APPENDIX A3
SUMMARY MINUTES FROM EXPERT PANEL MEETING ON JANUARY 11-12, 2005

[This Page Intentionally Left Blank]

Department of Health and Human Services
National Institutes of Health
National Institute of Environmental Health Sciences
Interagency Coordinating Committee on the Validation of Alterative Methods (ICCVAM)
Expert Panel Meeting

Summary Minutes of the Expert Panel Meeting to Assess the Current Validation Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants.

Introduction

A public meeting of an independent Expert Panel was convened on January 11-12, 2005, at the National Institutes of Health (NIH), Natcher Center, Bethesda, Maryland, to evaluate several i*n vitro* ocular irritation test methods. The purpose of this meeting was to assess the current validation status of the Bovine Corneal Opacity and Permeability (BCOP), Isolated Chicken Eye (ICE), Isolated Rabbit Eye (IRE), and Hen's Egg Test-Chorioallantoic Membrane (HET-CAM) test methods for identifying ocular corrosives and severe irritants. The meeting was organized by ICCVAM and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), and sponsored by the National Institute of Environmental Health Sciences (NIEHS) and the NTP. A comprehensive report of the Expert Panel is provided as an attachment to these minutes.

The following scientists served on the Expert Panel:

- Robert Scala, Ph.D., (Panel Chair), Tucson, Arizona, United States
- Sally S. Atherton, Ph.D., Professor, Medical College of Georgia, Augusta, Georgia, United States
- Roger Beuerman, Ph.D., Professor, Louisiana State University, New Orleans, Louisiana, United States
- June Bradlaw, Ph.D., International Foundation for Ethical Research, Rockville, Maryland, United States
- Ih Chu, Ph.D., Health Canada, Ottawa, Canada
- Henry Edelhauser, Ph.D., Professor, Emory University, Atlanta, Georgia, United States
- Donald Fox, Ph.D., Professor, University of Houston, Houston, Texas, United States
- Jim Freeman, Ph.D., Lab Director, ExxonMobil Biomedical Sciences, Inc., Annandale, New Jersey, United States
- Sidney Green, Ph.D., A.T.S., Graduate Professor, Howard University, Washington, DC, United States
- Frederick Guerriero, M.S., Senior Occupational Toxicologist, GlaxoSmithKline, King of Prussia, Pennsylvania, United States

- A. Wallace Hayes, Ph.D., D.A.B.T., F.A.T.S., F.I.Biol., F.A.C.F.E., E.R.T., Scientist, Harvard School of Public Health, Andover, Massachusetts, United States
- Hiroshi Itagaki, Ph.D., Deputy Director of JSAAE, Manager of Alternative Section, Shiseido Co., Ltd., Japan
- David Lovell, Ph.D., Reader in Medical Statistics, University of Surrey, United Kingdom
- Yasuo Ohno, Ph.D., D.J.S.T.S., Director of JSAAE, National Institute of Health, Japan
- Robert Peiffer, D.V.M., D.A.C.V.O., Senior Investigator, Merck Research Laboratories, West Point, Ohio, United States
- Lionel Rubin, V.M.D., D.A.C.V.O., Emeritus Professor of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania, United States
- Horst Spielmann, Dr. Med., Director and Professor, ZEBET at the BfR, Germany
- Martin Stephens, Ph.D., Vice President for Animal Research, Humane Society of the United States, Washington, DC, United States
- Katherine Stitzel, D.V.M., Consultant, West Chester, Ohio, United States
- Peter Theran, V.M.D., D.A.C.V.I.M., Vice President Animal Science, Massachusetts Society for the Prevention of Cruelty to Animals, Novato, California, United States
- Scheffer Tseng, M.D., Ph.D., Director, Ocular Surface Research and Education Foundation, Miami, Florida, United States
- Philippe Vanparys, Ph.D., Senior Research Fellow, Johnson and Johnson, Belgium

The following scientists were invited guests:

- Dr. Chantra Eskes, European Centre for the Validation of Alternative Methods, Ispra, Italy
- Mr. Robert Guest, SafePharm Industries, Derby, United Kingdom
- Dr. John Harbell, Institute for *In Vitro* Sciences, Gaithersburg, Maryland, United States
- Dr. Klaus Krauser, Abbott Laboratories, Abbott Park, Illinois, United States
- Mr. Menk Prinsen, TNO Nutrition & Food Institute, The Netherlands

The following ICCVAM agency representatives were present:

- Dr. Robert Bronaugh, (Ocular Toxicity Working Group), U.S. Food and Drug Administration
- Dr. Kailash Gupta, (OTWG), U.S. Consumer Product Safety Commission
- Dr. Karen Hamernik, (OTWG), U.S. Environmental Protection Agency
- Dr. Abigail Jacobs, (OTWG), U.S. Food and Drug Administration
- Ms. Deborah McCall (OTWG), U.S. Environmental Protection Agency
- Dr. Amy Rispin (OTWG), U.S. Environmental Protection Agency
- Dr. Leonard Schechtman, U.S. Food and Drug Administration

- Dr. Margaret Snyder, National Institutes of Health
- Dr. William Stokes, (OTWG), National Institute of Environmental Health Sciences
- Dr. Marilyn Wind, U.S. Consumer Product Safety Commission

The following additional members of the ICCVAM OTWG were present:

- Dr. Meta Bonner, U.S. Environmental Protection Agency
- Dr. Wiley Chambers, U.S. Food and Drug Administration
- Ms. Donnie Lowther, U.S. Food and Drug Administration
- Dr. Jill Merrill, U.S. Food and Drug Administration

The following members of the NICEATM Staff were present:

- Dr. David Allen, Integrated Laboratory Systems, Inc.
- Mr. Bradley Blackard, Integrated Laboratory Systems, Inc.
- Dr. Neepa Choksi, Integrated Laboratory Systems, Inc.
- Ms. Christina Inhof, Integrated Laboratory Systems, Inc.
- Ms. Linda Litchfield, Integrated Laboratory Systems, Inc.
- Ms. Debbie McCarley, National Institute of Environmental Health Sciences
- Dr. Raymond Tice, Integrated Laboratory Systems, Inc.
- Mr. James Truax, Integrated Laboratory Systems, Inc.

The following members of the public were present:

- Ms. Sara Amundson, Doris Day Animal League
- Dr. Daniel Bagley, Colgate-Palmolive
- Ms. Kathleen C. Cater, The Dial Corporation
- Ms. Nicole Cuellar, S.C. Johnson & Son, Inc.
- Dr. Rodger D. Curren, Institute for *In Vitro* Sciences, Inc.
- Ms. Sadhana Dhruvakumar, People for the Ethical Treatment of Animals
- Dr. Carol Eisenmann, Cosmetic, Toiletry and Fragrance Association
- Ms. Megha S. Even, Physicians Committee for Responsible Medicine
- Ms. Myra Karstadt, U.S. Environmental Protection Agency
- Mr. Ray Kemper, DuPont Haskell Lab
- Ms. Sue A. Leary, Alternatives Research and Development Foundation
- Dr. Dan Marsman, Procter and Gamble
- Mr. David J. McCanna, Bausch & Lomb
- Mr. Claude McGowan, Johnson and Johnson CPDW
- Dr. Pauline M. McNamee, The European Cosmetic, Toiletry, and Perfumary Association (COLIPA)
- Mr. Hidenori Meiseki, Dojindo
- Mr. Hans A. Raabe, Institute for *In Vitro* Sciences
- Dr. Michael W. Rohovsky, Johnson and Johnson
- Mr. Chad Sandusky, Physicians Committee for Responsible Medicine
- Mr. Dean Scott, Bureau of National Affairs News

- Ms. Judith E. Swanson, S.C. Johnson & Son, Inc.
- Dr. Kristina Thayer, National Institute for Environmental Health Sciences
- Dr. Kevin J. Trouba, Institute for *In Vitro* Sciences, Inc.
- Ms. Amanda Ulrey, Institute for *In Vitro* Sciences, Inc.
- Ms. Sarah B. Vieh, The Rose Sheet
- Dr. Sherry L. Ward, Physicians Committee for Responsible Medicine
- Mr. Keith Wyatt, National Eye Institute
- Mr. Gary Wnorowski, Product Safety Labs

The purpose of this meeting was to evaluate the validation status of *in vitro* test methods for identifying ocular corrosives and severe irritants. The Expert Panel was asked to evaluate four draft background review documents (BRDs) prepared by NICEATM.

The four BRDs reviewed and discussed were:

- Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Bovine Corneal Opacity and Permeability (BCOP) Test Method
- Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Hen's Egg Test-Chorioallantoic Membrane (HET-CAM) Test Method
- Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Isolated Chicken Eye (ICE) Test Method
- Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Isolated Rabbit Eye (IRE) Test Method.

Call to Order and Introductions

Dr. Robert Scala, Panel Chair, called the meeting of the Expert Panel (Panel) to order at 8:30 a.m. and asked each attendee to state their name and affiliation. Dr. Scala stated that the public would be given the opportunity to speak at various times during the meeting. Each speaker from the public would be limited to seven (7) minutes, and anyone addressing the group should state their name for the benefit of the transcriptionist.

Dr. William Stokes, Executive Secretary for the Expert Panel and the designated government official, read the Statement of Conflict of Interest and explained policies and procedures regarding confidentiality and avoidance of conflict of interest, as follows:

The members of this expert panel serve as individual scientists and not as representatives of any organization. Each member is to exercise judgment as to whether a potential conflict of interest might exist relative to one or more of the topics being discussed due to his or her occupational affiliation, professional activity or financial interest. Should there be a potential conflict of interest, the member is to recuse him or herself from participating in the discussion of panel recommendations and/or decisions on the topic. You will be signing a conflict of interest certification which declares that during this panel meeting you did not participate in discussion of panel recommendations and/or decisions that involve a particular matter that could have a direct and predictable effect on: 1) Any organization,

institution or university system in which a financial interest exists for yourself, spouse, parent, minor child or partner. 2) Any organization in which you, your spouse, parent, minor child or partner serves as an officer, director, trustee or employee or is otherwise similarly associated. 3) Any organization with which you, your spouse, parent, minor child or parent [sic] is negotiating or have any arrangements concerning prospective employment or other such associations. Panel members are asked to identify at the beginning of this meeting the nature of any such conflicts.

None of the Panel members declared a conflict of interest.

Overview of the ICCVAM Test Method and Evaluation Process

Dr. Stokes (Director, NICEATM, NIEHS) provided a brief overview of ICCVAM and NICEATM. ICCVAM was established as an *ad hoc* committee in 1994 in response to revisions in the 1993 NIH Revitalization Act (Public Law [P.L.] 103-43) that mandates that the NIEHS develop criteria for validation and regulatory acceptance of test methods, and develop a process to achieve regulatory acceptance of scientifically valid methods. The *ad hoc* committee issued its report in 1997, and the ICCVAM committee was formally established that year to implement P.L. 103-43 directives. In 2000, the ICCVAM Authorization Act (P.L. 106-545) established ICCVAM as a permanent committee.

The 15 member agencies of ICCVAM include those involved in regulatory and research activities (CPSC; Department of Agriculture [DOA]; Department of the Interior [DOI]; Department of Transportation [DOT]; U.S. Environmental Protection Agency [EPA]; U.S. Food and Drug Administration [FDA]; Occupational Safety and Health Administration [OSHA]) and those involved in non-regulatory research (Agency for Toxic Substances and Disease Registry [ATSDR]; Department of Defense [DOD]; Department of Energy [DOE]; National Cancer Institute [NCI]; NIEHS; National Institute for Occupational Safety and Health [NIOSH]; National Library of Medicine [NLM]; NIH).

The purposes of ICCVAM, as set forth in P.L. 106-545, are to:

- Increase efficiency and effectiveness of U.S. Federal agency test method review
- Eliminate unnecessary duplicative efforts and share experiences between U.S. Federal regulatory agencies
- Optimize utilization of scientific expertise outside the U.S. Federal Government
- Ensure that new and revised test methods are validated to meet the needs of U.S. Federal agencies
- Reduce, refine, or replace the use of animals in testing (i.e., 3Rs), where feasible

The duties of ICCVAM are to:

- Facilitate and provide guidance on test method development, validation criteria, and validation processes

- Consider petitions from the public for review and evaluation of validated test methods
- Facilitate acceptance of scientifically valid test methods
- Review and evaluate new or revised or alternative test methods applicable to regulatory testing
- Submit test method recommendations to U.S. Federal agencies
- Facilitate interagency and international harmonization of test methods

NICEATM is located at NIEHS in Research Triangle Park, North Carolina, and has the following responsibilities:

- Administers ICCVAM
- Provides operational and technical support for ICCVAM activities
- Communicates and partners with stakeholders
- Organizes test method peer reviews, expert panel meetings, and workshops
- Conducts independent validation studies, as resources permit

The definition of validation used by ICCVAM is that it is the process by which the reliability and relevance of a procedure are established for a specific purpose. Validation characterizes the usefulness and limitations of a test method for a specific purpose. Adequate validation is a prerequisite for regulatory acceptance.

The criteria for test method validation are:

1. Clear statement of proposed use and regulatory rationale
2. Biological basis/mechanistic relationship to effect of interest
3. Formal detailed protocol
4. Reliability adequately assessed
5. Relevance adequately assessed
6. Limitations described
7. All data (raw) available for review
8. Data quality: Ideally generated according to Good Laboratory Practices (GLPs)
9. Independent scientific peer review

The criteria for test method acceptance are:

1. Fits into the regulatory testing structure
2. Adequately predicts the toxic endpoint of interest
3. Generates data useful for risk assessment
4. Adequate data available for specified uses
5. Robust and transferable
6. Time and cost-effective
7. Adequate animal welfare consideration (3Rs)

The ICCVAM test method evaluation process also was described.

