Sina et al. 1995; Cassidy and Stanton 1997; Chamberlain et al. 1997; Bruner et al. 1998; Ubels et al. 1998, 2000, 2002, 2004; Cooper et al. 2001; Jones et al. 2001), as well as nine review articles (e.g., reports from a BCOP workshop) that discussed the assay and seven background articles (e.g., basis of the test method).

These studies were not included in previous sections of the BRD because they lacked sufficient information (e.g., substance names, *in vivo* data) for an evaluation of accuracy or reliability to be conducted. The first publication on the BCOP assay (Gautheron et al. 1992) was excluded because, for most of the substances tested in this study, only opacity results were reported. Eight studies lacked other necessary information with which to conduct an accuracy or reliability analysis. Vanparys et al. (1993), Rachui et al. (1994), Rougier et al. (1994), Cassidy and Stanton (1997), Bruner et al. (1998), Cooper et al. (2001), and Jones et al. (2001) lacked sufficient *in vivo* data. Sina et al. (1995) did not include the names of the substances tested. Additionally, the purpose of four studies (Ubels et al. 1998, 2000, 2002, 2004) was to investigate potential improvements of the BCOP assay, and test results were not compared to *in vivo* reference data.

In addition to these 14 studies, a retrospective evaluation of BCOP data was conducted by the Interagency Regulatory Alternatives Group (IRAG) (Chamberlain et al. 1997), in which eight data sets were submitted by eight laboratories on a total of 242 discrete (not all unique) substances. Due to the observation that at least two of the IRAG data sets had been published in reports reviewed in previous sections of the BRD, the data and other information in the IRAG report were not included to avoid duplication of BCOP studies and data. Additionally, detailed *in vivo* data, which are necessary for the analyses performed in **Section 6.0** of this document, were not received in response to NICEATM's requests for such data.

The correlative analyses conducted by Gautheron et al. (1994) and Balls et al. (1995) as a measure of test method performance are summarized below. These analyses were not included in **Section 6.0** since they are not relevant to an accuracy analysis of the BCOP data in relation to the EPA (1996), EU (2001), or GHS (UN 2003) classification systems for *in vivo* rabbit eye test data.

These 15 studies, presented in alphabetical order, are reviewed in the following subsections. In addition, summaries are provided for a 1997 workshop on the BCOP assay, along with two review articles on the test method. A list of recent poster presentations on the assay also is provided.

9.1.1 Balls et al. (1995)

Under the auspices of the British Home Office and Directorate General XI of the European Commission, a validation study on proposed alternatives to the *in vivo* rabbit ocular toxicity test method was conducted. The goal of the evaluation was to identify at least one non-whole animal test method that could be proposed to regulatory authorities as a replacement for the currently accepted *in vivo* eye irritation test method. A total of 52 substances were evaluated in 60 tests in two to five laboratories. Four of the test substances were evaluated at two different concentrations and two substances were evaluated at three different concentrations. The ocular irritancy potential of the test substances were ranked in terms of MMAS (which ranged from 0 to 108). *In vivo* data for 46 of the test substances, which were generated in compliance with OECD TG 405, were obtained from published sources. *In vivo* data for 14 of the test substances were obtained from concurrently conducted studies, which were in compliance with OECD TG 405. *In vivo* data in the report were presented as MMAS.

This study conducted correlative analyses of the BCOP scores and the *in vivo* MMAS scores. The Spearman's rank correlation test and Pearson's correlation analysis were used to compare *in vivo* MMAS with BCOP scores and mean adjusted BCOP scores (i.e., individual scores > 200 were adjusted to 200). Spearman's rank correlation coefficients and Pearson's correlation coefficients were calculated for each participating laboratory for all test substances and separately for water-soluble substances, surfactants, solids, solutions, and liquids. **Table 9-1** presents the correlation coefficients obtained for the different analyses. Mean opacity scores and mean permeability scores also were compared to *in vivo* MMAS scores; however, the results of these correlations are not included here.

9.1.2 Bruner et al. (1998)

Three variations of the original BCOP test method protocol were used in an attempt to optimize the assay for cosmetic formulations:

1. The exposure time was increased to 24 hours.
2. Test substances were applied and rinsed four times during the 24 hours of exposure.
3. Corneas were examined histologically.

Various cosmetic formulations were tested with different concentrations of organic acid in both water-in-oil and oil-in-water emulsions. The pH of the emulsion water phase was varied to test effects of pH on corneal injury. Effects of a metal oxide also were tested. The composition of the formulations was not revealed. *In vivo* rabbit eye test data were not reported; rather, human eye tolerance tests were performed for some formulations. Endpoints of the human studies were lacrimation, stinging, conjunctival redness, and conjunctiva and cornea fluorescein staining. The authors reported individual opacity and permeability results, and light microscopy data for the BCOP studies. Due to the lack of *in vivo* rabbit eye data and the fact that only mild formulations were evaluated, these data were not considered during the analysis of the performance of the BCOP (**Sections 6.0** and **7.0**).

Table 9-1 *In Vitro*/*In Vivo* Range of Correlations Reported in Balls et al. (1995)

Score index[1]	Range of Pearson's Correlation Coefficients[2]	Range of Spearman's Correlation Coefficients[2]
Full set of test substances (59)		
BCOP-Nonadjusted Scores	0.411 - 0.490	0.520 - 0.571
BCOP-Adjusted Scores	0.508 - 0.553	0.522 - 0.573
Substances soluble in water (30)		
BCOP-Nonadjusted Scores	0.477 - 0.625	0.326 - 0.448
BCOP-Adjusted Scores	0.446 - 0.554	0.326 - 0.448
Substances insoluble in water (18)		
BCOP-Nonadjusted Scores	0.160 - 0.336	0.581 - 0.690
BCOP-Adjusted Scores	0.359 - 0.446	0.582 - 0.690
Surfactants (12)		
BCOP-Nonadjusted Scores	0.772 - 0.895	0.685 - 0.825
BCOP-Adjusted Scores	0.772 – 0.895	0.685 - 0.825
Solids (20)		
BCOP-Nonadjusted Scores	-0.061 - 0.142	0.020 - 0.328
BCOP-Adjusted Scores	0.025 - 0.297	0.022 - 0.328
Solutions (14)		
BCOP-Nonadjusted Scores	0.586 - 0.771	0.546 - 0.689
BCOP-Adjusted Scores	0.558 - 0.775	0.543 - 0.693
Liquids (26)		
BCOP-Nonadjusted Scores	0.521 - 0.690	0.664 - 0.770
BCOP-Adjusted Scores	0.521 - 0.690	0.664 - 0.770

[1]Adjusted scores refers to the analysis in which individual BCOP scores > 200 were adjusted to 200. Balls et al. (1995) reports individual correlation coefficients for each laboratory.
[2]A correlation coefficient was calculated for each of the five participating laboratories; only the range of correlation coefficients obtained for the five laboratories is presented here.

9.1.3 Cassidy and Stanton (1997)

Six organosilicon compounds (siloxane polymers) were evaluated undiluted. The essential protocol components (e.g., preparation and treatment of corneas, opacity and permeability measurements) were the same as those used for Gautheron et al. (1994), except that the corneas were examined histologically. Five corneas were used per test substance, three corneas were used for an untreated control group, and two corneas were treated with a positive control (ethanol). The classification system is the same as that used for the Gautheron et al. (1994) study.

The test substances were hexamethyldisiloxane, polydimethylsiloxane, aminofunctional silicone A, aminofunctional silicone B, phenylsilsesquioxane fluid, and silicone ether. These substances are widely used in personal care formulations. The source and CASRNs of these compounds were not provided. Test substances were not coded.

The *in vivo* data were obtained according to OECD rabbit eye irritation testing guidelines, or their EU/EPA equivalent. However, only the *in vivo* irritancy grades are reported in the publication. Two nonirritants, two minimal irritants, one moderate irritant, one moderate to severe irritant, and one extremely severe irritant were evaluated, but it is not clear what *in vivo* ocular irritancy classification system was used for these classifications. NICEATM

contacted the study authors for the detailed *in vivo* data (i.e., raw scores for corneal opacity, iritis, conjunctival redness and chemosis for each animal) for this study; however, corporate clearance to release the data was not received. This study was not included in **Sections 3.0 – 7.0** of the BRD, because the raw *in vivo* scores for the rabbit studies, which are necessary to assign EPA (1996), EU (2001) and GHS (UN 2003) ocular irritancy classifications, could not be obtained.

Mean opacity and mean permeability values were reported, in addition to total BCOP scores (mean opacity value +15 x mean OD_{490} value). The BCOP scores were classified into the following three irritancy grades: nonirritant to mild (0-25), moderate (25.1-55), and severe (> 55).

For these six substances, the sensitivity and the specificity of the BCOP was 100%, using two classes of irritancy (nonirritant and irritant).

9.1.4 Chamberlain et al. (1997)

The eight laboratories that submitted BCOP data for the IRAG evaluation provided data on 242 substances encompassing a wide variety of chemical and product classes. These substances are summarized by laboratory in **Table 9-2** and represent the classifications reported in the IRAG evaluation study report. The specific substances tested were not provided in the IRAG evaluation. Neither were any physicochemical characteristics.

Table 9-2 Substances Tested in IRAG-Reviewed Studies

Laboratory Number	Number of Substances Tested	Substance Type or Class
Lab 3[1]	43	Full range of chemical classes; industrial raw materials and intermediates
Lab 4	12	Personal care products
Lab 5	21	Fragrance gels
Lab 6[2]	25	Surfactant-containing materials
Lab 7[3]	52	22 liquids; 22 solids; 8 surfactants
Lab 8	39	20 surfactants; 12 surfactant-based lotions; 7 shampoos
Lab 9	20	Industrial chemical intermediates
Lab 10	30	Miscellaneous organic chemicals from ECETOC (1992) database on eye irritants
Total:	242 discrete (not all unique) substances	

[1]Gautheron et al. (1992).
[2]From CTFA Phase III study (Getting et al. 1996).
[3]EC Interlaboratory study (Gautheron et al. 1994).

The protocols used by the different IRAG laboratories followed that used in Gautheron et al. (1994) with the following exceptions:

- The volume of test substance (both liquids and solids) applied to the cornea was reported as 0.5 mL or 0.75 mL.
- The exposure time of liquids varied (10, 30, or 60 minutes) depending on the laboratory.

- Some laboratories used positive controls (acetone in three laboratories for liquid test substances and imidazole in one laboratory for solids).
- Different laboratories used different numbers (3 to 6) of corneas per test substance.

Although these variations in BCOP protocols were described in Chamberlain et al. (1997), it was not noted which specific protocol was used by each of the eight laboratories.

Most submissions calculated a BCOP score that combined the opacity and permeability values using the same formula as the EC (Gautheron et al. 1994) and EC/HO (Balls et al. 1995) studies. However, one submission considered the opacity and permeability scores separately and assigned an *in vitro* irritation score (mild, moderate, severe) based on the greater of the two values. While the scoring procedure of the different laboratories was discussed, actual BCOP scores were not provided in the IRAG evaluation (Chamberlain et al. 1997).

The IRAG-reviewed studies used Pearson's correlation to compare the MAS of each laboratory's test substance set to the BCOP scores. The relationship between the BCOP scores and MAS for each laboratory test set was graphed on a scatterplot diagram and Pearson's correlation coefficients were determined. Pearson's correlation coefficients were calculated for some individual *in vivo* endpoints (e.g., cornea opacity, conjunctivae redness, conjunctivae discharge, swelling, days to recover, and iris); however, different *in vivo* endpoints were used for the correlation analysis of different laboratories.

The *in vivo* reference data used for the IRAG evaluation were submitted by each participating laboratory for the substances it had tested. Although the IRAG reviewers requested a description of the *in vivo* test method used by each laboratory, specific protocols or guidelines used to produce the *in vivo* eye irritation data were not discussed in the IRAG report (Chamberlain et al. 1997). *In vivo* MAS were used to produce scatterplots and perform the statistical analyses that compared the *in vivo* and *in vitro* data of each laboratory; however, only a range of MAS was reported for each laboratory.

BCOP data and results were presented in a way that maintained the confidentiality of the specific substances tested and the identity of the participating laboratories. Thus, neither original data nor individual BCOP scores were provided in the IRAG report. Instead, scores are graphically presented in scatterplots that compare the BCOP scores with the *in vivo* MAS of test materials for a specific laboratory. Pearson's correlation analysis was used to compare the MAS of each laboratory's test substance set to the BCOP scores. The data were analyzed according to guidelines developed by a separate IRAG working group (Scala and Springer 1997) for acceptance and evaluation of data submitted for comparing *in vitro* with *in vivo* data.

Pearson's correlation coefficients were calculated for each participating laboratory for the test substances evaluated by that laboratory. **Table 9-3** presents the correlation coefficients obtained for the different laboratories. Sensitivity, specificity, positive and negative predictivity, and false positive and negative rates were not determined or discussed. The

IRAG evaluation did not consider test method reliability in its assessment of the BCOP assay. It is not known whether the BCOP studies were conducted in compliance with GLP guidelines.

Table 9-3 Summary Evaluation of BCOP Data Submitted to IRAG

Laboratory (No. of Materials Tested)	Substance Type or Class	Range of *In Vivo* MAS	Pearson's Correlation (*r* value)
Lab 3 (43)	Full range of chemical classes; industrial raw materials and intermediates	0 - 81.5	0.72
Lab 4 (12)	Personal care products	1 - 28	0.78
Lab 5 (21)	Fragrance gels	17.9 - 40.0	0.35/0.31[1]
Lab 6 (25)	Surfactant-containing materials	0 - 40	0.79/0.79[2]
Lab 7 (52)	22 Liquids; 22 Solids; 8 Surfactants	0 - 84	0.66
Lab 8 (39)	20 Surfactants; 12 Surfactant-based lotions; 7 Shampoos	0 - 64	0.56
Lab 9 (20)	Industrial chemical intermediates	0 - 110	0.74
Lab 10 (30)	Miscellaneous organic substances from ECETOC (1992) database on eye irritants	1.67 - 108	0.55

[1]Pearson's correlation coefficients for BCOP scores at 10 minutes/30 minutes.
[2]Pearson's correlation coefficients for log BCOP/BCOP permeability only.

As described in the introduction to Section 9.1, the IRAG study was not included in previous sections of this document due to the observation that at least two of the IRAG data sets had been published in reports reviewed in previous sections of the BRD; the data and other information in the IRAG report were not included to avoid duplication of BCOP studies and data. Additionally, BCOP and detailed *in vivo* data, which are necessary for the analyses performed in **Section 6.0** of this document, were not received in response to NICEATM's requests for such data.

9.1.5 Cooper et al. (2001)

The BCOP assay was performed essentially as described by Gautheron et al. (1992), except that corneal swelling and histological evaluation were added as endpoints, and various exposure times and dilutions were used. Seven shampoo formulations of mild to extreme *in vivo* irritancy were evaluated. BCOP scores alone tended to underpredict the irritancy of the substances investigated; however, the authors noted that histological evaluation provided useful information.

In vivo data were not available for all substances. For some substances, modified Draize rabbit eye test data (MAS) were available. Mean opacity, mean permeability, and mean BCOP scores were reported. Additionally, corneal swelling percentages and results of the histological evaluation were reported. Since the *in vivo* test results were expressed as MAS, the data provided in this report could not be used to evaluate the accuracy of the BCOP for detecting ocular corrosives and severe irritants according to the GHS (UN 2003), EPA (1996), or EU (2001) classification systems. NICEATM contacted a representative from the corresponding author's organization for detailed *in vivo* data and was informed that these data were not readily available.

9.1.6 Gautheron et al. (1992)

This is the first publication on the BCOP assay. Many protocol components are the same as those for Gautheron et al. (1994); however, the protocol lacks some refinements used in the latter study, such as combining the opacity and permeability values into a total *in vitro* score, and assigning irritancy grades to test materials based on ranges of scores. For these reasons, the study was not included in the accuracy analyses (**Section 6.0**) of this document.

Forty-one liquids (e.g., alcohols, solvents, volatile organics, and other chemical classes with varying physicochemical characteristics) and six solids (acids, anionic surfactant, cationic surfactants) were tested for which chemical names are provided. Fifteen process intermediates also were tested but their structures/names were not provided.

The *in vivo* reference data used were from the published literature or from in-house studies. Data were standardized to four irritancy grades (mild, mild/moderate, moderate, and severe). Only opacity values were reported for the 47 reference substances; values were classified into four groups (mild: 0-20 opacity units; mild/moderate: 21-40; moderate: 41-70; severe: ≥ 71). Opacity and permeability values were reported for the 15 process intermediates. *In vitro* opacity grades were compared with *in vivo* irritancy grades for the 47 reference substances. There were six false negatives (SDS and some medium chain length alcohols). For opacity alone (44 substances), the Spearman's rank correlation coefficient was 0.73.

9.1.7 Gautheron et al. (1994)

The test method performance analyses conducted for this study are summarized here. An *in vitro/in vivo* comparison using BCOP and MAS scores was conducted as follows. Mean *in vitro* scores of the 52 test substances were compared first with *in vivo* MAS scores and then with day 1 scores using the Spearman rank correlation test. The correlation between BCOP and *in vivo* MAS scores was $r = 0.64$, while the correlation between BCOP and *in vivo* day 1 scores was $r = 0.73$. The authors decided to use *in vivo* day 1 scores for all further correlation calculations:

- *in vitro* opacity scores versus *in vivo* day 1 scores: $r = 0.67$
- *in vitro* permeability values versus *in vivo* day 1 scores: $r = 0.60$
- total BCOP scores for liquids plus surfactants versus *in vivo* day 1 scores: $r = 0.78$
- total BCOP scores for solids versus *in vivo* day 1 scores: $r = 0.62$

In vitro/in vivo comparison using irritancy categories: Substances tested *in vitro* were classified into two categories based on their *in vitro* score: irritant (score ≥ 25.1) and nonirritant (score ≤ 25.0). Substances categorized as nonirritant *in vivo* were those classified as practically nonirritant, minimally irritant, or mildly irritant with the Kay and Calandra system (Kay and Calandra 1962). Substances categorized as irritant *in vivo* were those classified as moderately irritant, severely irritant, and extremely irritant with the Kay and Calandra system. A two-by-two contingency table was constructed to determine concordance, sensitivity, and specificity values. The values for concordance, sensitivity, and specificity were the same, 85%, since the BCOP assay overpredicted and underpredicted four substances. The false positive rate was 15% (4/26 substances) and the false negative rate was 15% (4/26 substances). According to the study authors, the BCOP test method performed reasonably well at distinguishing irritating from nonirritating substances. **Table 9-4** provides a comparison of *in vivo* and *in vitro* data for irritants classified as severe or stronger in Gautheron et al. (1994) using either the Kay-Calandra (1962) or EEC (1984) classification systems.

Table 9-4 Comparison of *In Vivo* and *In Vitro* Data for Irritants Classified as Severe or Stronger in Gautheron et al. (1994) Using Either the Kay-Calandra (1962) or EEC (1984) Classification Systems

Substance Name (Physical Form)	*In Vivo*					*In Vitro*	
	MAS	Day 1 Score	Days to Reverse	K-C class[1]	EEC[2]	BCOP Score	BCOP Class[3]
Dibenzoyl-L-tartaric acid (S)	33.7	33.7	21	Mod	R41	120.5	Sev (I)
Sodium oxalate (S)	47.0	47.0	IRR	Sev	R36	4.8	Mild (NI)
Imidazole (S)	54.3	48.0	IRR	Sev	R36	87.9	Sev (I)
Quinacrine (S)	52.3	52.3	IRR	Extr	R36	31.1	Mod (I)
Hexadecyltrimethyl-ammonium bromide (SF)	69.0	49.7	IRR	Extr	R36	66.4	Sev (I)
Benzethonium chloride (SF)	76.3	67.0	IRR	Extr	R41	133.9	Sev (I)
Promethazine hydrochloride (S)	103.0	82.3	IRR	Max	R41	120.5	Sev (I)

Abbreviations: I = Irritant; IRR = Irreversible; MAS = Maximum average score; NI = Nonirritant; S = Solid; SF = Surfactant.