Dr. Stokes then described the history of *in vitro* alternatives for ocular irritation. Numerous methods were developed in the 1980s-90s, and numerous validation studies were conducted

in the 1990s. In 1993, a workshop was held by the Interagency Regulatory Alternatives Group (IRAG) that evaluated several *in vitro* test methods as replacements for *in vivo* tests. None of the test methods evaluated were considered a valid replacement. However, test guidelines (e.g., EPA [1998] and the Globally Harmonized System [GHS; UN 2003] tiered testing strategy) were modified to allow for the use of *in vitro* test methods following future validation and acceptance. Some countries in the European Union (EU) will accept positive results for classification of R41 (risk of serious damage to the eye).

A recent European Commission (EC) Directive (EU 2004) regarding IRE, BCOP, ICE, and HET-CAM was described. The directive states that, "These tests are not yet validated, and therefore not included in Annex V." However, positive results can be used to consider a substance a severe irritant and R41 applied with no further testing. But, "where a negative result is obtained, an *in vivo* test should subsequently be required, as the *in vitro* tests have not been shown to adequately discriminate between eye irritants and non-irritants."

Dr. Stokes also described the background and history of the ICCVAM evaluation of *in vitro* ocular irritation assays. In August 2003, EPA announced plans to nominate *in vitro* ocular toxicity test methods for review by ICCVAM. Emphasis was placed on those test methods that may be able to identify severe irritants without animal testing. ICCVAM's Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) unanimously recommended the methods as high priority for ICCVAM evaluation. In October 2003, EPA submitted a formal nomination of four ocular evaluation activities to NICEATM and ICCVAM. ICCVAM endorsed the four EPA nominations as high priority in January 2004. The highest priority was evaluation of *in vitro* screening methods for ocular corrosives/severe irritants. An OTWG was then established to coordinate the evaluation with NICEATM.

A *Federal Register* (*FR*) notice was published in March 2004 requesting public comment on the nominations, and data on chemicals evaluated by *in vitro* or *in vivo* ocular irritancy test methods. A second *FR* notice was published in April 2004 requesting nominations of scientific experts for an independent expert panel. Between April and October of 2004, the four BRDs on the BCOP, HET-CAM, ICE, and IRE test methods were prepared by NICEATM. In November 2004, a *FR* notice announced the dates of this meeting, the availability of the BRDs, and a request for public comments. A *FR* notice announcing the availability of additional data and analyses was published in December 2004.

Charge to the Expert Panel and Organization of the Review

Dr. Stokes explained the charge to the Expert Panel. The Panel was requested to evaluate, for each of the four test methods, the extent and adequacy that each of the applicable ICCVAM validation and acceptance criteria have been addressed, based on available information and data, or will be addressed in the proposed studies for the purpose of identifying ocular corrosives and severe irritants in a tiered testing strategy. The Panel was also asked to develop conclusions and recommendations on:

- Current usefulness and limitations of each of the four test methods for identifying ocular corrosives and severe irritants

- The test method protocol that should be used for future testing and validation studies
- Adequacy of proposed optimization and/or validation studies
- Adequacy of reference substances proposed for future validation studies

A tentative post-meeting timeline also was presented.

Acknowledgments

Dr. Stokes acknowledged the many individuals and organizations who helped with this review. These include the following invited test method experts: Menk Prinsen (TNO-CIVO Institutes, The Netherlands); Dr. Klaus Krauser (Abbott Laboratories, Abbott Park, Illinois, United States); Robert Guest (SafePharm Laboratories Ltd., Derby, United Kingdom); and Dr. John Harbell (Institute for *In Vitro* Sciences [IIVS]; Gaithersburg, Maryland, United States). An ICCVAM Working Group (OTWG) comprised of government scientists that is co-chaired by Drs. Karen Hamernik and Jill Merrill, worked with NICEATM to develop the questions that were addressed to the Panel. The OTWG also recommended experts to serve on the Panel and reviewed the BRDs for completeness. The OTWG will review the recommendations proposed by the Panel and develop draft ICCVAM recommendations. ICCVAM recommendations and the Panel's final report will be forwarded to the EPA and other Federal Agencies for consideration.

Welcome and Introduction to the Meeting by the ICCVAM Chair

Dr. Schechtman (Chair, ICCVAM; FDA) added his welcome to the Panel and the meeting attendees. He then briefly described the composition of the Panel, which was composed of scientists from Europe, Japan, Canada, and the United States, with expertise in toxicology, human and veterinary ophthalmology, biostatistics, pharmacology, anatomy and physiology, laboratory animal medicine, and pathology. Dr. Schechtman also discussed the importance of science-based expert panel recommendations. He noted that the advice of the Panel regarding the validation status of the *in vitro* ocular test methods, including their usefulness and limitations, will help guide:

- The formulation of ICCVAM/NICEATM recommendations regarding the validation status of the four *in vitro* ocular test methods of interest
- The conduct of any future studies that might be warranted that could help them satisfy ICCVAM's criteria for validation and acceptance of toxicological test methods and render any of these methods more acceptable for regulatory purposes
- Regulatory agencies on the use of data generated by these test methods that could help in their regulatory decision-making processes

Dr. Schechtman also briefly described regulation of food, drugs and cosmetics by the FDA. FDA is charged with protecting American consumers through the Federal, Food, Drug, and Cosmetic Act and its amendments. He recognized that today's regulations grew out of a series of health-related tragedies that caught the public's attention beginning in the early 1900's. The beginning of ocular irritancy testing in the United States also was described.

Overview of Current Ocular Toxicity Regulatory Testing Procedures

Ms. Debbie McCall (EPA) provided an overview of the current U.S. and European statutes and regulations that require ocular irritation testing. In the United States, the EPA, CPSC, FDA, and OSHA have authority to require testing of particular chemicals and products to determine their ocular irritation potential, as a result of various statutes and regulations (e.g., Toxic Substances Control Act [TSCA] and Federal Hazardous Substances Act [FHSA]). Testing guidelines are in place to aid the regulated community in meeting these testing requirements.

Ms. McCall then discussed the basic procedures for conducting the *in vivo* rabbit eye test. Relevant Testing Guidelines were also described and compared, to include those of the EPA, the EU, the FHSA, and the Organisation for Economic Co-operation and Development (OECD). All four test guidelines are based on the original method of Draize et al. (1944). FHSA requires the greatest number of animals in an initial test (six). EPA, EU, and OECD recommend up to three animals in an initial test, with the possibility of using only one animal for classifying a corrosive substance. The four test guidelines permit use of anesthetics, generally when pain is anticipated. EPA, EU, and OECD require studies to be carried out to 21 days to evaluate reversible/irreversible effects, while FHSA only requires observations out to three days. Irrigation of the eyes is allowed in all four test guidelines after 24 hours; EU and OECD allow for irrigation at one hour for solid substances.

All four ocular damage regulatory guidelines use the same rabbit eye scoring system. The eye of a treated rabbit is subjectively evaluated using the Draize method for three endpoints: corneal opacity, iris effects, and conjunctival effects. The scores for each of these endpoints were described in detail.

Ocular Hazard Regulatory Testing Requirements and Classification Schemes

Ms. Debbie McCall also provided an overview of ocular toxicity classification definitions and criteria among regulatory hazard classification systems (EPA, EU, GHS, FHSA). All current ocular toxicity classification systems are based on the Draize rabbit eye test method (Draize et al. 1944) and scoring system; however, the classification definitions and criteria vary among the systems.

The EPA classification system (1996) was described first. At least three animals per test are usually required for classification, with a one-animal screen permitted. The maximum score in any animal is used for classification of a substance (i.e., the most severe response is used). The EPA classifies substances into four ocular irritant categories, ranging from I to IV. 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. EPA labeling signal words, statements and protective equipment/actions for each of the four categories were described. For example, the signal word for Category I is Danger, and the statements required for labeling are “Corrosive. Causes irreversible eye damage. Do not get in eyes or on clothing.”

In the EU classification system (2001), at least three animals are usually required for classification, with a one-animal screen permitted for corrosive effects. There are two possibilities for classification:

- If a study includes > 3 animals, mean study values (each endpoint averaged over days 1-3 of the study for all animals) are used
- If a study includes 3 animals, individual animal mean values (each endpoint averaged over days 1-3) are used

Hazard classification of ocular irritation in the EU system corresponds to two risk phrases: 1) R36 denotes "irritating to eyes"; 2) R41 denotes "risk of serious damage to the eyes."

Ms. McCall proceeded to describe the GHS classification system (UN 2003). Classification is based on severity of effect and reversibility of the effect. The GHS 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 persistence, as Category 2A (irritating to eyes; reverses in 21 days), and Category 2B (mildly irritating to eyes; reverses within seven days). GHS labeling symbols, signal words and caution statements also were described.

The FHSA classification system (CPSC 1995) uses at least six animals per test. In this system, there are three categories: corrosive, irritant, or nonirritant. A classification of corrosive is used if one or more animals exhibit destruction or irreversible alterations at the site of contact. For irritants, the maximum score in any animal on any day is used for classification. The irritant classification depends on the incidence of test animals exhibiting a positive ocular response. Depending on the number of animals with positive scores in a study, additional testing may be carried out.

Ms. McCall compared and contrasted the four systems. The EPA, EU, and GHS systems allow for classification of corrosive based on a one-animal screen. If the initial animal indicates corrosivity, no additional testing is required. Classification according to EPA and FHSA is based on the most severe lesion in any animal on any day. However, the EU and GHS systems take into account the most severe mean scores over days 1-3, in addition to persistent lesions. All four systems have only one classification for ocular corrosives/severe irritants. However, there are different numbers of classifications for nonsevere irritants:

- EPA (Category II, III, or IV)
- EU (R36)
- FHSA (Irritant)
- GHS (Category 2A or 2B)

Organization of the Panel Review

During the course of the two-day meeting, the Panel addressed a detailed list of questions concerning the completeness of each BRD and the performance of each test method evaluated. The Expert Panel was subdivided into four groups (one group per test method).

Each subgruop was responsible for addressing the questions for the relevant BRD, and drafting responses for consideration by the entire Panel.

Prior to the presentations and discussions by each of the four groups, an invited test method expert presented information on the test method protocol for which he had expertise. A NICEATM staff member then provided a brief summary of the information contained in the test method specific BRD, including accuracy and reliability analyses for the test method, and proposed optimization and validation studies for the test method.

Each Panel group presented its draft responses for each of the questions assigned for the particular test method BRD. After each presentation, the entire Panel discussed the draft positions and offered additional comments and suggestions. The Chairman summarized the discussion for each question and sought consensus from the Panel on the topic. Public comments were accepted following the Panel's discussion of each BRD.

I. ISOLATED CHICKEN EYE (ICE) TEST METHOD EVALUATION

Primary Reviewers: Drs. Robert Scala, Roger Beuerman, June Bradlaw, Wallace Hayes, Robert Peiffer, Nancy Flournoy

Note: Due to a family emergency, Dr. Flournoy was unable to attend the Panel meeting. However, her comments and suggestions were included in the ICE Panel Report.

Overview of the ICE Test Method Procedure

Mr. Menk Prinsen (invited expert from TNO) provided an overview of the ICE test method. Included in his presentation was a description how the ICE test is conducted, and how the data are used to predict the ocular irritancy classification of a test substance. Mr. Prinsen indicated that the ICE test has been used at TNO since 1992 for eye irritation testing.

Mr. Prinsen indicated that the ICE test was adapted from the Isolated Rabbit Eye (IRE) test. He stated that he looked at the possibilities to use slaughterhouse animals as a replacement as an eye donor and we looked at several species, the bovine, the pig and the chicken. He described that the bovine and the pig for us were less suitable because the cornea was too thick, while the structure of the chicken cornea appeared comparable to that of the rabbit. The process of obtaining the chicken eyes was then described, along with the experimental setup, including the 11-chamber superfusion apparatus and the saline drip system. Mr. Prinsen estimated that the total cost of the experimental setup would be approximately $15,000 U.S. He then described how the chicken heads were transported from the slaughterhouse to the laboratory in humidified boxes at ambient temperature, and that the eyelids close spontaneously after death, providing a protective barrier for the corneal surface. He stated that the eyes could be dissected very quickly (approximately 10 seconds each), and placed in the superfusion apparatus under a saline drip at 32˚C for equilibration for 45 minutes.

Mr. Prinsen then detailed the experimental procedure, beginning with the pre-test measurements recorded to ensure the adequate integrity of each test eye. Mr. Prinsen next detailed the method of dosing, in which 30 µL of liquid or 30 mg of solid is applied to each of three eyes per test substance, with one eye remaining untreated as a control eye. He explained that the rationale for this quantity was based on the relative size of the chicken and rabbit, where the chicken eye is roughly one-third the size of the rabbit eye (which is dosed with 100 µL of liquid or 100 mg of solid). He then described how corneal opacity, corneal thickness, and fluorescein retention are measured with a slit-lamp microscope. Representative photographs of each endpoint were provided, along with video images of actual dosing and resulting opacity formation.

Mr. Prinsen then outlined the decision criteria used to assign an ocular irritancy classification using a categorization scheme for each endpoint. He described that a logical subdivision of the combined categories was used to derive an overall irritancy classification, and how these categories could be applied to yield an EU classification. He also noted that in addition to the combination of the three categories, an immediate corneal opacity score of three or

higher, a corneal opacity score of four throughout the test period, or if there is severe loosening of epithelium, would be criteria for assigning a severe irritant classification. Mr. Prinsen also indicated that histopathological effects may also be used to assign an irritancy classification, but the precise decision criteria for this endpoint were not provided.

After Mr. Prinsen completed his presentation, Dr. Scala invited the Panel to ask him questions on the procedural elements of the ICE test, a full record of which is available in the meeting transcript.

Summary of the ICE Test Method BRD

Dr. David Allen (NICEATM) presented a summary of the BRD for the ICE test method. Dr. Allen detailed that the primary data sources that were used in evaluating the performance of the ICE test were extracted from three publications (Prinsen and Koëter 1993, Balls et al. 1995, and Prinsen 1996). The number of substances evaluated for the EU (n=121) was the largest because the EU classification was given for certain substances for which individual animal data were not provided. Therefore, classification based on the GHS (n=92) and EPA (n=90) classification system was not feasible for all 121 substances. Fifteen chemical classes were tested; the most frequently being alcohols, acids, and surfactants. Also, fourteen product classes were tested, the most frequent being chemical/pharmaceutical intermediates, herbicides/pesticides, industrial chemicals, and soaps/surfactants/detergents.

Dr. Allen described the fact that the major ICE test method protocol variation among the three studies was the number of eyes tested per substance. Originally, five eyes per test substance were included, but that number was later reduced to three, reportedly with no effect on test method performance.

Dr. Allen also described that accuracy statistics were calculated for each test method protocol by report and where appropriate, classifications were pooled into one classification per substance as well as using individual studies where a balanced design existed. Overall accuracy was reported as 82% to 85% (depending on the classification scheme used). Likewise, the false positive rate was between 8% to 10%, while the false negative rate was between 30% to 40%.

Dr. Allen also presented the substances that were used to evaluate test method reliability. There were no substances tested in intralaboratory studies, but the Balls et al. (1995) study provided an interlaboratory reproducibility analysis (n=59 substances tested in four laboratories). This analysis was performed both quantitatively (coefficient of variation analysis) and qualitatively (the extent of agreement among laboratories). For the qualitative analysis, roughly 75 percent of the time, all four laboratories got the same outcome, while up to at least three out of four laboratories got the same answer approximately 90% of the time. For the quantitative analysis, median coefficient of variation (CV) values of approximately 35% was noted for all endpoints except corneal swelling, which was approximately 75%. Dr. Allen speculated that this discrepancy may be due to the use of different thickness measurements among the four laboratories, which could result in variability.

Dr. Allen closed by summarizing the draft test method proposals that were presented in the BRD. A proposed test method version was identified which evaluates corneal opacity, corneal swelling, fluorescein retention and morphological effects, along with a proposed standardized protocol based on a method of TNO. The only significant difference in the NICEATM-proposed protocol and the protocol used by TNO is the inclusion of additional eyes for negative controls where three eyes per negative control are proposed, as opposed to just one, in addition to concurrent positive controls and when appropriate, a benchmark control. Potential optimization studies that might enhance the performance of the test method were also identified. These included: 1) a retrospective analysis of the decision criteria that are used to identify corrosives; 2) an evaluation of the potential causes of the lower level of interlaboratory reproducibility for the corneal swelling endpoint; 3) additional evaluation of possible increased interlaboratory variability that was identified for particular chemical classes, albeit with relatively small numbers such as alcohols, acetates and esters and cationic surfactants; 4) determining the feasibility of introducing a quantitative measurement of corneal opacity; and 5) determining the utility of histopathology and when exactly it should be included. Once optimized, the protocol should undergo additional validation studies to further characterize the accuracy and reliability of the optimized method and that is a summary of the ICE test method BRD.

A discussion ensued regarding the use of CVs in the context of interlaboratory variability. Dr. Lovell stated that in this evaluation, caution should be observed in how these CV measures are interpreted. He noted that there will likely be significant variability in the CV values based on the range of endpoint scores. For example, you could easily get a zero value for the fluorescein retention value since that endpoint score ranges from 0 to 3. Conversely, the corneal swelling value has a much larger potential range, and thus would be expected to have greater variability.

Proposed Panel Recommendations for the ICE Test method

1.0 ICE Test Method Rationale

Dr. Scala presented the draft recommendations for the ICE test method for consideration and concurrence by the Panel. He noted that although the mechanistic basis of the ICE test is not known, this may not be of concern given the fact that correlation with irritancy classification was the predominant goal in this context for the assay. He also recognized that the anatomy and structure of the eyes of chickens, rabbits, and humans are different. He mentioned the differences between the ICE test and the *in vivo* rabbit test, and the endpoints that the ICE does not evaluate (i.e., conjunctival, iris effects, no assessment of reversibility, does not account for systemic effects). In a discussion among the Panel in a subsequent test method, they recommended that the ICE BRD should add discussion of cellular mechanisms of corrosion and severe irritation (e.g., necrosis, apoptosis) and relevance to *in vitro* testing, along with the role of responsive inflammatory cells in isolated rabbit eyes.

2.0 ICE Test Method Protocol Components

Dr. Scala continued by highlighting areas of concern in the protocol that were identified by the Group. These included: the potential for variability due to use of different depth measuring devices; temperature not being well controlled; the drip system, which appears difficult to control and results in removal of the tear film; the vertical position of the superfusion apparatus; randomization of the test eyes; length of exposure time; lack of divalent cations in the superfusion medium, and the number of test eyes per substance (n=3). Following a discussion, the Panel recommended that reference substances that are part of the performance standards developed for the validated test method should be identified.

3.0 Substances Used for Previous Validation Studies of the ICE Test Method

Dr. Scala then discussed the adequacy of the types and number of substances evaluated. He indicated that although only one of the studies evaluated used a coding system for the test substances, a lack of coding was not justification for excluding the remaining data.

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

Dr. Scala then continued by noting that the interpretation of the *in vivo* results appeared correct. However, he pointed out that the regulatory classification methods may be less than adequate for use in evaluating or making distinctions between *in vitro* methods and their suitability for chemical product class evaluations. He then noted that original study records were not available for any of the reports evaluated, but an evaluation of the results could be made and the quality of the studies otherwise appears to be adequate. Dr. Scala then summarized the extent of GLP compliance of the *in vivo* studies, along with the need for future studies to be conducting according to GLPs. He did note that the Primary ICE Reviewers believed that lack of GLP compliance was not an adequate basis for excluding data from the evaluation and that future validation studies should be conducted under GLP compliance and original study records should be readily available. He then recognized that most human eye data was from accidental exposures. During Panel discussion, Dr. Kathy Stitzel raised the point that animal testing did not necessarily precede human testing and Dr. Wiley Chambers (FDA) confirmed that point. Dr. Chambers also pointed out that irritating compounds are tested in controlled clinical trials. Dr. Martin Stephens then raised the point that there needs to be greater effort to obtain and consider information on human topical ocular chemical injury. Dr. Roger Beuerman recommended the Alabama Ocular Injury Registry, and Dr. Donald Fox recommended Fraunfelder's registry of ocular injury. Dr. Stitzel also noted that since this evaluation only deals with severe ocular irritants/corrosives, qualitative data would likely suffice as reference data.

Dr. Scala then acknowledged the ongoing debate over the variability of the *in vivo* rabbit eye test. He indicated that the potential variability of the *in vivo* rabbit data has not been adequately discussed in the BRD, and that the evaluation of an alternative method is unavoidably biased by the selection of the *in vivo* data used in that evaluation. Subsequent to a discussion, the Panel recommended that any optimization and validation studies should use existing animal data, if available, and that additional animal studies should only be conducted

if important data gaps are identified. Dr. Martin Stephens expressed a minority opinion that no animal testing was needed for this purpose, as the current database should be considered adequate.

5.0 ICE Test Method Data and Results

Dr. Scala indicated that the approaches used to evaluate the ICE data appear to adequately describe the accuracy and reliability of the test method. However, given the unavailability of original ICE data, a definitive statement regarding the adequacy of these approaches is not feasible.

6.0 ICE Test Method Accuracy

Dr. Scala then discussed the ICE test method accuracy evaluation. The overall false positive rate (8-10%) was considered adequate, but the acceptability of the false negative rate (30-40%) was less evident since this would result in corrosives/severe irritants to be tested *in vivo* (according to the tiered testing strategy). Dr. Scala stated that a comprehensive accuracy assessment in the absence of suitable human data should take account of the variability in the Draize test itself, such as the analysis by Dr. Joe Haseman that was distributed prior to the meeting. Subsequent to discussion, the Panel recommended an assessment based on ranking of experimental data for severity for both the reference method and the *in vitro* test. A minority opinion was expressed by Drs. Stephens and Peter Theran that the use of the term "accuracy" in this context is not appropriate because the *in vitro* test may in fact be a more accurate estimate of the human response. For this reason, the term "concordance" should be used in favor of "accuracy" when comparing the *in vitro* test to the *in vivo* rabbit eye test.

7.0 ICE Test Method Reliability (Repeatability/Reproducibility)

Dr. Scala then stated that the selection rationale for the substances used to evaluate test method reliability was considered adequate, and only one study was used for this analysis. Test method reliability analyses and conclusions were considered sound and appropriate, and both qualitative and quantitative evaluations of interlaboratory variability were conducted appropriately. Dr. Scala noted that no intralaboratory repeatability and reproducibility were conducted because of a lack of appropriate information.

8.0 ICE Test Method Data Quality

Next, Dr. Scala indicated that, given the lack of original records, caution should be exercised when evaluating these data, but that the lack of original records should not be used as a rationale for excluding these data. However, any future validation studies should include coded test substances of known purity, from a common source and centrally distributed; appropriate controls; and be conducted in compliance with GLP guidelines. He then discussed data quality audits, which were not feasible given the absence of original data, and that a more complete retrospective evaluation would be possible of such data were made available.

9.0 Other Scientific Reports and Reviews

No concerns were raised regarding this section. However, Dr. Scala did suggest that personal contacts by agencies to which ICE data have been submitted may be the best method to expedite acquiring more data. Furthermore, if such data are not received, additional *in vivo* studies may be necessary to compile an adequate reference database

10.0 Animal Welfare Considerations

Dr. Scala noted that although there is no additional inflicting of pain or distress to the animal as a result of the testing procedures, because chickens do not come under the Animal Protection Act, there is still a need to ensure their humane treatment.

11.0 Practical Considerations

Dr. Scala then discussed the transferability of the test method, which does not appear to be a significant obstacle to its use. However, he did indicate that specifications for the custom-built superfusion apparatus must be readily available. Following discussions, the Panel recommended that a training video and other visual media on the technical aspects of the assay be produced and that training approaches in the application of this test method should be developed/implemented. The relative cost of the ICE test and the *in vivo* rabbit eye test were considered comparable, and the ICE test can be completed in much less time than the full *in vivo* test (extending out to 21 days). However, during discussions Drs. Stitzel and Itagaki raised the point that a corrosive or severe irritant may be detected within a few hours using a single rabbit, and thus the reduction in time afforded by the ICE would not always be applicable.

12.0 Proposed ICE Test Method Recommendations

Finally, Dr. Scala summarized the draft recommendations for the ICE test method. He stated that the ICE test method appears to be useful in the identification of ocular corrosives from severe irritants in a tiered testing strategy with the following limitations: alcohols tend to be over predicted; surfactants tend to be under predicted; and solids and insoluble substances may be problematic because they may not come in adequate contact with the corneal surface. He also highlighted that the reliability of the ICE test has not been adequately assessed.

A discussion ensued regarding the context of these statements and how they relate to the optimization/validation studies that have also been recommended. Dr. Stokes noted the difference between declaring a test method as useful versus saying that it has been fully validated. He stated that a test method may be used if it is considered to be useful, and once it has been through adequate validation, it will be mandated for routine use before an *in vivo* test. He continued that alternatives have to be considered by animal care and use committees (at least in the United States). Therefore, by stating that a test method can be used, then institutional animal care and use committees are going to have to ensure that it has been considered before a rabbit test is done. He concluded that this doesn't imply that a mandate

for their use will occur, but they need to be considered and some rationale will have to be provided why they are not being proposed for use now.

Notwithstanding the above conclusions, the Panel agreed on following recommendations: a formal evaluation of the optimum number of eyes per test substance should be conducted; a standardized scoring scheme for histopathology, along with identification of the appropriate circumstances to include in such an evaluation, should be developed; and reference photographs for all subjective endpoints should be provided. During discussions, the recommendation from the Dr. Edelhauser to install centering lights on the optical pachymeter to make it easier to take precise corneal thickness measurements was agreed to. Dr. Fox also recommended an evaluation of the impact of delayed use of chicken eyes on performance. Dr. Scala continued by identifying a number of optimization studies considered useful to enhance the performance of the ICE test. These included:

- optimizing the decision criteria to reduce the false negative rate while maintaining a low false positive rate
- determining the utility of rotating the superfusion apparatus to a horizontal position
- determining the utility of including divalent cations in the assay medium
- determining the optimum mechanism for handling differences in corneal swelling values for test substances from different laboratories

The Panel also recommended that reference substances should be identified that can be used as part of performance standards, and that NICEATM/ICCVAM should facilitate the development of a histopathology scoring system for corneal damage (with visual aids). Finally, the Panel recommended that any optimization and validation studies should use existing animal data, if available; that additional animal studies should only be conducted if important data gaps are identified; and that such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing). Dr. Stephens again expressed a minority opinion that there is sufficient data so that additional animal testing for this purpose is not warranted.

Panel Vote on the ICE Report

Dr. Scala concluded this discussion with a vote among the Panel members. He noted that everyone on the Panel, with the exception of Dr. Stephens, was in agreement with the conclusions and recommendations for the ICE test method. Dr. Stephens's dissenting opinions are noted in the relevant sections above.

Public Comment Session

1. Dr. Rodger Curren (IIVS)

Dr. Curren presented a public comment that dealt with the variability of the Draize eye test. He indicated that the question at hand is whether the *in vitro* tests are good enough to replace the *in vivo* test. He stated that the first step would be knowing as much as possible about what the performance of the test that is going to be replaced. He was therefore critical of the

BRD as having, "virtually a complete absence of discussion about the performance of the animal test that we were looking at." He questioned as to how judgments about how well the *in vitro* test performs can be made in the absence of information about the animal test. He continued by stating that given the variability of the animal test, a severe irritant outcome doesn't mean that that chemical is severe by some degree of a physical-chemical property that never changes, because if it is tested again, a different outcome may result.