[1]Kay and Calandra (1962): Mod = Moderately irritant; Sev = Severely irritant; Extr = Extremely irritant; Max = Maximally irritant.

[2]EEC (1984) risk categories for ocular irritancy: R36 = Irritant; R41 = Severely irritant.

[3]BCOP data were grouped into three classes (Mild irritant = 0-25; Moderate irritant = 25.1-55; and ≥ 55.1) and two classes (Nonirritant [NI] = < 25.0 and Irritant [I] = ≥ 25.1).

Regarding interlaboratory reproducibility, this study found that 82.7% of the test substances were classified the same by all laboratories when using a three-category system. In this system, substances were classified into one of the following categories: mild irritant (BCOP score 0-25), moderate irritant (25.1-55) and severe irritant (≥ 55.1).

9.1.8 Jones et al. (2001)

The BCOP assay was performed as described by Gautheron et al. (1992), except that corneal swelling and histological evaluation were added as endpoints, and various exposure times and dilutions were used. Ten shampoos containing anionic or amphoteric surfactants and seven conditioner formulations containing cationic surfactants were evaluated. *In vivo* irritant categories (mild, moderate, substantial) were based on Draize scores and any other information that was available, such as market history. NICEATM contacted the corresponding author for detailed *in vivo* data and was informed that these data were not readily available. Mean opacity, mean permeability, and mean BCOP scores were reported. Additionally, corneal swelling percentages and some histological results were reported. BCOP classifications correlated poorly with the *in vivo* irritancy categories used for this set of substances. The assay overpredicted the irritancy of the conditioners, but could discriminate between shampoos with different *in vivo* irritancies.

9.1.9 Rachui et al. (1994)

The BCOP protocol used for this study was essentially the same as Gautheron et al. (1994). Thirty-eight cosmetic and personal care test materials obtained from Maybelline, Inc. were tested. Examples include creams, refresher sprays, oil sprays, lotions, shower gels, bath oils, eyeliners, mascara, and eye creams. Mean BCOP scores (opacity + 15 x O.D.) were reported and classified into 3 grades: mild (0-25); mild/moderate (25.1-55); and severe (≥ 55.1).

This study was not included in the accuracy analyses (**Section 6.0**) because the *in vivo* data were obtained from a modified Draize eye irritation protocol (i.e., 0.03 mL of test substance). Scores for 24, 48, and 72 hours were reported and irritant grades assigned; however, the *in vivo* ocular irritation classification scheme was not described. *In vivo* data were not available for all substances tested *in vitro*. For 32 substances, 24-, 48-, 72- hour scores are reported. Seventeen substances were classified as nonirritants, and two were classified as mild. The method of comparing *in vivo* and *in vitro* results is not clear or well-described. However, the study reports that BCOP grades correlated to available *in vivo* grades for 25 of 28 (89%) substances, without clearly explaining how these results were obtained or providing sufficient *in vivo* data to verify the accuracy calculation.

9.1.10 Rougier et al. (1994)

The essential protocol components (e.g., preparation and treatment of corneas, opacity and permeability measurements) were the same as those used by Gautheron et al. (1994). However, the authors did not note the number of corneas used per test substance, whether any controls were used, or whether the substances were tested undiluted or diluted. Spearman's correlation coefficients were calculated for BCOP scores and *in vivo* MAS for the 20 surfactants and the 21 cosmetic formulations. No other measures of accuracy were noted. An *in vitro* classification system was not provided

Twenty surfactants and 21 cosmetic formulations were evaluated. The surfactants were identified, and included nonionic, anionic, amphoteric, and cationic types. The types of cosmetic formulations included eye make-up remover, make-up remover, shampoos, and one shower gel. The components of the formulations were not provided. Seven of the

surfactants were purchased from Sigma; however, the sources of the other materials were not provided. CASRNs were not provided. Test substances were not coded.

Historical *in vivo* data from in-house Draize rabbit eye tests were used as reference data. MAS and the average score at seven days are reported for each substance in the publication. Irritancy grades were not provided. Detailed *in vivo* data were not obtained for this study, which prevented its inclusion in earlier sections of this document. NICEATM could not readily find current contact information for the study authors.

BCOP scores (opacity value +15 x O.D. value) were reported for all substances tested, but irritancy grades were not assigned. Spearman's correlation coefficients were calculated for BCOP scores and *in vivo* MAS for the set of 20 surfactants and the set of 21 cosmetic formulations. The Spearmann's correlation coefficient for BCOP scores and MAS for the 20 surfactants was 0.75. The Spearmann's correlation coefficient for BCOP scores and MAS for the 21 surfactant-based cosmetic formulations was 0.79.

The performance characteristics of the BCOP assay for all 41 substances using two classes of irritancy (nonirritant and irritant) was reported as: concordance = 93% (38/41 substances); sensitivity = 91% (20/22 substances); false negative rate = 9% (2/22 substances); specificity = 95% (18/19 substances); false positive rate = 5% (1/19 substances).

9.1.11 Sina et al. (1995)

The protocol was identical to Gautheron et al. (1994) except that 0.5 mL of test material was applied to corneas and exposure was for 30 minutes. Thirty-seven test substances representing a broad range of pH, solubility, and *in vivo* irritation potential were tested. Most substances were synthetic intermediates isolated during manufacture of pharmaceuticals. Chemical names, structures, and classes are not provided in paper.

Few details were provided on the conduct of the *in vivo* reference studies. However, MAS and Kay/Calandra classifications were reported for each substance. Mean BCOP scores (opacity + 15 x O.D.) were reported and classified into four grades: nonirritating/ mild (0-15); mild (> 15-25); moderate (> 25-55); severe (≥ 55.1). The correlation between BCOP and *in vivo* classes for 36 substances was 89%. Specificity (36) was reported as 90%. Sensitivity (36) was reported as 88%. The Spearman correlation coefficient for *in vitro* and *in vivo* scores (32 substances) was 0.74. The Pearson correlation coefficient for *in vitro* and *in vivo* scores (32 substances) was 0.62. NICEATM contacted Dr. Sina for additional data and information on the various studies he published; it was found that many of the *in vivo* studies were stored on microfiche in company archives, so additional data were not readily available.

9.1.12 Ubels et al. (1998)

This study investigated the effect of hydration on corneal opacity using the modified BCOP assay reported by Casterton et al. (1996). The authors note that corneal opacity can result from an increase in corneal hydration (i.e., corneal swelling or edema) or from damage to the cornea (e.g., precipitation of corneal proteins), and that it might be useful to distinguish between these two causes of opacity, since the former is sometimes reversible while the latter

is often irreversible. The study evaluated 10 substances previously studied by Casterton et al. (1996) that are known to cause opacity in the BCOP assay. Hydration measurements (i.e., comparison of wet and dry cornea weights), corneal light absorbance at 570 nm, and light and electron microscopy data were reported. The authors concluded that corneal hydration measurements would be a useful addition to the BCOP assay.

9.1.13 Ubels et al. (2000)

This study investigated the effect of reduced treatment times (30 seconds and 1 minute) on corneal opacity, permeability, and hydration using a modified BCOP assay (Casterton et al. 1996). Effects of irritants on the corneal endothelium were also examined. This study examined 13 substances previously studied by Casterton et al. (1996). Hydration measurements, corneal light absorbance at 570 nm, and electron microscopy data were reported. For most of the substances, the reduced treatment times resulted in lower corneal opacity and hydration values. The authors suggested that the shorter exposure times might provide results in the BCOP assay more predictive of human response to eye irritants. Based on the electron microscopy data, the authors also found that certain irritants damage the corneal endothelium. Some endothelial damage also was found for untreated corneas that had simply been mounted in the BCOP corneal holder, suggesting the need for optimization of the corneal holder.

9.1.14 Ubels et al. (2002)

This study represents a continuation of the work reported in Ubels et al. (2000). It describes the design and use of a redesigned corneal holder. The authors note potential limitations of the conventional corneal holder: 1) it has a circular opening 17 mm in diameter, yet the bovine cornea is oval shaped and has dimensions of about 24 mm vertically and 30 mm horizontally; 2) it has flat inner surfaces, whereas the bovine cornea is convex or curved. These elements of the corneal holder reportedly force the bovine cornea into an unnatural shape when mounted in the holder, causing the cornea to wrinkle. The authors also noted damage to all three corneal cell layers (epithelium, stroma, and endothelium) where the cornea comes in contact with the circular edge of the holder opening.

Recognizing some of the potential limitations of the conventional corneal holder, the authors designed a new corneal holder with dimensions that better fit the bovine cornea and maintain its natural shape during the BCOP assay. The new holder was designed to contact the 2 to 3 mm rim of sclera left around the bovine cornea during dissection, rather than the corneal tissue. The authors report that this refined corneal holder does not cause wrinkling of the mounted bovine cornea, nor does it damage the cell layers around the edge of the cornea.

The following test substances were studied in this evaluation: acetone, 1% benzalkonium chloride, isopropanol, and 30% trichloroacetic acid. Hydration measurements, corneal light absorbance at 570 nm, and electron microscopy data were reported.