Dr. Curren referenced what he considered to be the best animal data that he could find to bias it towards the best animal data (the Cosmetics, Toiletry, and Fragrance Association [CTFA] evaluation study of surfactants and surfactant-containing materials conducted in 1993-94). Citing an evaluation done by Dr. John Harbell, he noted that there is underprediction of the severe irritants among these substances of varying degrees. During a subsequent discussion, it is revealed that this analysis is included in the BCOP BRD as Appendix H so that it could remain intact, as it was not part of the performance analysis conducted in the BRD. Dr. Tice also noted during the discussion that Dr. Harbell's overall figure for the underprediction rate of the *in vivo* test (19%) is comparable to that of Dr. Haseman's estimate (8% to 18%), which also is near the value obtained using the Weil and Scala (1971) data (15%).

2. Sara Amundson (Doris Day Animal League)

Ms. Amundson provided general comments on the BRDs and their subsequent review by the Panel. She expressed concern that there seemed to be a tendency for the BRDs to lead the Panel in very specific directions with regard to recommendations and she hoped that was carefully considered as the Panel moves forward with their recommendations. She also noted that accuracy is not relevance and that relevance is validation, a concept that she stated has reached international agreement. She continued by indicating that the fact that this was left out of the direct information communicated to the Panel in the Executive Summary of each BRD was unconscionable.

Ms. Amundson then referenced S.C. Johnson's written comments that indicated that they have utilized the BCOP, one of the test methods under consideration during the meeting, for the past 15 years or so. She continued that they have made internal decisions and have submitted data to EPA for regulatory decisions. She noted that the potential outcome of the meeting was that the ability of companies like S.C. Johnson to continue to use a test method that they have been able to utilize both internally and externally for regulatory decision making purposes for a great number of years would be hindered.

She closed by advising the Panel to keep in mind that if there is confirmatory testing that is required under any sort of strategy or paradigm that is put forward, the concerns about regulatory and internal corporate decision-making may actually stunt the submission of test methods to the ICCVAM for consideration.

II. BOVINE CORNEAL OPACITY AND PERMEABILITY (BCOP) TEST METHOD EVALUATION

Primary reviewers: Drs. Kathy Stitzel (Group Chair), Ih Chu, Henry Edelhauser, Hiroshi Itagaki, Lionel Rubin, Scheffer Tseng, David Lovell

Overview of the BCOP Test Method Procedure

Dr. Harbell (Invited Expert, IIVS) began his presentation by acknowledging Dr. Pierre Gautheron and Dr. Joseph Sina of Merck Research for much of the initial work on the BCOP assay. He then described the corneal lesions associated with eye irritation *in vivo*. There are four major types of corneal lesions:

- Epithelial cell loss -- partial or full thickness loss over some fraction of the cornea
- Stromal swelling -- results from either a change in hydration state, which is reversible, or protein denaturation, which is poorly reversible
- Death of keratocytes -- depth of injury or loss of these cells is associated with increased severity of the lesion and initiation of inflammation
- Endothelial cell loss -- these cells do not regenerate in humans

Dr. Harbell also described the common modes of chemical action in eye irritation. Membrane lysis can occur when surface-active agents solubilize membrane lipids or when organic solvents extract lipids. A second mode of chemical action involves protein coagulation or denaturation, which can result from exposure to acids and certain solvents. Saponification can result from alkalis and is often progressive. Lastly, alkylation or oxidative damage to macromolecules can result from reactive materials such as bleaches and peroxides.

Dr. Harbell then explained the "depth of injury" model first introduced by Drs. Maurer and Jester, who proposed that depth of injury in the cornea is predictive of the degree and duration of injury. The conclusion of these authors is that the changes in the early post-exposure period (i.e., the first 3 hours, or so) are predictive of the degree and duration of injury *in vivo*. This model provides a link between the *in vivo* situation, whether human or rabbit, and the *in vitro* testing situation.

Using the *ex vivo* cornea, as done in the BCOP assay, provides appropriate cellular and structural targets for eye irritation testing, allows determination of depth of injury, and responds to the various modes of action of irritants on the cornea. Factors that impact exposure in the eye also were described. Binding or trapping of substances in the eye will increase irritation, while vaporization, dilution, and flushing will decrease irritation. Rapid binding and/or penetration will favor increased irritation potential.

Dr. Harbell then described the procedures involved in conducting the BCOP assay. A local slaughterhouse supplies the bovine eyes. The eyes are transported in a bucket in Hanks' balanced salt solution (HBSS), and the bucket is maintained over ice. The time between first

slaughter and the time of arrival at the laboratory is between four and five hours. After arrival at the laboratory, the cornea is inspected for defects, scratches and opacity; the quality of the eyes is very important. The cornea is then excised. Corneas are held in HBSS, then are mounted very carefully on a holder. The holder has an O-ring side and a non-O-ring side. The endothelium is placed on the O-ring; it is critical that once placed on the O-ring, the cornea is not moved. The other half of the corneal holder is placed on top of the cornea, and the two chambers are screwed together. The chambers are filled with medium, and the system is allowed to equilibrate. Opacity is then measured quantitatively with an opacitometer. Sodium fluorescein is used to measure damage to the epithelium. If histology is conducted on the corneas, the corneas are fixed after the sodium fluorescein evaluation is completed.

One of the key features of the BCOP assay is complete control over exposure conditions. One can control the exposure concentration as well as the exposure time at the specified concentration, and ensure that exposure is over the whole corneal surface. Control over the post-exposure (expression) period is another important feature.

The *in vitro* score is an algorithm developed by Merck for pharmaceutical intermediates. It uses the opacity score, which ranges from 0 to 700, plus the net optical density (OD) of fluorescein multiplied by 15 (i.e., *in vitro* score = opacity + 15 x OD).

Certain chemicals do not induce direct opacity (e.g., anionic and nonionic surfactants) and so only the permeability score is used. Positive controls are always used at IIVS. Benchmark materials are used whenever possible.

The protocols for liquid and solid test substances were also described. Liquids are normally tested at 100%, while solids are tested at 200 mg/mL. The exposure period for liquids is normally 10 minutes, while that for solids is 4 hours. After the liquid is rinsed off the cornea there is a 2-hour "post-exposure" period, but there is no post-exposure period for solids.

Dr. Harbell then explained that exposure and post-exposure times can be modified to address certain chemical classes or expected consumer exposure scenarios. Reactive chemistries (e.g., peroxides) require extended post-exposure incubation because *in vivo* there is a delayed response. Something with a delayed onset, where apoptosis or delayed necrosis of keratocytes is a factor, requires a longer post-exposure period and histology.

Experience with the European Commission/Home Office (EC/HO) chemicals has shown that when the mode of action is not known, the most conservative approach is to address the reactive chemistry mode by using extended post-exposure incubation and histology. The more rapid changes will also be detected with this approach.

After Dr. Harbell completed his presentation, Dr. Scala invited the Panel to ask him questions on the procedural elements of the BCOP assay.

Summary of the BCOP Test Method BRD

Dr. Neepa Choksi (NICEATM) first discussed the current U.S. regulatory status of BCOP. ICCVAM agencies were surveyed by NICEATM, and the EPA and FDA indicated that BCOP data have been submitted to these agencies for consideration. She proceeded to describe the primary data sources for the BCOP BRD. Nine studies were used for the accuracy and/or reliability analyses in the BRD. In addition to these nine reports, there were 31 studies (discussed in Section 9) that were not used in the BRD analyses for a variety of reasons, including the lack of appropriate *in vivo* data or quantitative *in vitro* data.

The database used for the accuracy analyses included 166 different substances or formulations, representing 15 chemical classes and 20 product classes. The test method protocols used to generate BCOP data were similar to each other, but not identical. Some of the protocol variations include different storage conditions of the bovine eyes during transport, use of positive controls, and analysis of the resulting data.

Dr. Choksi then described the different BCOP data analysis methods. The most commonly used analysis method is the *In Vitro* Irritancy Score (IVIS), which is equal to the opacity value plus 15 times the optical density value. An IVIS >55.1 is considered a severe irritant. A few laboratories use the endpoint (opacity or permeability) with the highest score. One of the studies in the BRD analyzed permeability data only for substances that produce significant permeability without appreciable opacity. The distribution of BCOP tests among analysis methods was briefly described.

Accuracy analyses were performed to determine the ability of BCOP to correctly identify ocular corrosives and severe irritants as defined by GHS (Category 1), EPA (Category I), and EU (R41). Accuracy statistics were calculated for:

- each BCOP study that had acceptable *in vitro* and *in vivo* data
  - by test substance
  - by test
- pooled data from studies with similar protocols

In addition, false negative and false positive rates were calculated for different chemical classes and available physicochemical properties (liquid/solid).

The accuracy using pooled studies was relatively good for all three classification systems, ranging from 77% to 80%. False positive rates for the pooled data ranged from 17% to 23%, while false negative rates ranged from 22.5% to 27% for the three classification systems. The analyses by chemical class showed that the false negative (14%) and false positive (5%) rates were rather good for surfactant-containing formulations (n=34). Liquids had lower false negative (18%) and false positive (21%) rates, than the corresponding rates (33% and 29%) for solids. A major limitation of the BCOP accuracy analysis is that for a majority of the chemical classes (63%; 20/32), only a small (two or less) number of substances were tested. Another limitation is the limited information on the physicochemical properties for some test substances.

For the reliability analyses, BCOP data were available to evaluate intralaboratory repeatability and reproducibility, as well as interlaboratory reproducibility. For the intralaboratory analyses, quantitative CV analyses were conducted. For the interlaboratory analyses, a CV analysis and a qualitative analysis that evaluated extent of agreement between testing laboratories were performed. Intralaboratory repeatability was evaluated for three studies (Dr. Sina's submission, Swanson et al. 1995, Southee 1998) by calculating the CV of the IVIS obtained for replicate corneas exposed to the same test substance. For substances predicted as severe irritants, the mean %CV ranged from 8.2 to 11.1 for the three studies. Intralaboratory reproducibility was evaluated for two studies (Gettings et al. 1996 and Southee 1988), which tested substances in two or more independent experiments (trials). For the Gettings study, which evaluated permeability only for 25 substances in three different trials, the mean and median %CV for the permeability value was 33.4 and 29, respectively. In the Southee study, which evaluated 16 substances in two or more trials in three laboratories, the mean %CV for the IVIS ranged from 12.6 to 14.8 for the three laboratories. As far as the interlaboratory reproducibility, the classification agreement among laboratories was very good. The interlaboratory %CV values for IVIS were also good for substances predicted as severe, but were much higher when the datasets included moderate to mild irritants, which have lower scores and, thus, greater "noise" in the data. No limitations were identified for the BCOP reliability analyses.

The draft BRD on BCOP included a number of proposals. A recommended BCOP version was identified which evaluates opacity and permeability, as well as histology on a case-by-case basis. A standardized protocol was proposed for the recommended version of the BCOP test method; this protocol is based on the method used by the Institute for *In Vitro* Sciences. The only significant difference between the two protocols is that the recommended protocol in the BRD lacks the detailed histology procedures provided in a separate IIVS protocol on histology for the BCOP assay. The decision criteria recommended are those previously described by Gautheron et al. (1994). Proposed optimization studies include a retrospective analysis of decision criteria used to identify corrosives and severe irritants, an evaluation of possible increased interlaboratory variability for specific chemical classes appearing more variable (e.g., alcohols), an evaluation of reduced exposure times for alcohols and possibly other volatile solvents, and determination of the utility of histopathology and when it should be used in the BCOP assay. Once optimized, additional studies are recommended to further assess the accuracy and reliability of BCOP, so that the applicability domain can be better defined and data gaps are filled.

Proposed Panel Recommendations for the BCOP Test Method

1.0 BCOP Test Method Rationale

The Panel agreed that use of living corneal tissue is a good model. Opacity is an important endpoint, but the proposed protocol doesn't differentiate the different mechanisms of opacity. Permeability measures integrity of the cornea and adds important information on degree of injury. Limitations of the BCOP test method are that the method evaluates only corneal effects, it may not identify materials that cause serious corneal injury without changes to opacity or permeability, it doesn't evaluate damage to limbal stem cells, and it doesn't model

protective mechanisms (e.g., blinking, tearing) that affect the outcome of *in vivo* studies. The Panel discussed the importance of the limbus for evaluating damage to the human eye; it is part of the conjunctiva and the most important issue when classifying human eye damage.

The endpoint of corneal opacity is measured in both BCOP and the accepted *in vivo* method. However, BCOP does not measure changes in iris and conjunctiva, and systemic toxicity is not identified by BCOP. Although reversibility is not evaluated in BCOP *per se*, initial depth of injury *in vitro* may be useful to evaluate potential long-term effects. The Panel, therefore, recommended that the BCOP BRD include a discussion of the work of Maurer and Jester, which provides evidence that initial changes can predict long-term effects. The BRD should also discuss the human clinical experience with injury scales that are used to predict long-term effects from immediate injury.

2.0 BCOP Test Method Protocol Components

The Panel agreed with the protocol components in the BRD, with the exception of the following items:

Eyes

- Discourage use of antibiotics since they are not effective at 4º C, could potentially increase permeability of the epithelial cells, and possibly cause drug-compound interactions; however, to inhibit bacterial growth the eyes must be kept cold
- The recommended storage time of 4-5 hours may be too restrictive
- Bovine Spongiform Encephalopathy (BSE) is a risk, thus, appropriate precautions should be taken

Solvent for preparing solutions

- Use 0.9% NaCl, *not* sterile water
- Osmolarity and pH of solutions should be known

Corneal culture medium

- Minimum Essential Medium with Fetal Bovine Serum is not necessary
- Balanced salt solutions should be acceptable

Optimize corneal holder

- It should clamp on the sclera and not the cornea
- Holder should maintain curvature of cornea

Exposure

- Optimize exposure duration for 'volatile solvents'
- Exposure method for solids is problematic

Rinsing procedures

- Recommend optimizing these procedures especially for viscous materials and solids

Histopathology

- Must be added unless the substance is from a class of materials known to be accurately predicted using only opacity and permeability in the BCOP assay
- A grading system for histopathology is needed

Reference substances

- Identification of reference substances that are part of the performance standards developed for the validated test method

Controls Needed

- Positive, negative and benchmark controls are needed
- Each laboratory must establish acceptable ranges

Reexamine Prediction Model

- Is a calculated score advisable/necessary?
- Optimize to identify severe irritants
- The BRD should identify the decision criteria for identifying ocular corrosives and severe irritants and discuss rationale for development

Additionally, the Panel discussed some issues surrounding BSE. It takes a full day after slaughter to determine whether cattle have BSE, but the eyes must be used before then. Another point is that calves up to 16 months do not have BSE, so some companies are trying to use calves eyes instead of eyes from mature animals. It was recommended that NICEATM obtain the BCOP data obtained using calf eyes for further evaluation.

3.0 Substances Used for Previous Validation Studies of the BCOP Test Method

The Panel agreed that the number and classes of substances tested in previous validation studies were acceptable. However, materials known to be severe eye irritants in humans should be confirmed to be severe in BCOP. Since available data indicate alcohols, ketones, and solids are problematic in BCOP, better chemical characterization and physicochemical data on all the test substances are needed.

The Panel considers coding procedures to be important, since data quality could be affected if they are not used. The coding procedures used in the previous validation studies were considered adequate.

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

The Panel agreed that the *in vivo* rabbit eye test method protocols used to generate reference data in the studies cited in the BRD were appropriate. However, the use of the three regulatory classification systems to evaluate *in vitro* methods was questioned by the Panel. Regarding the data quality of the *in vivo* studies, the lack of original study records was a concern of the Panel, but was not considered serious enough to prevent use of the data. Also,

the BRD should include more information on the extent of GLP compliance of the *in vivo* studies.

With respect to the availability of relevant human ocular toxicity information, the Panel recommended confirming that current ocular hazard classification schemes by examining Poison Control Center databases, Department of Labor data, and by reviewing published case reports. Also, the Panel stated greater effort should be put forth to obtain and consider information on human ocular injury from chemical exposures.

Regarding the accuracy and reliability of the *in vivo* rabbit eye test, the Panel agreed that the potential variability of the rabbit eye data was not adequately discussed in the BRD. Therefore, it was recommended that the BRD be modified to include discussion of several publications that address this issue, such as Weil and Scala (1971). An effort should be made to confirm *in vivo* classifications using other data sources such as the Registry of Toxic Effects of Chemical Substances or the International Uniform Chemical Information Database. Any optimization and validation studies should use existing animal data, if available. A majority of the Panel agreed that additional animal studies should only be conducted if important data gaps are identified and such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing). However, Drs. Stephens and Theran had a minority opinion that no additional animal testing should be conducted for this purpose.

5.0 BCOP Test Method Data and Results

The Panel agreed with the BRD assessment of the information presented in Section 5.0 of the document.

6.0 BCOP Test Method Accuracy

With respect to the accuracy evaluation of the test method for identifying ocular corrosives and severe irritants as defined by the EPA (1996), the EU (2001), and the GHS (UN 2003), the Panel agreed that the currently used BCOP assay, with the addition of histology, is acceptable to assess the ability of materials to cause corrosive or severe injury to the eye as part of the screening strategy described in the BRD. However, based on the data presented, the assessment of alcohols, ketones, and solids with the protocol as written is problematic.

The Panel further discussed that alcohols were overclassified in three (BCOP, ICE, IRE) of the four test methods under review. In the BCOP test method, the 10-minute exposure protocol for liquids does not appear suitable for alcohols, but it was mentioned that a 3-minute exposure protocol might be more suitable. The Panel recommended that it would be useful to do a retrospective evaluation of any data obtained from the 3-minute exposure protocol.

The Panel agreed that the accuracy parameters must indicate that the values are a concordance comparison with the results of a single rabbit eye test. Because there is potentially a 10-20% misclassification rate for severe eye irritants in the rabbit eye test, the

Panel discussed the need to correct the performance statistics of BCOP for the Draize test variability. The misclassification in the *in vivo* test method would affect the false negative and false positive rates of the BCOP assay, and must somehow be accounted for in the performance statistics.

Regarding the strengths and limitations of the test method, the Panel stated that the effect of colored substances was not discussed in the BRD. Also, to better determine if certain physicochemical properties are problematic in the test method, consideration should be given to exploring physicochemical effects by using a structure activity or structure property relationship program.

In addition to the BRD analyses conducted, the Panel recommended an assessment based on ranking of experimental data for severity for both the reference method and the *in vitro* test.

7.0 BCOP Test Method Reliability (Repeatability/Reproducibility)

The Panel agreed with the BRD assessment of the selection rationale for the substances used to evaluate test method reliability. Regarding the analyses and conclusions regarding intralaboratory repeatability and intra- and inter-laboratory reproducibility, the Panel agreed that the data from existing studies was extensively reviewed and considered in the BRD. The data indicated acceptable levels of intra- and inter-laboratory variability. However, the Panel stated that the use of CVs should be used with care with this data. Optimization of the protocol may decrease variability.

Positive control data were presented in the BRD; however, negative control data were not included.

With respect to the effect minor protocol changes might have on the recommended test method protocol and the transferability of the test method, the Panel stated that the data indicate the test method is transferable. At what point minor protocol changes will be sufficiently significant to require further validation could not be determined with the information provided.

8.0 BCOP Test Method Data Quality

The Panel agreed that coding of chemicals should be used for all subsequent validation studies. While spot checks of data not part of multilaboratory validation studies could be conducted, the Panel believes this is not necessary. The lack of original notebook data was of some concern to the Panel but not sufficient to remove the data from consideration. The Panel noted that recent information indicates that raw data may be available for many, if not all, of the studies in accuracy evaluation of the BRD.

9.0 Other Scientific Reports and Reviews

The Panel agreed that the BRD adequately identified relevant data from other published or unpublished studies. The BRD also adequately summarized the conclusions published in

independent peer reviewed reports or other independent scientific reviews. It is possible that more data could be obtained by working with trade associations, but much of the data in the BRD is from these types of efforts, so whether more data could be obtained is unclear.

10.0 Animal Welfare Considerations (Refinement, Reduction, Replacement)

The Panel agreed that the BCOP will reduce the numbers of animals exposed to severe irritants. Also, the BCOP will aid in classifying some substances as severe without further animals tests.

11.0 Practical Considerations

The Panel agreed that the BRD adequately addresses the facilities, major fixed equipment, and availability of other supplies needed to conduct the BCOP test method. The required level of training and expertise to conduct BCOP were also adequately considered. However, the Panel thought the description of training of technicians for the *in vivo* test may be incorrect – proficiency in the *in vivo* test is demonstrated the same way as for BCOP.

The Panel recommended development of a training video and other visual media on the technical aspects of the assay. Also, training approaches in the application of this test method should be developed and implemented.

The discussion of the test method cost in the BRD should be modified to reflect the written public comments submitted by S.C. Johnson in December 2004.

The Panel noted that for very corrosive substances and some severe irritants, an *in vivo* evaluation may be completed within 4 hours in the rabbit eye test. Thus, it is not always true that *in vivo* evaluations would be extended to 21 days.

12.0 Proposed BCOP Test Method Recommendations

Regarding the recommended version of the BCOP test method, the Panel suggested confirming with several active laboratories that the proposed changes are workable.

Regarding the Panel's conclusions on the recommended standardized BCOP test method protocol, the Panel discussed at length whether the BCOP assay can be considered "valid", "validated", "acceptable", "useful" or whether the "validation criteria have been met" to identify ocular corrosives and severe irritants. However, even after the lengthy discussion, the Panel was uncertain what terminology to use to describe the conclusions reached about the recommended standardized BCOP test method protocol. The Panel Chair ultimately decided that this section of the BCOP report (Section 12.2) required a Panel vote.

A majority of the Panel agreed with the following conclusions regarding the recommended standardized BCOP test method protocol:

- For the purpose of detecting severe eye irritants in the testing scheme outlined in the BRD, the BCOP test method presented is useful in identifying ocular

corrosives and severe irritants, as described in the BRD, with the exception of:
  - Alcohols, ketones, and solids
  - Histopathological examination must be added, unless the substance is from a class of materials known to be accurately predicted using only opacity and permeability in the BCOP assay
- There is a need to confirm that the BCOP test identifies substances known to cause serious eye injury in humans
- Negative, positive, and benchmark (where possible) controls should be included
- Eyes from young adult cattle should be used
- Users should be aware of the risk of BSE and other zoonoses and use proper precautions
- 0.9% sodium chloride should be used as the standard diluent and rinse
- the osmolarity and pH of test solutions should be determined

Although Dr. Freeman abstained from voting on Section 12.2 since he believed the discussion had not been satisfactorily resolved due to time constraints, he agreed to provide a written abstaining opinion. Drs. Stephens and Theran were opposed to the language presented for Section 12.2 because they felt there was undue pressure on the group to "back off" the issues; they agreed to prepare a written dissenting opinion.

With respect to the recommended BCOP optimization studies in the BRD, the Panel made the following suggestions as future improvements to the test method:

- Using the larger holder designed by Ubels
- Reexamining the calculated total score
- Optimize media used to bathe the eyes
- Optimize rinsing procedures
- Consider use of younger animals
- Discourage the use of antibiotics

Optimization studies will be necessary to ensure any changes to the protocol will decrease the variability of the test method.

With respect to the recommended BCOP validation studies in the BRD, the Panel believes that validation studies are not necessary except, possibly, for solids, alcohols, and ketones. It is possible that submission of additional historical data for these types of substances may be sufficient. The Panel also stated that validation is not required for the addition of histopathology or changes in scoring system.

The Panel made one additional comment with respect to the BCOP test method. It was suggested that consideration be given to use of porcine eyes, since the porcine eye is a better model for human eyes than the bovine eye.

Panel Vote on the BCOP Report

The Panel Chair asked the Panel members to indicate whether they accepted the revised document that had been presented, with all the revisions that had been placed in it. A show of hands indicated that all Panel members, except for Drs. Freeman, Theran, and Stephens, accepted the BCOP report. Drs. Theran and Stephens stated that they voted no because of their opinion that animals should not be used for future validation studies, and because of their minority opinion on Section 12.2. Dr. Freeman stated that he voted no because he thought "the question at hand [in Section 12.2] was the most important question that we were asked to address over the two days that we have been here … if we are not taking a position to state clearly that the criteria have been met, this was so close that we should have specified what criteria were not met." Drs. Freeman, Theran, and Stephens indicated they agreed with the Panel on all other parts of the BCOP report.

Public Comment Session

1. Dr. Sherry Ward (Physicians Committee for Responsible Medicine)

Dr. Sherry Ward prefaced her comments by stating that she is a member of an animal advocacy organization, but is also a scientist with more than seven years of experience working in an industrial *in vitro* toxicology laboratory, where she spent significant time working on the validation and development of *in vitro* ocular toxicology test methods.

One important point she thought was missing from the BRDs is the importance of these methods to industry. If the methods are approved as screening assays, this will give companies the flexibility they need to choose the proper method or methods that are compatible with the testing of their products.

Another issue that she thought could have been better clarified in the BRDs is the potential to optimize *in vitro* decision criteria to look at how to reduce over- or under-prediction of the methods. It would have been helpful to have some of this analysis in the BRDs because this may have helped the Ocular Expert Panel to make a better decision on whether any of the particular methods should have been recommended for validation at this time with the ICCVAM. Without this information, it is really hard to make that decision.

Additionally, in the United States, a method has to be validated before it can be considered for acceptance by regulatory agencies. That is one of the ICCVAM regulatory-acceptance criteria. Inclusion in the Institutional Animal Care and Use Committee (IACUC) reviews due to this designation would be helpful, but this seems to be a minor application for saying the methods are acceptable, and it is not sufficient to ensure their use.

Dr. Ward stated that her organization hoped to see progress on one or more of the methods being recommended for validation by the ICCVAM and would like ICCVAM to act to validate the methods as soon as possible. Her organization also strongly opposes any additional animal testing for conducting new optimization or validation studies and requests that the isolated rabbit eye protocol contain stronger wording prohibiting the use of

laboratory rabbits. The word "should" in the recommendation is not a very strong statement to keep from using laboratory rabbits.

2. Ms. Nicole Cuellar (S.C. Johnson & Son)

Ms. Nicole Cuellar provided a corporate perspective on use of the BCOP assay. For the past 10 years, S.C. Johnson has put significant effort into trying to reduce the use of animals in hazard assessment and frequently has used alternative assays for product development and labeling decisions. The company currently uses the BCOP in a weight-of-evidence approach for hazard classification and labeling purposes for its non-registered products. A benchmark-related approach, using the BCOP assay, has been used with a variety of different product types. For non-registered products, such as air fresheners and cleaning products, the company has used alternative assays for labeling and hazard. For registered products, the company has used alternative assays more in product-development situations and worker safety assessments. By combining in-house historical data, toxicology information on raw materials, and postmarket surveillance, the company is comfortable using the BCOP alone, without *in vivo* testing, on non-registered products. This assay is an indispensable tool that the company has used for addressing the potential irritation of S.C. Johnson products.

Ms. Cuellar then described how S.C. Johnson conducts the BCOP assay; a standard protocol is used with concurrent benchmarks, controls, and histology if needed. The exposure and post-exposure times are chosen to be appropriate for the formula or chemical class or the test material. Each formula is carefully matched with a specific benchmark material for which the irritancy potential is well understood. Histology is conducted on both the test sample and the benchmark, for a complete assessment of degree and depth of injury. Histology is conducted under the following situations: to understand new chemistries and formulas; to investigate known chemistries with delayed effects; for chemicals where the mode of action is not easily predicted or the complete picture is needed. It is also used to further characterize damage not obvious from the standard BCOP endpoints and to resolve borderline cases.