9.1.15 Ubels et al. (2004)

This study is a continuation of the authors' evaluation of the utility of a redesigned corneal holder for use in the BCOP assay. Previous studies have suggested that the new holder is an improvement over the conventional holder based on comparisons of corneal opacity,

hydration, and endothelial morphology (Ubels et al., 2000, 2002). This study provides a comparison between the conventional holder and the redesigned holder with respect to corneal permeability. The effects of acetone, isopropyl alcohol, 1% sodium hydroxide, 30% trichloroacetic acid, and 30% sodium dodecylsulfate on corneal permeability were compared between the two corneal holders. The authors contend that the lack of damage seen with the redesigned holder (as opposed to the damage to the cornea reportedly induced by the conventional holder) reduces the level of permeability, as well as reducing measurement variability.

9.1.16 Vanparys et al. (1993)

The essential protocol components (e.g., preparation and treatment of corneas, opacity and permeability measurements) were the same as those used for Gautheron et al. (1994). Six corneas were used per test substance and three corneas were used for an untreated control group. The classification system differed slightly from Gautheron et al. (1994) in that: nonirritant = BCOP score of 0 to 3; mild irritant = 3.1 to 25; moderate irritant = 25.1 to 55; and severe irritant > 55. Concordance, specificity, and sensitivity were calculated for two scenarios: 1) the *in vivo* and *in vitro* irritancy grades were divided into two groups; 2) the irritancy grades were divided into four groups.

Fifty pharmaceutical and commercially available substances were evaluated representing both liquids (miscible and immiscible) and solids (soluble and insoluble). Examples include piperidines, epoxides, furans, thiazoles, nitrophenyls, imidazole, Tween 20 & 80, shampoos, and alcohols. Nine of the substances were in-house compounds (i.e., candidate drugs) and 15 were pharmaceutical process intermediates. Chemical names and physical state are provided in the publication. Test substances were not coded. CASRNs and the source of materials were not provided.

For the in-house substances and the pharmaceutical intermediates, historical *in vivo* data from the Draize test were available at Janssen Pharmaceutica. For the commercially available substances, *in vivo* data were obtained from the literature or from Draize tests (OECD 1987) performed at J. Simon Laboratories. The *in vivo* ocular irritancy grades of the test substances were nonirritant (13 materials), mild (6), mild/moderate (2), moderate (10), and severe (19) based on an internal classification scheme (not an accepted regulatory classification system). The only *in vivo* data in the publication were these irritancy grades.

Mean opacity and mean permeability values were reported, in addition to total BCOP scores (mean opacity value +15 x mean O.D. value). The BCOP scores were classified into the following four irritancy grades: nonirritant (0 to 3), mild (3.1 to 25), moderate 25.1 to 55), and severe (> 55). Concordance, specificity and sensitivity were calculated for two scenarios: 1) *in vivo* and *in vitro* irritancy grades were divided into two groups; and, 2) irritancy grades were divided in four groups

When *in vivo* and *in vitro* irritancy grades were grouped into two categories (negative = nonirritant and mild irritants; and positive = moderate and severe irritants), the concordance was 96% (48/50 substances), specificity was 95% (18/19 substances), and sensitivity was

97% (30/31 substances). The false positive rate was 5% (1/19 substances) and the false negative rate was 3% (1/31 substances).

When four *in vivo* and *in vitro* eye irritancy grades (nonirritant, mild, moderate, and severe) were used, 36 of 50 (72%) *in vivo* grades were correctly predicted with the BCOP assay. Twelve (24%) substances (alcohols and other highly permeable substances) were overpredicted in the BCOP assay, while two (4%) solids were underpredicted.

Of the 19 substances classified as severe irritants *in vivo* by the investigators, the BCOP assay correctly predicted all 19 as severe irritants.

9.1.17 1997 Bovine Corneal Opacity and Permeability Technical Workshop

In November 1997, the Institute for *In Vitro* Sciences (IIVS) convened a workshop that addressed the state-of-the-art of the BCOP assay with a focus on how it met certain regulatory acceptance criteria. The proceedings of this workshop were published in 1998 (In Vitro & Molecular Toxicology 11(4):315 to 351) in an article entitled "Report from the Bovine Corneal Opacity and Permeability Technical Workshop – November 3-4, 1997." This report summarizes the talks and discussions of the workshop, which included:

- An Historical Perspective (summarized by J.F. Sina and P. Gautheron)
- The Bovine Corneal Opacity and Permeability Assay: An Alternative Protocol (summarized by P. Casterton)
- Considerations for Histological Examination of Bovine Corneal Tissue (summarized by M.G. Evans)
- The Bovine Corneal Opacity and Permeability Assay: Observations on Assay Performance (summarized by J.W. Harbell and R.D. Curren)
- Experience with Other Isolated Eye Models: Isolated Rabbit Eye (IRE) (summarized by L. Earl)
- The Use of Prediction Models with Non-Animal Eye Irritation Tests (summarized by L. Bruner)
- Workshop Summary (summarized by R.D. Curren and J.W. Harbell)

In the Workshop Summary, Drs. Curren and Harbell addressed several of the criteria used by ICCVAM to assess the validation status of an alternative test method. The BCOP assay was discussed in terms of its scientific and regulatory rationale, the relationship of the test method endpoints to the biologic effect of interest, available protocols, extent of intra- and inter-laboratory variability, test method performance using reference chemicals (prediction models), and assay constraints.

Regarding the discussion of available protocols, the authors noted that the original test method protocol was designed to assess the potential eye irritation of pharmaceutical intermediates. As use of the assay spread to different laboratories and testing of different types of materials, the protocol changed to accommodate the different physical and chemical characteristics of different test substances. Certain aspects of the protocol, such as exposure and post-exposure times, can vary depending on the test material or objective of the study. The authors concluded that it is very likely that no single exposure protocol and prediction

model could provide accurate prediction of ocular irritation across all chemical classes and physical forms of test substances.

The authors also noted that histopathological evaluation of the corneas appears to be very useful; however, further development and refinement of this procedure was recommended at the time of the publication. Histology allows for an evaluation of the depth and type of injury, which could be used to evaluate the potential for recovery.

Regarding variability in the BCOP assay, the authors noted that reproducibility within and among laboratories appeared to be acceptable based on a number of in-house evaluations and multinational studies. Proposed sources of variability include variations in technical approach and potential differences in the corneas related to their source.

Constraints of the assay also were discussed. The authors noted that some laboratories have reported a decline in quality of the bovine corneas obtained during the summer months. The thicker epithelial layer of the bovine cornea, in comparison with human and rabbit corneas, was noted as a possible constraint that could potentially lead to an underestimation of irritancy for some substances. Also, the authors noted limitations of the currently used opacitometer, which provides a center-weighted reading of corneal opacity; they recommended development of a more accurate device for measurement of corneal opacity that could account for opacity over the whole surface of the cornea.

9.1.18 Review Articles on the BCOP Assay

Sina (1994) reviewed the steps taken by Merck Research Laboratories (West Point, Pennsylvania) to validate the BCOP assay for the purpose of screening chemicals (e.g., pharmaceutical intermediates and raw materials) to which workers would be exposed in a pharmaceutical manufacturing setting. The author discussed the initial development of the BCOP assay for this purpose, the results of an interlaboratory evaluation, and how results from the BCOP assay compare to other alternative eye irritation test methods.

Sina and Gautheron (1994) reviewed their experiences with developing a test battery to evaluate ocular irritation of substances. The BCOP assay, three cytotoxicity assays, and a few inflammation assays (e.g., chemotaxis, arachidonic acid cascade) were evaluated. In a study of 43 in-house materials representing a variety of chemical classes (aromatic and organic acids and bases, alcohols, esters, peptides, inorganic salts), the authors found that the accuracy of the BCOP *In Vitro* Irritancy Score in predicting Kay-Calandra class was greater than 80%. However, two of the false negatives in the BCOP assay resulted from substances that produced no irritation in the rabbit eye test until after 48 hours. The cytotoxicity assays did not perform very well across a range of chemical classes. The authors noted that the inflammation assays were still under development.

9.1.19 Poster Presentations

Over the past five years, numerous poster presentations have been given on the BCOP assay, which depict the ways in which the assay has evolved in recent years. Although it is not possible to summarize all of these presentations here, they are listed below by year of presentation to show that the assay has been applied to many different types of substances

(e.g., alkaline dry detergent products, hypochlorite-containing solutions, fragranced formulations, oxidizing/reactive cleaning products, petrochemical products, and fragrance mixtures). In studying different types of substances with the BCOP assay, optimal exposure and post-exposure times have been defined for certain types of substances. For example, a protocol using a 25% (v/v) aqueous dilution and 30-minute exposure was recommended for the surfactant-based products tested by Cater et al. (2001). Some of these posters (e.g. Curren et al. 2000a, 2000b) also demonstrate the usefulness of adding histopathological assessment to the BCOP assay. Another significant use of the BCOP assay has been to compare results of a product series with a selected "benchmark" that had been previously tested *in vivo* and had a well-established market history (e.g., Cater et al. 2001). A majority of the poster presentations can be obtained from the Institute for *In Vitro* Sciences (Gaithersburg, Maryland; website: http://www.iivs.org/).

2000

Curren R, Evans M, Raabe H, Ruppalt R, Harbell J. 2000a. Correlation of histopathology, opacity, and permeability of bovine corneas exposed *in vitro* to known ocular irritants. Veterinary Pathology 37(5):557.

Curren RD, Evans MG, Raabe HA, Ruppalt RR, Harbell JW. 2000b. An histopathological analysis of damage to bovine corneas *in vitro* by selected ocular toxicants. Presented at the 2000 Society of Toxicology meeting.

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.
(*Note:* S.C. Johnson & Son, Inc. submitted *in vitro* and *in vivo* data to NICEATM for this poster. This study was included in **Sections 3.0 – 7.0** of this document.)

2001

Cater KC, Raabe HA, Mun G, Harbell JW. 2001. Corporate validation program for predicting eye irritation of surfactant formulations *in vitro*. The Toxicologist 60:99.

Rees WM, Swanson JE, Burdick JD, Hilgers DS, Harbell JW. 2001. Evaluating toxic synergism in hypochlorite-containing solutions using the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 60:99.