In conclusion, Ms. Cuellar stated that she appreciated the enormous effort that has gone into the production of the BRD and review of the data for support of this assay. S.C. Johnson submitted five datasets for this evaluation, and is very supportive of this effort. The company respectfully requests that its comments be considered due to the wealth of investigation and its application by S.C. Johnson.

3. Dr. Rodger Curren (IIVS)

Dr. Roger Curren first addressed a few of the points addressed during the Panel recommendations for the BCOP test method. He stated that IIVS has negative control data available. Also, regarding the issue of whether the protocol is acceptable to laboratories other than IIVS that are running the BCOP test method, one of the things IIVS has done over the years is to have workshops just for users of the BCOP assay, where we have discussed, in general, among our laboratories, what protocols are appropriate.

Dr. Curren then addressed three issues related to the BCOP test method evaluation. He first described how histology was found to be useful in identifying the severe irritancy of a few substances that did not produce significant opacity or permeability in the EC/HO validation study of BCOP. IVIS, histology slides, *in vivo* Maximum Average Score (MAS) values, and EU and EPA classifications were presented for quinacrine and sodium oxalate. The substances are classified as Category I and R41, respectively, in the EPA and EU systems, but in the EC/HO study, they were predicted as mild in the BCOP assay (IVIS <25). However, histologically, these substances produce severe effects. For sodium oxalate, there is severe destruction of the epithelium and penetration of the material into the stroma. For quinacrine, there are severe changes in the stroma, and the endothelial cells are damaged. It is reasonable to call these lesions severe. Adding histology to this study improved the sensitivity from 81% to 90% for the EPA classification system.

Dr. Curren also discussed the prevailing definitions of accuracy versus concordance. He reviewed the American Society for Testing and Materials (ASTM) standard practice report, which has years of experience with validation-type exercises. The ASTM report defines accuracy as "expressing the closeness of a test result to a true value or an accepted reference value." An accepted reference value must be of fairly high quality. For the purposes of this Panel evaluation, many scientists think of the true value as a human result, with the animal as only an imprecise surrogate. Dr. Curren believes it is more accurate to refer to the BRD accuracy analyses as performance of the *in vitro* methods relative to the rabbit for eye irritation. The statistics in the BRDs form a set of performance statistics, not accuracy measurements. Concordance has often been the historical way to do it, from the original Cooper statistics. Performance statistics are measures of the concordance with the rabbit-test results.

Finally, Dr. Curren suggested considering the work of the biostatistician, Feinstein, from Yale. Dr. Feinstein talks about how sensitivity and specificity are not user statistics. The relevant statistics may be slightly different than looking at sensitivity and specificity.

4. Ms. Sara Amundson (Doris Day Animal League)

Ms. Sara Amundson provided some clarification on the recommendation to use eyes from 18- to 48-month old cattle. Based on her experience as an animal advocate and her knowledge of the slaughtering industry, the preponderance of beef cattle are going to be in that age range naturally, without having to optimize this as a goal set within the test-method protocol. It is not economically feasible to have beef cattle that are going to be more aged than that, from a rancher's perspective. She asked the Panel to keep in mind that this information is generally available directly from the slaughterhouse.

5. Ms. Sadhana Dhruvakumar (People for the Ethical Treatment of Animals [PETA])

Ms. Sadhana Dhruvakumar was pleased with the presentation of the Panel's recommendations and the very positive conclusions. She understood from the Panel presentation that the BCOP is acceptable for use, with certain caveats, and that the caveats

can be resolved retrospectively, with more existing data, which is something that can be done pretty quickly. She also requested that the Panel consider using the term "considered scientifically valid," or better yet, "validated" instead of using the term "acceptable for use." She believed that would have greater clarity than just saying "acceptable for use," if validity is indeed what the panel means by that term.

III. ISOLATED RABBIT EYE (IRE) TEST METHOD EVALUATION

Primary reviewers: Drs. James Freeman (Group Chair), Sally Atherton, David Lovell, Yasuo Ohno, Horst Spielmann, Peter Theran

Overview of the IRE Test Method Procedure

Mr. Robert Guest (SafePharm Laboratories; Derby, United Kingdom), an invited expert with many years of experience using the IRE test method, provided an overview with background information, technical aspects, a description of the ocular scoring system and a discussion of the decision criteria for identification of an ocular corrosives or severe irritants.

Dr. Guest indicated that the IRE test method was similar to the ICE test method previously discussed by the Panel. He indicated that SafePharm Laboratories has been performing toxicology studies for over 30 years and that 300 to 400 eye irritation tests are performed each year on a variety of test substances. Data from these tests are submitted to regulatory bodies worldwide and are used for occupational safety assessment. A tiered-testing strategy was used in accordance with OECD Test Guideline (TG) 405, EPA, and other regulatory guidelines. Initially, physicochemical properties, reactivity, corrosivity and other factors are considered before testing. Once tested, a positive outcome results in labeling as a severe irritant/corrosive. A negative outcome results in testing in a single animal, then a decision is made to follow-up with a full study, if necessary. The original IRE protocol was prevalidated using 14 chemicals, most from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) reference bank, followed by a study with GlaxoSmithKline (GSK) using a total of 30 test substances. The principal advantage of the IRE is the use of the main target tissue of the rabbit eye (cornea) without pain or distress. He pointed out that the weighted Draize ocular scoring system places most of the weight (73%) on corneal scores. In addition to a qualitative endpoint used *in vivo* such as corneal opacity, quantification of corneal swelling is possible in the IRE. Corneal opacity (area and intensity), corneal swelling (percent increase in corneal thickness), and uptake of sodium fluorescein (area and intensity) are measured routinely. Histopathology is not performed routinely, but is available if necessary.

Dr. Guest discussed the technical aspects of the IRE test which are outlined as follows:

- The perfusion apparatus consisted of 11 custom-built temperature-controlled chambers with four temperature probes to monitor temperature (1200 British pounds [1740 Euros, $2268 U.S. dollars at the current exchange rates]).
- A portable slit-lamp and an ultrasonic pachymeter are used.
- Other standard laboratory equipment/instruments are needed such as temperature-controlled baths, peristaltic pumps, etc.
- New Zealand White Rabbit weighing 2.5-4 Kg are used.
- Control eyes are occasionally obtained from animals that have undergone skin testing, but would have had either no reaction or a mild reaction.
- Corneal pachymetry is performed at the optical center and at four other points, namely the 3, 6, 9, and 12 o'clock positions. During this procedure, the

animals do not even blink and there is no distress or damage to the cornea as a result of this measurement.

- Eyes are carefully dissected, placed in holders held in place by jaws, and then equilibrated for 30 minutes after placing the holders in the chambers with the bath temperature maintained at 32°C.
- Eyes are examined prior to testing and any with corneal effects or fluorescein penetration are rejected.
- Following equilibration, the isolated eyes are perfused with saline for hydration.
- Before application of test substances, the eyes are re-evaluated by slit-lamp for corneal effects, corneal thickness is measured again and a fluorescein examination is conducted. Eyes with an increase in corneal swelling greater than 10% are rejected.
- For application of the test substance, the holders are removed from the chamber and then placed to keep the eyes in a horizontal position. A volume of 0.1 mL or weight of 100 mg is applied to the surface of the cornea. For solids, the material is either gently compacted into a 0.1 mL volume in a syringe with the tip cut off or weighed in a gelatin capsule and sprinkled over the cornea.
- The corneas are evaluated macroscopically and by slit-lamp and scored for opacity and area. The mean corneal thickness from five measurements is calculated. The advantage of the optical pachymeter is that measurements are not precluded by corneal opacity or dye uptake. Corneal evaluation and thickness measurements with epithelial changes (i.e., mottling, stippling, sloughing, or ulceration) are evaluated under diffuse illumination (1, 2, and 3 hours) and a more detailed evaluation using a slit-lamp biomicroscopic examination is then conducted (1, 2, 3, and 4 hours) to look at the stroma and endothelium.

The ocular scoring system of Hackett and McDonald (1991) was used; a more detailed scoring system than the Draize. At 4 hours, fluorescein uptake was scored for area and intensity.

The decision criteria used to identify a test substance as a severe ocular irritant or corrosive were developed retrospectively, by looking at the data generated in-house. A severe irritant/corrosive is identified as any test substance that produces a maximum corneal score (in three tested eyes) of 3 or greater (opacity x area), maximum fluorescein penetration of 4 or greater (area x intensity), mean corneal swelling (n=3) equal to or greater than 25%, or any single incidence of disruption of the epithelium (stippling, mottling, etc.) in which the control eyes did not respond. If any of these criteria are met, the substance is labeled a severe irritant/corrosive and it is not tested *in vivo*. This testing strategy has been in use since 1999 and has resulted in a reduction in the number of animals exposed to severe irritants. Use of the IRE in conjunction with other *in vitro* tests (e.g., human reconstituted tissue models), would, hopefully, result in replacement of the rabbit eye test.

After Mr. Guest completed his presentation, Dr. Scala invited the Panel to ask him questions on the procedural elements of the IRE assay.

Summary of the IRE Test Method BRD

Dr. Allen (NICEATM) thanked Mr. Truax (NICEATM) for compilation of the BRD. Dr. Allen pointed out that although the IRE test is currently accepted in the EU as a screen for the identification and labeling of ocular corrosives and severe irritants, IRE test data have not been submitted to U.S. regulatory agencies. Dr. Allen indicated that there were four primary sources of data for the IRE test method (CEC 1991; Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). Dr. Allen showed the number of test substances used for each of the regulatory classification systems (GHS, EPA and EU) and he noted that the numbers with an EU classification were higher, because some *in vivo* data were provided with assigned EU classifications, but the *in vivo* raw data were unavailable to allow conversion of the data to the other classification systems. No intralaboratory data were available for analysis in any of the reports, although interlaboratory data (three to four laboratories) were provided in the CEC (1991) and Balls et al. (1995) reports. The EU classifications assigned to eight compounds on the basis of skin corrosivity test results or pH extremes were excluded, since the accuracy analysis is base don the ability of the ICE test method to correctly predict ocular responses in the intact rabbit.

The database consisted of 149 test substances with 124 chemicals and 25 products or formulations. Fifteen chemical classes (heterocyclics/aromatics, acetates/esters, and formulations were the most common) and 14 product classes (industrial chemicals, soaps, and surfactants were the most common) were identified. The endpoints from the recommended IRE protocol in the BRD were corneal opacity, corneal swelling, fluorescein penetration, and evaluation of epithelial integrity. However, the Gettings et al. (1996) study measured only corneal swelling, while Balls et al. (1995) evaluated the first two parameters, the CEC (1991) study used the first three endpoints, and Guerriero et al. (2004) used all four endpoints. Accuracy of each test method was evaluated based on use of all three regulatory classification systems, when possible. The Guerriero et al. (2004) data demonstrated accuracy in the 77-78% percent range with false positive rates of 33 to 34% and false negative rates of 0, albeit with a small number of compounds tested (n=36-44). The small numbers of test substances within any particular chemical class (n=2 to 3) made it difficult to evaluate the performance of any specific chemical classes.

Reproducibility analyses indicated that in the Balls et al. (1995) study using 59 substances and four laboratories, 59% of the time all four laboratories were in agreement with respect to the outcome. Three of four laboratories were in agreement 85% of the time. In the CEC (1991) study with 21 substances among three laboratories, all three laboratories were in agreement 81% of the time. Three of four laboratories were in agreement 95% of the time. If limited to severe irritants, the four laboratories in Balls et al. (1995) were in agreement 100% of the time and the three laboratories in the CEC (1991) study were in agreement 83% of the time. In the Balls et al. (1995) study, a wide range of %CV values (0-200%) using corneal opacity and corneal swelling as endpoints were obtained; the median %CV values was 43.4% and 49.7%, respectively. If only GHS Category 1 substances (i.e., severe

irritants) were considered, median %CV's were 35% to 40%. For the CEC (1991) data, median %CV's were in the range of 24 to 43% (endpoint times were slightly different for some laboratories in this study). For GHS Category 1 irritants, %CV's ranged from 15 to 30%.

A recommended standardized protocol using corneal opacity, corneal swelling, fluorescein penetration, and evaluation of epithelial integrity was proposed in the BRD. This is essentially the SafePharm Laboratories protocol described by Guerriero et al. (2004) with additional positive, negative, and reference controls included.

Proposed Panel Recommendation for the IRE Test Method

1.0 IRE Test Method Rationale

Dr. Freeman pointed out that the Panel concurred with the description in the BRD, but recommended inclusion of a discussion of potential cellular mechanisms of ocular corrosion and severe irritation. This discussion should include the relevance of necrosis and apoptosis to *in vitro* testing, the role of resident and migrating inflammatory cells and their products in ocular irritation *in vivo*, and the consequence of having an incomplete response *in vitro*. The BRD needs to be updated to reflect the basis of the test method as a correlation of descriptive observations or toxicity, rather than mechanistic. In addition, the IRE Panel suggested that additional studies such as microscopy and immunohistochemistry might add to the accuracy of the test method.

The regulatory use and rationale was thoroughly covered. However, it should be noted that the IRE test method does not account for effects on the iris and conjunctiva, nor does it account for reversibility of corneal effects or systemic effects or identification of slow-acting irritants.

The IRE Panel recommends consideration of the use of microscopy or histopathology to improve sensitivity and scope. These efforts may provide insights into early markers of effect or identify transient *versus* progressive changes.

2.0 IRE Test Method Protocol Components

There is a limited dataset using the recommended IRE protocol. The recommended protocol enhancements do appear to improve the accuracy of the test. The recommended protocol, however, has not been directly assessed across other laboratories. In addition, the decision criteria (prediction model) may need to be modified by use of a statistical paradigm (e.g., discriminant analysis) to enhance performance (i.e., to reduce the false positive rate without appreciably increasing the false negative rate). In addition, positive and negative controls and reference substances should be identified from a validated reference substance list such as that being prepared by the Expert Panel Reference Substances Subcommittee.

The IRE Panel recommends that appropriate sources of rabbit eyes be defined. Acceptable rabbit strains should be identified. Acceptable storage and transport conditions (e.g.,

temperature limits, time limits, required buffers, salt or other solutions, containers, etc.) of the isolated eyes for shipment should be defined.

The IRE Panel agrees with the BRD that corneal opacity and swelling, fluorescein penetration, and epithelial integrity should be used as endpoints in the IRE test method. In addition, identification of the reference substances that are part of the performance standards developed for the validated test method is recommended. Data should be collected according to GLP-compliant procedures. Finally, the BRD should clarify the orientation of the eye during application of the test material. Other considerations include application of confocal microscopy or histopathology to detect changes at the cellular level, quantification of the observation (e.g., counting pixels) where possible, and use of descriptive statistics based on individual scores. Finally, the statistical algorithm or rationale used to establish the decision criteria should be more clearly defined and stated in the BRD.

Any further additions to the test method should be backed by a specific rationale. The recommended protocol was adequately covered in the BRD. Consideration of the use of histopathology and defined, validated reference substances should be included as previously described by the IRE Panel. The types and numbers of substances used for prior validation have been adequately described in the BRD. However, further optimization or validation would require use of the reference substances list being developed by the Expert Panel subgroup.

3.0 Substances Used for Previous Validation Studies of the IRE Test Method

The Panel agreed that the number and classes of substances tested in previous validation studies were acceptable.

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

The IRE Panel noted that the Draize test has been used basically unchanged for decades. However, it was suggested that the Draize test could be improved *vis à vis* by use of some of the technology being considered for use in the *in vitro* studies.

The IRE Panel considered the interpretation of the results to be correct as described in the BRD. However, a question was raised by the Expert Panel regarding the adequacy of using regulatory classification systems for evaluating *in vitro* methods and the suitability for chemical or product class evaluations.

Issues regarding data quality were adequately written in the BRD. The IRE Panel felt that if evaluation of the results can be made and the quality of the study appears to be adequate, then lack of original study records does not raise undo concern about a study. Reference to data quality and use of GLPs is covered in the BRD. If the work is performed in a well-established laboratory, then no distinction between GLP-compliant or non-compliant studies is required, and a lack of GLP-compliance *per se* is not a sufficient criterion for exclusion of data of evaluation of performance.

Previously, the Expert Panel had expressed the need for greater effort to find and consider human topical ocular chemical injury data. However, it was recognized that limited data are available and dose and exposure rates would be difficult to quantitate. In addition, no scoring or time course data would likely be available for comparison to an *in vitro* test method.

The IRE Panel indicated that more discussion of the variability of the *in vivo* data is needed in the BRD. The question is how concordant is rabbit data with human data. There is a need to develop an acceptable reference standard, since we do not know if inaccuracy results from inconsistencies in the *in vitro* test method or from misclassification based on a single *in vivo* result.

Further optimization or validation studies should use existing animal data, if possible. Additional animal studies would be used only if data gaps are identified. Such studies should be carefully designed to garner as much information as possible and maximize the amount of pathophysiological information obtained.

Dr. Stevens expressed and provided a written minority opinion that no additional animal tests should be performed for this purpose.

5.0 IRE Test Method Data and Results

The Panel agreed with the BRD assessment of the information presented in Section 5.0 of the document.

6.0 IRE Test Method Accuracy

The recommended protocol includes the additional parameters that enhance accuracy (e.g., Guerriero et al. 2004). No additional datasets were produced with that method. The statistical methods were limited, but appear to be appropriate for descriptive toxicology data, and the conclusions on reliability in the BRD appear to be sound.

Documentation of the data quality was adequate. The studies using the recommended protocol were conducted according to the spirit of GLP.

The reference studies analyzed in the BRD were independent efforts, so lot-to-lot consistency really did not apply here. The consistency was controlled and described within three of the four studies, but not described in the fourth. The stability of chemicals over each study's time frame was not discussed in the BRD.

This section was adequately described in the BRD. Accuracy results summarized in Section 6.1 (Tables 6.1 to 6.3) of the BRD provide a correct overview of performance as reported in the literature, as well as discordant results. Accuracy appears to be improved, based on a small n, with the recommended method, resulting in a false negative rate of 0 and a false positive rate of 33 to 34%.

Draize variability must be included in the discussion in the BRD. There is a weakness in this evaluation. That is, there is a lack of a common protocol for all of the studies analyzed. However, the IRE Panel found it encouraging that the accuracy appeared to improve in the protocol that has become the recommended protocol.

This section was adequately described in the BRD. However, the BRD needs to be revised to assure that the temporal sequence of the studies described is consistent with the publication dates. In Section 6.3 (Tables 6.4 and 6.5), the source of the *in vivo* and *in vitro* data and appropriate author information should be included, and the datasets used to calculate the irritancy classifications identified.

Differences in reproducibility of the Draize test must be taken into account when comparing the predictive value of *in vitro* alternatives. Other relevant information (e.g., weight-of-evidence approach) may clarify the performance of the IRE. It was noted that variability of the Draize test for corrosives or severe irritants might not occur to the same extent as it does for milder irritants.

7.0 IRE Test Method Reliability

The BRD should incorporate information from Bland and Altman (1986) that discusses statistical comparison of methods with poor reproducibility. Information from the European Centre for the Validation of Alternative Methods (ECVAM) skin-irritation prevalidation study on repeatability and reproducibility should be obtained and incorporated where relevant to the ocular test systems. Information from Dr. Sebastian Hoffman's detailed variability analysis comparing SD's and CV's for two skin-irritation models, where relevant, should be incorporated. A strategy to evaluate reliability in any future optimization and validation testing should be developed and implemented.

The IRE Panel concurred with the BRD. No data were provided for multiple studies from a single laboratory. Neither intralaboratory repeatability nor reproducibility could be assessed. Quantitative interlaboratory reproducibility was assessed in two of the four studies, which used slightly different protocols. The recommendation is that reproducibility analyses should be conducted from studies using the recommended protocol and the approved list of reference substances.

The availability of historical negative and positive control data was appropriately covered. However, positive controls have not been consistently used. In future studies, this information should be tracked.

There appears to be no impediments to minor protocol changes or to transferability of the IRE test method. It may be useful to contrast results developed using the SafePharm recommended protocol *versus* earlier renditions. Good agreement across the board with *in vivo* data would suggest that existing data from all of the protocols could be used as validation data. Any differences in protocols used for future studies should be specifically justified.

8.0 IRE Test Method Data Quality

The IRE Panel concurred with the BRD. A lack of GLP compliance *per se* is not an exclusion criterion. Although not all studies were considered GLP-compliant, the reviewed data appear to be of satisfactory quality.

This was covered in the BRD adequately. Verification of accuracy of data against original data records is beyond the scope of the IRE assessment. The impact of deviations from GLP guidelines in Section 8.3, are appropriately covered. Noncompliance with GLP was not considered a mandatory exclusion criterion. All laboratories that performed the studies were considered reputable.

This was well covered in the BRD. The original raw *in vitro* data for all studies was not available for review. This data cannot be audited retrospectively. The quality of the institution, the reputation of the individual researcher, and evidence of reproducibility of results must guide our confidence in the accuracy of the data.

9.0 Other Scientific Reports and Reviews

This was adequately covered in the BRD. The submitted Procter and Gamble Enucleated Rabbit Eye Test (ExRET) and Low Volume Eye Test (LVET) data were not readily transferable to other studies for regulatory classification, and thus were excluded from the overall analysis. Reviews of all relevant published IRE studies were included in the BRD.

The IRE Panel felt that the conclusions reached on the report summaries were adequate and complete.

Appropriate measures were taken as described in the BRD. A *FR* notice was sent and study authors were contacted to request original IRE data and *in vivo* reference data. The IRE Panel acknowledged that obtaining this data was a difficult process.

10.0 Animal Welfare Considerations (Refinement, Reduction, Replacement)

This was appropriately covered in the BRD. It is important to determine the availability of rabbits to be used for this purpose. Furthermore, rabbits should not be raised or sacrificed specifically for use in this test. Currently, most U.S. Federal regulatory agencies do not permit prior use of animals for ocular testing. Therefore, the availability of eyes from an abattoir may be a factor for further development of this test method. The test method could be considered a partial replacement, under the 3R's, if eyes indeed were available.

11.0 Practical Considerations

The IRE Panel felt that transferability of the IRE test method could be readily achieved.

Training needs to be conducted with experienced personnel. Training videos and visual aids would be useful as discussed at other sessions.

This was adequately described. However, the cost for conducting the tests in each country should be obtained to reflect differences at the current exchange rates. For example, a laboratory in the United Kingdom may run the test with controls at $1074.00 using the current exchange rate for both *in vitro* and *in vivo* eye tests, but the actual cost in the U.S. may be significantly higher due to labor costs and other factors.

The *in vivo* test may take up to 21 days, whereas the *in vitro* IRE test takes up to 4 hours.

12.0 Proposed IRE Test Method Recommendations

The IRE Panel concurred with the BRD recognizing that the recommended version of the protocol was only conducted in one laboratory, and limited data were generated using that protocol. The most appropriate version of the IRE test method was selected using the additional endpoints identified.

The appropriate source of rabbit eyes needs to be defined. Currently, not all U.S. regulatory agencies will accept ocular data from studies in which the rabbits were used for other experimental purposes. The IRE Panel recommends that the U.S. practice be revised and updated, if possible.

Although the decision criteria for identification of ocular corrosives/severe irritants are defined, a rationale for them and/or a discussion of statistical algorithm(s) used in their development should be provided.

A standardized scoring system for histopathology should be more clearly defined to maximize the likelihood of obtaining reproducible results. In addition, reference photographs for all subjective endpoints should be developed to aid in training and transferability.

The recommended IRE test method appears to be capable of identifying ocular corrosives and severe irritants in a tiered-testing strategy. However, the database classifiable as GHS is small (n=36) and there is lack of data on reproducibility. In order to accept IRE data for classification purposes, the database needs to be expanded to corroborate the current results. However, it may be possible to look at existing data such as the CEC (1991) data to expand the database, since they used three of the four recommended endpoints and the accuracy was similar to that of Guerriero et al. (2004). The low false negative rate (0%) is encouraging meaning that either few or no animal tests would need to be performed. The false positive rate (33 to 34%) is relatively high. This false positive rate could be reduced by optimization of the decision criteria through employment of appropriate statistical methods (e.g., discriminant analysis).

Additional Expert Panel Discussion/Recommendations

- For consistency in histopathology assessment, it was recommended that the Panel request the Society of Toxicological Pathologists or NICEATM to set

up a standardized ocular histopathology grading system with visual aids and publish the information as soon as possible to provide uniformity in ocular tissue evaluation.

- To provide consistency among the BRDs, text in Section 8.4 concerning the need for laboratory notebooks containing the original raw data, Drs. Green and Freeman, modified the text to read, "…availability and review of raw data would improve the confidence in the data."
- In response to a concern regarding a potential issue of whether data used in the BRD should be subjected to a GLP audit, particularly if data was obtained from a single laboratory, Dr. Scala noted that in the ICCVAM documents the criteria for validation are presented (i.e., page 3) and indicate that, "All data supporting the assessment of the validity of the test method must be available for review." The final language should, therefore, be consistent with that statement.
- A change in wording was recommended for Section 6.3 of the IRE BRD to "Also recognize that the variability of the Draize test for corrosives or severe irritants is lower."
- Concern was raised that the IRE test method was not being endorsed for use in a tiered-testing strategy as the other test methods were and that there was a recommendation for further testing. It was pointed out that the test appears to be useful for identification, but the numbers to support it are low and the data needs to be corroborated. There was also concern that the data came from a single laboratory and no intra- or inter-laboratory reproducibility was available using the recommended protocol. However, it was stressed that the ICE data also came from a single laboratory, and this was not objected to.
- The current U.S. regulatory policy for acceptance of eye data from animals with prior treatment is inconsistent and it was recommended that this position be clarified and a proposal made to regulatory agencies to reconsider the use of such animals.
- Since consensus for use of the terms Prediction Model and Data Interpretation Procedure were not obtained at OECD or other meetings, Dr. Stokes suggested for now that the term "decision criteria" should be followed by "prediction model" in parentheses.
- Additional data on the IRE should be requested from other companies performing the test.
- The Panel continued to discuss specific points in the IRE BRD to harmonize wording across other BRD recommendations, where necessary and to reword some language of the proposals in IRE presentation for consensus agreement (See Expert Panel transcript for details).

Panel Vote on the IRE Report

The Panel Chair asked the Panel members to indicate whether they accepted the revised document that had been presented, with all the revisions that had been placed in it. A show of hands indicated that all Panel members, except for Dr. Stevens, accepted the revised IRE report. Dr. Stevens indicated stated that he voted no, because of his opinion that animals

should not be used for future validation studies, and because of his minority opinion on Section 4.0. Dr. Stevens indicated that he agreed with the Panel on all other parts of the IRE report. Dr. Guerriero abstained from voting, because he had a conflict of interest as a principal user of the test method.

Public Comment Session

1. Dr. Pauline McNamee (Procter & Gamble, Co; representing The European Cosmetic, Toiletry, and Perfumary Association [COLIPA])

Dr. Pauline McNamee congratulated ICCVAM on the tremendous amount of work, both in terms of scope and depth, which went into compilation of all four BRDs for this very important activity on eye irritation.