2002

Burdick JD, Merrill JC, Spangler TC, Moyer GO, Harbell JW. 2002. Use of histological examination in bovine corneal opacity and permeability (BCOP) assay for assessing the ocular irritation potential of fragranced formulations. The Toxicologist 66:244.

Cater K, Nusair T, Merrill JC, Harbell JW. 2002. Exploratory *in vitro* eye irritation study of marketed alkaline laundry detergents by BCOP assay and pH/reserve alkalinity (RA) parameters. The Toxicologist 66:244.

Cuellar N, Merrill JC, Clear ML, Mun G, Harbell JW. 2002. The application of benchmarks for the evaluation of the potential ocular irritancy of aerosol fragrances. The Toxicologist 66(1-S):243-244.
(*Note:* S.C. Johnson & Son, Inc. submitted *in vitro* data and other information to NICEATM for this poster. See **Section 9.2** for a summary of this poster and **Appendix G** for the submitted information.)

2003

Cater K, Mun G, Moyer G, Merrill J, Harbell JW. 2003. Exploratory *in vitro* eye irritation study of marketed alkaline dry laundry detergents by BCOP assay and pH/reserve alkalinity (RA) parameters. The Toxicologist 72:220.

Cuellar N, Lloyd PH, Swanson JE, Merrill JC, Clear ML, Mun G, Harbell JW, Bonnette KL. 2003. Evaluating the eye irritancy of solvents in a simple fragrance mixture with the bovine corneal opacity and permeability assay. The Toxicologist 72:312.

Gran BP, Swanson JE, Merrill JC, Harbell JW. 2003. Evaluating the irritancy potential of sodium percarbonate: a case study using the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 72:220.
(*Note:* S.C. Johnson & Son, Inc. submitted *in vitro* data and histology figures to NICEATM for this poster. See **Section 9.2** for a summary of this poster and **Appendix G** for the submitted information.)

Swanson JE, White BT, Gran BP, Merrill JC, Harbell JW. 2003. Evaluating oxidizing/reactive cleaning products in the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 72:220-221.

2004

Bailey P, Freeman J, Phillips R, Merrill J. 2004. Evaluation of the BCOP assay as a predictor of ocular irritation of petrochemical products. Presented at the 2004 Society of Toxicology meeting.
(*Note:* ExxonMobil Biomedical Sciences, Inc. submitted *in vitro* and *in vivo* data to NICEATM for this poster. This study was included in **Sections 3.0 – 7.0** of this document.)

Cater K, Patrick E, Harbell J, Schilcher S. 2004. Comparison of *in vitro* eye irritation potential by BCOP assay to erythema scores in human eye sting test of surfactant-based formulations. Presented at the 2004 Society of Toxicology meeting.

Cuellar N, Lloyd PH, Swanson JE, Merrill JC, Mun G, Harbell JW, Bonnette KL. 2004. Phase Two: Evaluating the eye irritancy of solvents in a simple fragrance mixture with the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 78(S-1):Abstract No. 1306.
(*Note:* S.C. Johnson & Son, Inc. submitted *in vitro* data and histology figures to NICEATM for this poster. See **Section 9.2** for a summary of this poster and **Appendix G** for the submitted information.)

Swanson JE, Rees WM, Hilgers DS, Merrill JC, Harbell JW. 2004. Managing toxic synergism in hypochlorite-containing cleaners using the bovine corneal opacity and permeability (BCOP) assay. Part II. Presented at the 2004 Society of Toxicology meeting.

9.2 Other Scientific Reports Received in Response to a *Federal Register* Notice

In addition to the BCOP studies identified from the literature search, several studies were obtained in response to two *FR* Notices (Vol. 69, No. 57, pp. 13859-13861, March 24, 2004, and Vol. 70, No. 38, pp. 9661-9662; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original BCOP test method data and *in vivo* reference data. In response to these requests, *in vitro* test method data were submitted by Johnson & Johnson Pharmaceutical Research & Development, L'OREAL, and S.C. Johnson & Son, Inc. In these three reports, insufficient *in vivo* reference data precluded their use in an assessment of the performance characteristics of BCOP compared to the GHS (UN 2003), EPA (1996) and EU (2001) ocular irritancy classification systems. IIVS submitted replicate experiment data for the BCOP results reported in Gettings et al. (1996); these data were used for an analysis of intralaboratory reproducibility in **Section 7.0**. IIVS also submitted additional analyses of the *in vivo* and BCOP data reported in Gettings et al. (1996). Johnson & Johnson Pharmaceutical Research & Development submitted data for 20 chemicals tested in the BCOP assay, comparing corneas from adult animals (> 24 months) to those of young animals (6 to 8 months). These data were provided to evaluate the impact of age of source animals for test eyes on the BCOP assay. Details of these studies are included below.

9.2.1 S.C. Johnson & Son, Inc.

In addition to two datasets included in the accuracy and reliability analyses of this document, S.C. Johnson & Son, Inc. submitted three other datasets on the BCOP assay:

1. an evaluation of the potential ocular irritancy of aerosol fragrance formulations with the BCOP assay
2. the application of benchmarks for evaluation of the ocular irritancy of solvents in a simple fragrance mixture
3. an evaluation of reactive chemistry formulations using the BCOP assay

These three datasets are provided in **Appendix G,** and briefly summarized here.

The first dataset (**Appendix G1**) provides data and supporting information for a poster presentation given by Cuellar et al. (2004) on use of the BCOP assay to study the influence of solvents on the ocular irritation potential of fragrance mixtures. The study evaluated one fragrance, six solvents, and six solvent/fragrance mixtures. In this study, the protocol was modified in the following ways: exposure times of one and three minutes were used to evaluate the test substances; and, post-exposure times of 2-, 4- and 20-hours were used for different aspects of the study. In addition, a histopathological evaluation was performed on the treated corneas. A modified rabbit eye irritation test was conducted on the same substances tested *in vitro*. Four animals were treated per test substance. After 24 hours, ocular tissues were harvested for three animals for a histopathological assessment. The remaining animal was examined up to 28 days to evaluate recovery of ocular lesions.

The authors concluded that the choice of solvent can have a major influence on the ocular irritation potential of fragrance mixtures. Some solvents in a simple fragrance produce mild irritation, while other different solvents can produce severe irritation. The authors noted that the time course of tissue scores *in vivo* was similar to the time course of the histological changes in BCOP. It was also noted that morphological changes in the keratocytes were found in both the isolated bovine corneas and the rabbit eye treated corneas.

The second dataset (**Appendix G2**) provides data and supporting information for a poster presentation given by Cuellar et al. (2002) demonstrating how the BCOP assay can be used to evaluate new formulations in relation to an appropriate reference benchmark for which the ocular irritation potential is well-characterized. This study evaluated specific aerosol formulations in comparison to ethanol/fragrance benchmarks.

The third dataset (**Appendix G3**) provides data and supporting information for a poster presentation given by Gran et al. (2003) that described use of the BCOP assay to evaluate the potential eye irritancy of sodium percarbonate, a commonly used substance in cleaning products. Sodium percarbonate is highly reactive, producing corneal epithelial peeling and other types of irritation in the rabbit eye test. The standard BCOP protocol for solids was not used in this investigation of sodium percarbonate. Based on past experience with the BCOP assay, the eye irritancy potential of more aqueous-soluble solids such as laundry powders using the standard solids protocol is vastly overpredictive of the outcome resulting from accidental human exposure. Furthermore, experience has shown that reactive/oxidizing chemistries (such as bleach, percarbonates and peroxides) have a delayed toxicity response in the assay necessitating increased post-exposure observation time.

The question the investigators faced in this case study of sodium percarbonate was what protocol parameters were needed to model the bolus exposure for an extended period that occurs in the Draize eye irritation protocol, as well as what might be expected to be a realistic maximum exposure in humans. The following parameters were chosen: a 50% suspension of the solid with a 30-minute exposure time to model the *in vivo* exposure and 10-minute exposure time to model maximum accidental human exposure. While post-exposure time in the BCOP assay is typically two hours, times of four hours and 20-24 hours were chosen for this study.

Using the protocol considerations discussed above, the BCOP assay was able to adequately predict the irritancy potential of two different concentrations of sodium percarbonate for both a realistic human exposure scenario and an *in vivo* exposure scenario. Reduction of sodium percarbonate concentration predictably reduced the irritancy potential of the end-use formulation. Histology as a third endpoint in the BCOP assay was critical in evaluating the depth and degree of injury.

9.2.2 L'OREAL

L'OREAL Advanced Research provided a dataset for an in-house porcine corneal opacity and permeability (PCOP) assay, as well as some data from the BCOP assay. The dataset includes PCOP data and *in vivo* MAS scores for 50 liquid and water-soluble compounds, and data from both the PCOP and BCOP assays for 23 substances for which there was historical

in vivo data in the form of MAS scores. The authors note that the PCOP protocol is essentially that described by Gautheron et al. (1992), with the exception of some changes related to using a different species. Detailed *in vivo* data were requested from the submitters, but they indicated that they did not have the individual irritation scores for individual animals. This data submission is provided in **Appendix G4**.

9.2.3 IIVS

Dr. John Harbell submitted supplementary analyses for the BCOP study conducted for the CTFA Phase III evaluation of surfactant-based personal care products (Gettings et al. 1996). Dr. Harbell performed a bootstrap analysis of the *in vivo* rabbit eye studies performed for this evaluation, and compared the results of the bootstrap analysis to the permeability values obtained for the 25 test substances. Six rabbits were used in the *in vivo* eye irritation studies performed for the CTFA evaluation. However, a three rabbit eye irritation protocol is now accepted for use by the OECD and the EPA. Thus, the bootstrap analysis involved determining all of the possible three-animal combinations that result from the six animal test, assessing what the classification of the study would be according to the GHS system for each three animal combination, then determining the percentage of agreement among the 20 different possible three-animal combinations. For highly irritating substances and substances that were nonirritating, the extent of agreement among the 20 combinations was high. However, for substances that produced irritation in between these two extremes, the extent of agreement was more variable.

Several graphical representations of the *in vivo* data generated for the CTFA Phase III study were provided to include the average opacity score, average iris score, average redness score, and average chemosis score obtained for each substance in the rabbit eye test. The variability of these scores for each test substance was also depicted on the graphs.

Finally, there are three graphs that show the permeability values obtained for each test substance versus the results of the bootstrap analysis discussed above.

The bootstrap analysis and graphs are provided in **Appendix G5**.

9.2.4 Johnson & Johnson Pharmaceutical Research & Development

BCOP results from tests conducted with 20 substances using corneas from adult animals (i.e., > 24 months) and young animals (i.e., 6 to 8 months) were provided, along with the reported EU and GHS classification for each substance (**Table 9-5**). The submitters state that one of the test substances (acetone) needs to be repeated, due to discordant results with this substance relative to an earlier study. For this reason, only 19 of the 20 test substances were considered in the evaluation below. Corneas (3/test substances) were treated for 10 minutes followed by a 120-minute recovery period. Medium was removed from the anterior compartment and replaced by 1 ml of a 0.4% sodium-fluorescein solution. Corneas were incubated in a horizontal position for 90 minutes at 32°C in a water-bath. After incubation, medium from the posterior chamber was removed and its optical density (OD) determined with a spectrophotometer at 490 nm, and the IVIS calculated. Experiments with corneas from young animals were performed with a specially designed cornea holder, which has a smaller diameter than the traditional holder.

Based on the data summarized in **Table 9-5** (and excluding acetone, as indicated above), and regardless of which *in vivo* classification was used (i.e., EU or GHS), the overall accuracy of the BCOP using eyes from adult animals (> 24 months) was 68%, with false positive and false negative rates of 31% and 33%, respectively. By comparison, the overall accuracy of the BCOP using eyes from young animals (6-8 months) was 74%, with false positive and false negative rates of 19% and 67%, respectively. These results provide evidence that the performance of the BCOP using eyes from younger animals may not be significantly different than using eyes from older animals. However, given their smaller size relative to those from adult animals, from which corneas are typically obtained for the BCOP, a specially designed corneal holder (i.e., smaller diameter) is required for using younger corneas.

Table 9-5 Substances Used to Evaluate the Use of Corneas from Animals of Different Ages in the BCOP Assay

Test Substance	CASRN	*In Vivo* (EU)[1]	*In Vivo* (GHS)[1]	BCOP (> 24 months)				BCOP (6-8 months)			
				Opacity	Perm.	IVS	Class	Opacity	Perm.	IVS	Class
3,3-Dimethylpentane	562-49-2	NI	NI	0.6	0.01	0.8	NON	0.0	0.02	0.3	NON
3-Methoxy-1,2-propanediol	623-39-2	NI	NI	-0.3	0.0	0.2	NON	0.6	0.02	0.9	NON
Polyethylene glycol 400	25322-68-3	NI	NI	-0.3	0.0	-0.3	NON	0.0	0.08	1.1	NON
Glycerol	56-81-5	NI	NI	-1.0	0.01	-0.9	NON	-0.7	-0.01	-0.8	NON
Methyl cyclopentane	96-37-7	NI	NI	1.0	0.43	7.5	MILD	1.3	0.26	5.2	MILD
Tween 20	9005-64-5	NI	NI	0.0	0.01	0.1	NON	0.0	-0.01	-0.1	NON
Methyl iso-butyl ketone	108-10-1	NI	NI	6.6	1.07	22.7	MILD	5.7	0.83	18.1	MILD
Toluene	108-88-3	NI	NI	6.3	3.18	54	MOD	6.0	1.46	28.0	MOD
Methyl amyl ketone	110-43-0	NI	NI	5.3	1.8	32.3	MOD	4.0	0.99	18.8	MILD
2-Methyl-1-pentanol	105-30-6	NI	2B	12.0	4.3	76.6	SEV	8.6	1.94	37.7	MOD
Ethanol	64-17-5	NI	2B	16.0	2.34	51	MOD	16.3	1.83	43.8	MOD
Sodium hydroxide (1%)	1310-73-2	R36	2B	99.7	4.16	162	SEV	135.7	3.74	191.8	SEV
Triton X-100 (5%)	9002-93-1	R36	2B	4.3	3.81	61.5	SEV	4.7	3.7	60.1	SEV
1-Octanol	111-87-5	R36	2B	10.0	5.24	88.6	SEV	10.3	1.53	33.3	MOD
2-Ethyl-1-hexanol	104-76-7	R36	2B	4.3	1.76	30.6	MOD	2.3	0.86	15.3	MILD
n-Hexanol	111-27-3	R36	2A	15.3	3.73	71.2	SEV	14.0	3.62	68.2	SEV
Acetone[2]	67-64-1	R36	2A	39	2.95	83.2	SEV	91.3	2.86	134.2	SEV
Cyclohexanol	108-93-0	R41	1	15.3	5.04	90.7	SEV	11.6	2.13	43.6	MOD
Cetylpyridinium bromide (6%)	140-72-7	R41	1	11.7	1.01	26.8	MOD	15.0	1.66	39.9	MOD
Benzalkonium chloride (10%)	8001-54-5	R41	1	92.2	4.22	155.4	SEV	105.7	4.05	166.5	SEV

[1]*In vivo* classification provided by Johnson & Johnson Pharmaceutical Research and Development

[2]Data excluded due to reported technical difficulties with this substance, which requires retesting (no data received from retest)

CASRN=Chemical Abstracts Service Registry Number; IVS=*In vitro* score; MILD=Mild irritant (IVS=3.1-25); MOD=moderate irritant (IVS=25.1-55); NI=Nonirritant; NON=Nonirritant (IVS ≤ 3); Perm.=Permeability; SEV=Severe irritant (IVS > 55.1)

[This Page Intentionally Left Blank]

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)

With respect to these animal welfare considerations, the BCOP assay both refines and reduces the use of laboratory animals bred specifically for the purpose of toxicity testing. This assay uses isolated corneas from cattle that have been slaughtered for the food industry or for other nonlaboratory purposes. Since isolated tissues are treated in the assay, treatment-related pain and suffering are avoided in live animals. By using slaughterhouse by-products, the BCOP assay also reduces the use of laboratory animals (i.e., substances that are identified as ocular corrosives or severe irritants *in vitro* would be excluded from testing *in vivo*).

10.2 Requirement for the Use of Animals

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

[This Page Intentionally Left Blank]

11.0 PRACTICAL CONSIDERATIONS

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

11.1 Transferability of the BCOP Test Method

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

11.1.1 Facilities and Major Fixed Equipment

The facility requirements necessary to conduct the BCOP test method include a standard laboratory setup for nonsterile tissue culture, and proximity to an abattoir or other bovine eye supplier so that eyes can be obtained from freshly slaughtered animals. The major equipment necessary to conduct the test method is readily available and includes an opacitometer, cornea holders, a UV/VIS spectrophotometer or microplate reader, and a water bath or forced air incubator. Suppliers and estimated costs of this equipment are summarized in **Table 11-1** to the extent this information was available.

If histopathology is included as a component of the BCOP test method, the testing facility may choose to process the tissue in-house, whereby the facility would need equipment for tissue processing, sectioning, and staining. This specialized equipment would add significant cost to the major equipment required for the BCOP assay. Most likely, if a facility is not already equipped to prepare tissue slides for histological evaluation, this portion of the test method could be outsourced to an appropriate contractor as is done at IIVS (Harbell J, personal communication).

In contrast, the *in vivo* rabbit eye test requires a facility that meets the approval of applicable State and Federal regulations to house live laboratory animals. The primary expense for equipping a facility to conduct the *in vivo* rabbit test would be the acquisition of an adequate animal room and associated housing (e.g., cages, bedding, food, water, etc.) for boarding animals during the study.

11.1.2 General Availability of Other Necessary Equipment and Supplies

The remaining equipment and supplies necessary to conduct the BCOP test method (e.g., dissection equipment, micropipettors, petri dishes, volumetric flasks) are readily available in most scientific laboratories or can be obtained from any of several scientific laboratory equipment vendors.

Table 11-1 Suppliers and Costs of Major Equipment for the BCOP Assay

Equipment	Supplier/Manufacturer(s)	Estimated Costs[1]
Opacitometer	Stag Bio (Clermont, France) Spectro Design (Riom, France)	***Not yet provided***
Cornea holders	Stag Bio (Clermont, France) Spectro Design (Riom, France)	***Not yet provided***
UV/VIS Spectrophotometer	Beckman, Fisher Scientific, Perkin Elmer, Thermo Spectronic	Starting at $6000 for an 8-cell holder unit with a spectral range from 200 -1100 nm
Microplate reader	Bio-Rad, Bio-Tek Instruments, Cambrex, Fisher Scientific, Molecular Devices, PerkinElmer	Starting at $6500 for a 96-well plate absorbance reader with a spectral range from 400 -750 nm
Water bath	e.g., Brinkmann Instruments, Fisher Scientific, Neslab	2 L capacity ~ $590 5 L capacity ~ $750 10 L capacity ~ $875 20 L capacity ~ $1100
Incubator - forced air	e.g., Fisher Scientific, Precision, Thermo Electron	2.2 cu. ft. ~ $2100 3.75 cu. ft. ~ $2230 5.0 cu. ft. ~ $2460

[1]Estimated costs based on 2004 catalog prices.

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

11.2 BCOP Test Method Training Considerations

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

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

A training period of between two to three months is usually required for a technician with general laboratory skills to conduct all aspects of the assay independently and proficiently (Harbell J, personal communication). A training video or other visual media to provide guidance on performing the assay may be considered useful. During the training period the technician would learn how to:

- dissect the cornea from the bovine eye
- identify corneas with defects
- mount the cornea in a corneal holder without damaging the epithelium or endothelium

- add assay medium and test substances to the appropriate chamber of the corneal holder
- properly time and conduct incubations
- calibrate and use the opacitometer
- rinse the cornea without damaging it
- perform quantitative opacity readings
- conduct the permeability evaluation

There are currently no known proficiency criteria used to ensure that personnel are performing the test method competently. Rather, this must be demonstrated through experience with the oversight of an experienced supervisor. All of the tasks in the BCOP assay are technically simple to perform. When a technician has mastered all aspects of the protocol, and can independently conduct the assay, such that the positive control falls within its historical range, the technician has essentially demonstrated proficiency in the assay.

In contrast, the *in vivo* test method requires training in the care and handling of laboratory animals. Possibly the most difficult aspect of the *in vivo* test method to master is the subjective assessment of corneal opacity, iritis, conjunctival chemosis, conjunctival redness, and discharge as endpoints in the evaluation of ocular irritancy. The laboratory personnel must be adequately trained to accurately and consistently identify these endpoints. It is not known what, if any, proficiency requirements are in place for the *in vivo* test.

11.3 Cost Considerations

The current cost for a GLP compliant BCOP assay at IIVS is $1400 per test substance (Harbell J, personal communication). This cost includes both positive and negative controls. Histology can be added to the standard BCOP assay for $650 and includes both slides and photographs. Another source reports that a typical GLP compliant BCOP study for one sample with benchmarks and histology costs about $4,500, and this includes two time courses and one benchmark (Cuellar N and Swanson J, personal communication). In comparison, a GLP-compliant EPA OPPTS Series 870 Acute Eye Irritation test (EPA 1998) in the rabbit ranges from $765 for a 3 day/3 animal study up to $1665 for a 21 day/3 animal study at MB Research Laboratories (MB Research laboratories, personal communication). While the cost of the BCOP assay includes concurrent positive controls, the *in vivo* rabbit test method does not include equivalent controls.

11.4 Time Considerations

Use of the BCOP test method would significantly reduce the time needed to assess the ability of a test substance to induce ocular corrosivity or severe irritancy, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days. Depending upon the severity of ocular effects produced by a test substance, the method can be extended for up to 21 days. Comparatively, the standard BCOP test method can be completed in about five hours for liquid test substances and seven hours for solid test substances, once the bovine eyes arrive at the testing facility. However, one source notes that the turnaround time from initiation of the

study to receipt of the final report is similar for the BCOP assay and the *in vivo* rabbit eye test (Cuellar N and Swanson J, personal communication).

12.0 REFERENCES

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

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.

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

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

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

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

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

Bruner LH, de Silva, O, Earl LK, Easty DL, Pape W, Spielmann H. 1998. Report on the COLIPA Workshop on Mechanisms of Eye Irritation. ATLA 26:811-820).

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

Burdick JK, Merrill JC, Spangler TC, Moyer GO, Harbell JW. 2002. Use of histological examination in bovine corneal opacity and permeability (BCOP) assay for assessing the ocular irritation potential of fragranced formulations. The Toxicologist 66:244.

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

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

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

Cassidy SL, Stanton E. 1997. In vitro eye irritation studies on organosilicon compounds. J Toxicol - Cut and Ocular Toxicol 16(1):45-60.

Casterton PL. 1998. The bovine corneal opacity and permeability assay: An alternative protocol. In Vitro Mol Toxicol 11(4):337-341.

Casterton PL, Potts LF, Klein BD. 1996. A novel approach to assessing eye irritation potential using the bovine corneal opacity and permeability assay. J Toxicol - Cut and Ocular Toxicol 15(2):147-163.

Cater KC, Raabe HA, Mun G, Harbell JW. 2001. Corporate validation program for predicting eye irritation of surfactant formulations in vitro. The Toxicologist 60:99.

Cater K, Nusair T, Merrill JC, Harbell JW. 2002. Exploratory in vitro eye irritation study of marketed alkaline liquid laundry detergents by BCOP assay and pH/reserve alkalinity (RA) parameters. The Toxicologist 66:244.

Cater K, Mun G, Moyer G, Merrill J, Harbell JW. 2003. Exploratory in vitro eye irritation study of marketed alkaline dry laundry detergents by BCOP assay and pH/reserve alkalinity (RA) parameters. The Toxicologist 72:220.

Cater K, Patrick E, Harbell J, Merrill J, Schilcher S. 2004. Comparison of in vitro eye irritation potential by BCOP assay to erythema scores in human eye sting test of surfactant-based formulations. Poster presentation at the Society of Toxicology 2004 meeting.

Cerven D, Moreno O. 1998. Bovine corneal opacity and permeability test validation as an alternative to the Draize eye irritation test. In: Advances in Animal Alternatives for Safety and Efficacy Testing (Eds. H. Salem and SA Katz). Taylor & Francis:Washington, DC, 261-267.

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

Chan P-K, Hayes AW. 1985. Assessment of chemically induced ocular toxicity: A survey of methods. In: Toxicology of the Eye, Ear, and Other Special Senses (Ed. AW Hayes). Raven Press:New York, New York, 103-143.

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

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

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

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

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

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

Cuellar N, Merrill JC, Clear ML, Mun G, Harbell JW. 2002. The application of benchmarks for the evaluation of the potential ocular irritancy of aerosol fragrances. The Toxicologist 66(1-S):243-244.

Cuellar N, Lloyd PH, Swanson JE, Merrill JC, Clear ML, Mun G, Harbell JH, Bonnette KL. 2003. Evaluating the eye irritancy of solvents in a simple fragrance mixture with the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 72:312.

Cuellar N, Lloyd PH, Swanson JE, Merrill JC, Mun G, Harbell JW, Bonnette KL. 2004. Phase Two: Evaluating the eye irritancy of solvents in a simple fragrance mixture with the bovine corneal opacity and permeability (BCOP) assay. Poster presentation at the Society of Toxicology 2004 meeting.

Curren RD, Harbell JW. 1998. In vitro alternatives for ocular irritation. Environ Health Perspect 106:485-492.

Curren R, Evans M, Raabe H, Ruppalt R, Harbell J. 2000. Correlation of histopathology, opacity, and permeability of bovine corneas exposed in vitro to known ocular irritants. Veterinary Pathology 37(5):557.

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

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

Doughty MJ. 1997. Evaluation of the short-term bovine eye storage protocol for the enucleated eye test. Toxicol In Vitro 11:229-240.

Doughty MJ. 2004. Further assessment of the size, shape and surface features of superficial cells of the bovine corneal epithelium, using scanning electron microscopy. Current Eye Research 28(3):203-214.

Doughty MJ, Petrou S, Macmillan H. 1995. Anatomy and morphology of the cornea of bovine eyes from a slaughterhouse. Can J of Zool 73 (11): 2159-2165.

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

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

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

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

ECETOC. 1998. Eye Irritation – Reference Chemicals Data Bank. Technical Report No. 48(2). 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.

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

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

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

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L255:1-333.

EU. 2004. 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].

Evans MG. 1998. Considerations for histopathological examination of bovine corneal tissue. In Vitro Mol Toxicol 11:335-337.

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

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

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

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

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

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

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

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

Gautheron P, Dukic M, Alix D, Sina JF. 1992. Bovine corneal opacity and permeability test: An in vitro assay of ocular irritancy. Fundam Appl Toxicol 18:442-449.

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

Gettings S, Lordo R, Hintze K, Bagley D, Casterton P, Chudkowski M, Curren RD, Demetrulias JL, Dipasquale LC, Earl LK, Feder PI, Galli CL, Glaza SM, Gordon VC, Janus J, Kurtz PJ, Marenus KD, Moral J, Pape WJW, Renskers KJ, Rheins LA, Roddy MT, Rozen MG, Tedeschi JP, Zyracki J. 1996. The CTFA evaluation of alternatives program: An evaluation of in vitro alternatives to the Draize primary eye irritation test. (Phase III) surfactant-based formulations. Food Chem Toxic 34:79-117.

Gran BP, Swanson JE, Merrill JC, Harbell JW. 2003. Evaluating the irritancy potential of sodium percarbonate: a case study using the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 72:220.

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

Harbell JW, Curren RD. 1998. The bovine corneal opacity and permeability assay: Observations on assay performance. In Vitro Mol Toxicol 11:337-341.

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

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjöström M. 1996. Recommendations of the application of biostatistical methods during the development and validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. ATLA, 24:511-530.

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

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

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

INVITTOX Protocol 98. 1996. Bovine Corneal Opacity and Permeability Assay (Method of Gautheron).

INVITTOX Protocol 124. 1999. Bovine Corneal Opacity and Permeability Assay – SOP of Microbiological Associates Ltd.

Jester JV, Li L, Molai A, Maurer JK. 2001. Extent of corneal injury as a basis for alternative eye irritation tests. Toxicol *In vitro* 15:115-130.

Jones PA, Budynsky E, Cooper KJ, Decker D, Griffiths HA, Fentem JH. 2001. Comparative evaluation of five in vitro tests for assessing the eye irritation potential of hair-care products. ATLA 29:669-692.

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

Kinoshita J. 1962. Some aspects of the carbohydrate metabolism of the cornea. Inves Ophth. 1:178-186.

Kuckelkorn R, Schrage N, Keller G, REdbrake C. 2002. Emergency treatment of chemical and thermal eye burns. Acta Ophth Scand. 80:4-10.

Leopold IH. 1945. Local toxic effect of detergents on ocular structures. Arch Opthalmol. 34:99-102.

Lewin L, Guillery H. 1913. Die Wirkungen von Arzneimitteln und Giften auf das Auge. Hirschwald, Berlin. 2nd edition.

Marsh RJ, Maurice DM. 1971. The influence of non-ionic detergents and other surfactants on human corneal permeability. Exp Eye Res. 11:43-48.

Marzulli F, Ruggles D. 1973. Rabbit eye irritation test: collaborative study. Journal of the Association of Analytical Chemists 56:905-914.

Maurer JK, Li HF, Petroll WM, Parker RD, Cavanagh HD, Jester JV. 1997. Confocal microscopic characterization of initial corneal changes of surfactant-induced eye irritation in the rabbit. Toxicol Appl Pharmacol 143:291-300.

Maurer JK, Parker RD, Jester JV. 2002. Extent of corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. Reg Tox Pharmacol 36:106-117.

Maurice DM. 1957. The structure and transparency of the cornea. J Physiol 136:263-286.

McCulley JP. 1987. Chemical injuries. In: Cornea; scientific foundations and clinical practice (Smolin G, Thoft RA, eds.). Boston:Little, Brown, 527-542.

McDonald TO, Seabaugh V, Shadduck JA, Edelhauser HF. 1987. Eye irritation. In: Dermatotoxicology. (Marzulli FN, Maibach HI, eds). Washington:Hemisphere Publishing Corporation, 3rd edition, 641-696.

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

Millichamp NJ. 1999. Species specificity: Factors affecting the interpretation of species differences in toxic responses of ocular tissues. In: Ophthalmic Toxicology, Second Edition, Target Organ Toxicology Series (Ed. GCY Chiou). Philadelphia:Taylor and Francis, 89-114.

Mishima S, Hedbys BO. 1968a. Physiology of the cornea. Int Opt Clin. 8(3):527-560.

Mishima S, Hedbys BO. 1968b. Measurement of corneal thickness with the Haag-Streit pachymeter. Arch Ophthal. 80:710-713.

Muir CK. 1985. Opacity of the bovine cornea in vitro induced by surfactants and industrial chemicals compared with ocular irritancy in vivo. Toxicol Lett 24:157-162.

Nakano M. 1958. Effect of various antifungal preparations on the conjunctiva and cornea of rabbits. Yakuzaigaku. 18:94-99.

NIOSH. 2004. Work-Related Injury Statistics Query System. Available: http://www2a.cdc.gov/risqs/ [accessed 26 October 2004].

North Dakota State University Extension Service. 1999. Value-based beef cattle production. Available: http://www.ext.nodak.edu/extpubs/ansci/beef/as1163-2.htm#Physiological [accessed on 21 October 2004].

Nourse WL, Tyson, CA, Bednarz RM. 1995. Mechanisms of Mild Ocular Irritation. Toxicol In Vitro. 9(6):967-976.

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

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

OECD. 1996. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Methods. ENV/MC/CHEM/TG (96)9. OECD, Paris.

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

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

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

Parish W. 1985. Ability of in vitro (corneal injury-eye organ-and chorioallantoic membrane) tests to represent histopathological features of acute eye inflammation. Food Chem Toxicol. 23:215-227.

Pasquale LC, Hayes AW. 2001. Chapter 18. Acute Toxicity and Eye Irritancy. In: Principles and Methods of Toxicology. 4th Ed. (Hayes AW, ed.). Philadelphia:Taylor and Francis, 853-915.

Pfister R. 2005. Chemical Trauma. In: Smolin and Thoft's, The Cornea: Scientific Foundations and Clinical Practice (Foster CS, Azar DT, and Dohlman CH eds). New York:Lippincort Williams and Wilkins, 781-796.

Rachui SR, Robertson WD, Duke MA, Paller S, Ziets GA. 1994. Predicting the ocular irritation potential of cosmetics and personal care products using two in vitro models. In Vitro Toxicol 7(1):45-52

Rees WM, Swanson JE, Burdick JD, Hilgers DS, Harbell JW. 2001. Evaluating toxic synergism in hypochlorite-containing solutions using the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 60:99.

Rougier A, Cottin M, DeSilva O, Catroux P, Roguet R, Dossou KG. 1994. The use of in vitro methods in the ocular irritation assessment of cosmetic products. Toxic In Vitro 8(4):893-905.

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

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

Sina JF. 1994. Validation of the bovine corneal opacity-permeability assay as a predictor of ocular irritation potential. In Vitro Toxicol 7(3):283-290.

Sina JF. 1998. Report from the bovine corneal opacity and permeability technical workshop – November 3-4, 1997. In Vitro Mol Toxicol 11(4):315-316.

Sina JF, Gautheron PD. 1994. A multitest approach to evaluating ocular irritation in vitro. Toxicol Methods 4(1):41-49.

Sina JF, Gautheron PD. 1998. Report from the bovine corneal opacity and permeability technical workshop (November 3-4, 1997): An historical perspective. In Vitro Mol Toxicol 11:316-326.

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.

Solti J, Freeman JJ. 1988. Effect of reducing the number of animals in acute toxicity/irritation tests on U.S. and European labeling requirements. The Toxicologist 8:263.

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.

Springer J, Chambers W, Green S, Gupta K, Hill R, Hurley P, Lambert L, Lee C, Lee J, Liu P, Lowther D, Roberts C, Seabaugh V, Wilcox N. 1993. Number of animals for sequential testing. Food and Chemical Toxicology 31:105-109.

Suker GF. 1913. Injury to cornea from oxalic acid. Opthalmol Rec. 23:40-47.

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.

Swanson JE, White BT, Gran BP, Merrill JC, Harbell JW. 2003. Evaluating oxidizing/reactive cleaning products in the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 72:220-221.

Swanson JE, Rees WM, Hilgers DS, Merrill JC, Harbell JW. 2004. Managing toxic synergism in hypochlorite-containing cleaners using the bovine corneal opacity and permeability (BCOP) assay, part II. Poster presentation at the Society of Toxicology 2004 meeting.

Swanston DW. 1985. Assessment of the validity of animal techniques in eye irritation testing. Food Chem Toxic 23:169-173.

Talsma D, Leach C, Hatoum N, Gibbons R, Roger J-C, Garvin P. 1988. Reducing the number of rabbits in the Draize eye irritancy test: A statistical analysis of 155 studies conducted over 6 years. Fundamental and Applied Toxicology 10:146-153.

Tchao R. 1988. Trans-epithelial permeability of fluorescein in vitro as an assay to determine eye irritants. In Alternative Methods in Toxicology, Volume 6, Progress in In Vitro Toxicology (Ed. AM Goldberg). New York:Mary Ann Liebert, Inc., 271-283.

Thoft RA. 1979. Chemical and thermal injury. Int Ophthalmol Clin Summer. 19(2):243-256.

Toxic Substances Control Act of 1976. 1976. Public Law 94-469.

Ubels JL, Erickson AM, Zylstra U, Kreulen CD, Casterton PL. 1998. Effect of hydration on opacity in the bovine cornel opacity and permeability (BCOP) assay. J Toxicol - Cut and Ocular Toxicol 17(4):197-220.

Ubels JL, Pruis RM, Sybesma JT, Casterton PL. 2000. Corneal opacity, hydration, and endothelial morphology in the bovine cornea opacity and permeability assay using reduced treatment times. Toxicol In Vitro 14:379-386.

Ubels JL, Paauw JD, Casterton PL, Kool DJ. 2002. A redesigned corneal holder for the bovine cornea opacity and permeability assay that maintains normal corneal morphology. Toxicol In Vitro 16:621-628.

Ubels JL, Ditlev JA, Clousing DP, Casterton PL. 2004. Corneal permeability in a redesigned corneal holder for the bovine cornea opacity and permeability assay. Toxicol In Vitro 18:853-857.

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

Unger WG. 1990. Review: Mediation of the Ocular Response to Injury. J Ocular Pharmacol 6(4):337-353.

Unger WG. 1992. Pharmacological and neural bases for eye irritation. Ann NY Acad Sci. 641:176-186.

van Goethem F, Hansen E, Sysmans M, Vanparys P. 2002. Development of a new opacitometer for the bovine corneal opacity and permeability assay. Poster presentation at the Fourth World Congress on Alternatives and Animal Use in the Life Sciences, August 11-15, 2002, New Orleans, Louisiana.

Vanparys P, Deknudt G, Sysmans M, Teuns G, Coussement W, Van Cauteren H. 1993. Evaluation of the bovine corneal opacity-permeability assay as an in vitro alternative to the Draize eye irritation test. Toxicol In Vitro 7(4):471-476.

Vaughn D, Asbury T, Riordan-Eva P. 1999. General Ophthalmology. 15th ed. McGraw-Hill: New York.

Wilhelmus KR. 2001. The Draize eye test. Surv Ophthalmol 45(6):493-515.

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

Williams SJ, Graepel GJ, Kennedy GL. 1982. Evaluation of ocular irritancy potential: Interlaboratory variability and effect of dosage volume. Toxicol Letters 2:235-241.

Wilson SE, Mohan RR, Mohan RR, Ambrósio, Jr. R, Hong JW, Lee JS. 2001. The corneal wound healing response: Cytokine-mediated interaction of the epithelium, stroma, and inflammatory cells. Prog Retin and Eye Res. 20(5): 625-637.

York M, Lawrence RS, Gibson GB. 1982. An in vitro test for the assessment of eye irritancy in consumer products – preliminary findings. Int J of Cosm Sci 4:223-234.

13.0 GLOSSARY[1]

Accuracy[2]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance" (see also "two-by-two" table). Accuracy is highly dependent on the prevalence of positives in the population being examined.

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

Benchmark substance: A substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties:

- a consistent and reliable source(s)
- structural and functional similarity to the class of substances being tested
- known physical/chemical characteristics
- supporting data on known effects
- known potency in the range of the desired response

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

Blepharitis: Inflammation of the eyelids.

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

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

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

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

---

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

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

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

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

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

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

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

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

Corneal opacity: 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."

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

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

Ocular: Of or relating to the eye.

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

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

Opacitometer: An instrument used to measure "corneal opacity" by quantitatively evaluating light transmission through the cornea. The instrument has two compartments, each with its own light source and photocell. One compartment is used for the treated cornea, while the other is used to calibrate and zero the instrument. The difference between

photocell signals in the two compartments is measured electronically as a change in voltage, and is displayed digitally, generating numerical opacity values with arbitrary units.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.