Members of COLIPA compiled a list of technical comments on the IRE test method. The first related to use of a statistical method for determination of the decision criteria that was adequately addressed in the Panel recommendations. COLIPA also welcomes consideration of the use of histopathology as an endpoint for consideration in the IRE, as well as in the ICE and BCOP test methods. This effort would build on the initial work by Jim Maurer and Rosemarie Osborne. COLIPA requested that the Panel consider the recommendations supporting the proposed IRE protocol on the number of eyes tested, use of concurrent positive and negative controls, or reference substances to reflect the practical limitations associated with the ability to perfuse and assess all of these eyes at any one time, and also in terms of the ability to appropriately time treatments, measurements, and other functional aspects of the test.

COLIPA recognized that application of the standardized protocol decision criteria to all of the datasets was problematic, because the number and type of endpoints varied from study to study and differences in scoring scales were used. However, COLIPA suggested that an effort be made to contact the authors of the studies to determine if data could be obtained for use in a weight-of-evidence approach for the evaluation of all data in the IRE BRD.

COLIPA reiterated previous public comments that it very much welcomes the continued clarity resulting from this meeting and in further development of the BRDs and encourages further efforts to ensure that 1) the most comprehensive data package available is used in the BRDs, 2) those data sources are used in the overall test evaluation, and 3) additional work needed due to the complexity of the protocols involved and limitations of existing data would be done after examination of the relationship between the experimental protocols, adjustment of the decision criteria, and subsequent interpretation of the data. COLIPA was strongly convinced that a weight-of-evidence approach could be applied to the post-hoc evaluation of these test methods. Furthermore, COLIPA welcomed a retrospective analysis to determine 1) what is needed to move forward, 2) identify specific research needs on mechanisms of chemically-induced eye irritation, and 3) lead to further optimized methods and/or new methods such as those currently being explored and researched in the COLIPA eye-irritation research program.

COLIPA closed by commenting that they stand ready to continue to collaborate with ICCVAM and ECVAM on this and future activities in the area of eye irritation.

Ms. McNamee indicated that Procter & Gamble has used farm-raised a rabbits as a source of eyes for an *in vitro* test for years and would be willing to provide information to ICCVAM on the suitability and acceptability of those eyes for use under the circumstances of removal, transport, and use in the laboratory, if that data is useful to them.

2. Ms. Sadhana Dhruvakumar (PETA)

Ms. Dhruvakumar expressed concern about use of the term accuracy, which she indicated was defined as "the quality of nearness to the truth or the true value." Regardless of the definition in the glossary of the BRD, the term accuracy implies that the rabbit data are the truth. Since we have come a long way from assuming that the animal data are the gold standard, our language should reflect that and use the dictionary definition of accuracy.

The second comment was related to agreement with comments from Ms. Sarah Amundson that the documents far overstep their bounds and led the process. The scope and content of the BRD should have been to present the data to the Expert Panel and allow them to consider that the method were 1) scientifically valid and useful today, 2) unscientifically valid in certain circumstances or with certain limitations, or 3) they are scientifically invalid and are not ready to be useful in any circumstance today. For example, any considerations of improvements or optimization should have been secondary to where these test methods stand in terms of usefulness. Instead, conclusions and additional optimization and validation studies were proposed for every method in the BRDs. The Panel was therefore focused on possible improvement, but did not draw clear conclusions on the current validity of the test methods. This was a step backwards for these test methods and a disservice to them.

These methods have been in use for approximately 20 years. Furthermore, these tests have been accepted for use in a tiered-testing strategy by some European countries for approximately a decade, and are currently accepted by all EU countries because of mutual data acceptance. Several have had extensive laboratory validation studies of them. These methods have stood the test of time, which is also the only test of accuracy that has been applied to the *in vivo* rabbit eye test. These *in vitro* tests have already proved their utility. The fact that these tests were evaluated for use in identification of ocular corrosives/severe irritants should have been a "slam-dunk." If the Panel cannot agree that even one of these test methods is valid for use as a partial replacement right now, without years of additional work, there is little hope of ever getting to a point where mild irritants can be assessed to provide complete replacement of the Draize test. If the process is perceived to take these tests backward, rather than forward, ICCVAM will not be in a position to receive new nominations.

PETA requests that the Panel consider whether or not these tests are scientifically valid and potentially useful as a positive screen for ocular corrosives and severe irritants in any definable set of circumstances, and deemed to be valid for a specific purpose. Also, existing data should be used retrospectively and combined with the scientific judgment of the Expert

Panel to validate this test method. If none of these test methods are judged to be valid for current use, faith in the ICCVAM process will be significantly affected.

3. Dr. Robert Guest (SafePharm Laboratories)

Although an invited expert on the IRE test method, Dr. Guest asked to have time to provide public comment. He indicated that a comment made in a previous session on the statement that use of coded or blinded substances was a GLP requirement and he clarified that, as far as he was aware, it was not.

Dr. Guest commented on the use of rabbit eyes from animals that had been used previously for other tests. Typically, at SafePharm laboratories, rabbits used for skin testing of mild or nonirritant substances are routinely used as a source of eyes in the IRE. Dr. Guest asked that consideration of the use of these animals, as unacceptable sources of rabbit eyes, should be reconsidered. He noted that numerous controls are in place. The animals are allowed sufficient time to recover, certain limits or conditions for *in vitro* use are imposed, and the eyes are carefully examined by slit-lamp *in vivo* and again after enucleation. These eyes remain viable and in good condition for the test and there is data available to support this conclusion. There are certain substances that may have delayed effects on eyes or exhibit systemic toxic effects on local exposure. However, safeguards to avoid use of such substances are in place. Any corneal effects produced by such substances would be identified on examination. The advantage of this method is that the eyes are fresh and animals are spared.

With respect to false positives, the IRE may identify ocular corrosives and severe irritants or even some less irritant substances may be overlabelled. This does not mean that a nonirritant will be classified as an irritant. The company that submits the data for testing must eventually decide on labeling, the level of tolerance they have for false positives, a decision often based on additional weight-of-evidence information.

Yesterday, there was a question pertaining to why more laboratories do not perform the IRE test. Several laboratories do run the IRE test and at least one runs the ICE test as well, but generally the regulators do not ask for the data. If regulators asked for the data, companies would perform the test. If the test is not required, few companies will apply resources unless it is necessary. There is a commitment in most laboratories to reduce the use of animals that drives the development of alternatives. There are pressures to reduce animal use and efforts are underway in the EU to change legislation due to the chemicals-notification programs, such as the Registration, Evaluation and Authorisation of Chemicals (REACH) system, and the 7th amendment to the Cosmetics Directive, to require use *in vitro* tests for testing of these products. Something has to be done now, and this meeting has given us a fantastic opportunity to look at these methods and approve them. Is there a regulatory barrier that would preclude parallel use of an *in vitro* test method with other data to extend or expand the database for at least certain types of products?

Dr. Scala noted that one comment made on regulatory barriers and another regarding classification of false positives were beyond the scope of this Expert Panel, although regulators certainly will consider the implications.

4. Dr. John Harbell (IIVS)

Dr. Harbell noted that the Expert Panel might be considered the supreme court of science, since what is done here and the statements made about how the tests are evaluated will be the models for how this is done in the future. It is the Expert Panel at this public meeting that provides the checks and balances for the executive branch, the ICCVAM/NICEATM. In 1992, a similar meeting was held to design submission criteria for the IRAG program with Drs. Bradlaw, Gupta, Green, and Wilcox. In that meeting, Dr. Scala stressed the importance of continuous evaluation of data rather than categorization, and the importance of looking at individual animals and their performance. Then in 1996-1997, we met as stakeholders to design the ICCVAM program. Before federal funding, this was an informal committee, but the paradigms for validation were laid down at that time by all of the stakeholders. In that document, relevance was defined in the glossary as the extent to which a test is meaningful and useful for a particular purpose, that is the extent to which a test method will correctly predict or measure a biological effect of interest. Although animal welfare is a noble cause, it is not the only driving force for development of an *in vitro* test method. Dr. Scala's work in the predictive capacity of the Draize test, and subsequent work by Dr. Marzulli in 1973 indicate that the predictive capacity of the Draize test is limited. For example, there is a phrase in the paper by Dr. Marzulli, "Furthermore, collaborative results indicate that additional study to identify and eliminate sources of variability is necessary before reproducible results with regard to comparison of degrees of irritancy can be obtained." We have moved from relevance to using the word "accuracy" that leads to a fundamental change. This change is reflected in Criterion 4 in the BRDs that says "the potential for the proposed test method to provide improved prediction of adverse health or environmental effects compared to the current test methods accepted by regulatory agencies." This is a criterion for consideration by ICCVAM. The response of ICCVAM/NICEATM is, "It is proposed that the current animal test provides a suitable assessment for eye-irritation potential in humans." There is not complete agreement on this statement.

The last point regards a paper called *Ophthalmological Perspectives on Eye Irritation Testing*, which says, "We note that two major themes should permeate all future work to further development of alternative tests. First, we unanimously agree that the Draize rabbit eye test method as currently used should not be considered the primary standard for the evaluation of new test methods." Dr. Harbell noted that two of the authors, Drs. Chambers and Edelhauser were members of this Panel. He asked that the Panel consider these points of view and add them to the deliberations.

IV. HEN'S EGG TEST – CHORIOALLANTOIC (HET-CAM) TEST METHOD EVALUATION

Primary reviewers: Drs. Shayne Gad, Donald Fox, Martin Stephens, Frederick Guerriero, Sidney Green, Philippe Vanparys, Nancy Flournoy

Overview of the HET-CAM Test Method Procedure

Dr. Klaus Krauser provided an overview of the HET-CAM test method. Dr. Krauser discussed that the test method was first proposed by Professors Luepke and Kemper both from Germany. In 1988, a validation project was started in Germany, and funded by the government of the Federal Republic of Germany, to evaluate the validity of the HET-CAM test method. In 1992, the protocol used in this validation project was published in INVITTOX (Protocol No. 47), which is in the ERGATT/FRAME databank. Dr. Krauser noted that there had been other validation efforts conducted over the years, as well.

Dr. Krauser stated that a defined dose of a test substance is applied to the chorioallantoic membrane (CAM) of fertilized and incubated hens' eggs and the CAM is then evaluated for up to 300 seconds for the development of defined endpoints. Dr. Krauser then reviewed the evaluation of endpoints either involves determining the time elapsed until the first appearance of the endpoints or the severity of each endpoint is evaluated after a certain time.

Dr. Krauser then provided a brief review of the CAM. He stated that the CAM was a vascularized respiratory membrane that surrounds the developing bird embryo. It is composed of three parts. Dr. Krauser stated that the blood vessels that are present in one of the CAM layers are branches from the embryo-allantoic arteries and veins and they form a capillary bed. He then reviewed drawings to further describe the locations of the various parts of the chicken embryo.

Dr. Krauser stated that most current HET-CAM protocols used White Leghorn hen's eggs that were fresh and not older than 7 days. The eggs are fertile and clean and between 50 and 60 grams. Eggs are typically candled prior to use to ensure viability. Dr. Krauser stated that commercially available incubators with an automatic rotating device are used and that the eggshells are usually opened with small saws or dentist rotary saws.

Dr. Krauser stated that the historical negative control substances were typically saline solution, or other vehicles. Positive control substances were sodium hydroxide or 1% sodium dodecyl sulfate.

Dr. Krauser then reviewed the preparation of the test system. He discussed the age of the eggs used, the incubation temperature, the relative humidity, rotation of the eggs, removal of the eggshells, treatment with the test substance, test substance volume, exposure duration, and endpoint evaluation. Dr. Krauser noted that of all the endpoints that could be evaluated, the evaluation of hyperemia was the most variable. Dr. Krauser stated that the endpoints currently evaluated in the HET-CAM test method are hemorrhage, lysis, and coagulation. He

then provided additional detail about the observations made for each of these endpoints during conduct of the test method.

Dr. Krauser then provided detailed information on the measurements of the endpoints, number of replicates used in the conduct of the test method, number of repeat experiments conducted during studies, calculation of irritancy potential and/scores, decision criteria, and acceptance criteria.

Summary of the HET-CAM Test Method Background Review Document

Dr. Choksi (NICEATM) described the analysis of the HET-CAM test method presented in the HET-CAM BRD. She indicated that ICCVAM agencies were surveyed and to the best of their knowledge HET-CAM data had not been submitted to U.S. regulatory agencies for ocular irritation purposes. Dr. Choksi stated that there were 10 studies that were evaluated in the BRD. Of those studies, there was no information on intralaboratory reproducibility. Three different studies were used to evaluate interlaboratory reliability. Additional data had been received since the draft BRD was published. The information would be included into the revised HET-CAM BRD. In addition to the 10 studies, there were 39 additional studies that were not evaluated because comparative *in vivo* data were not available or the *in vitro* data was qualitatively described. These reports were described in Section 9 of the BRD.

Dr. Choksi stated that there were five different analysis methods that have been described in the literature in the 10 studies evaluated. In the HET-CAM BRD, there were 246 substances evaluated in 253 tests. Most of the substances tested by the IS(A) and IS(B) analysis methods were formulations while the other three analysis methods evaluated mostly chemicals or pharmaceutical intermediates. Chemical classes evaluated, where there were at least three substances or more per class, were alcohols, carboxylic acids, amines, and formulations. Product classes evaluated, where there were at least three substances or more per class, included cosmetics, solvents, hair shampoos, and soaps and surfactants. Dr. Choksi provided a breakdown of the number of testing laboratories that tested the substances.

Dr. Choksi provided variations between the testing protocols used by the studies. Variations included differences in incubation time, temperature and humidity, the amount of volume tested on the CAM, whether the substance was rinsed from the CAM, and the endpoints evaluated. Endpoints evaluated in the studies described in the HET-CAM BRD included hemorrhage, lysis coagulation, hyperemia, and dilation.

Dr. Choksi reviewed of the accuracy of the HET-CAM test method, when compared to the GHS classification scheme. Of the four analysis methods described (IS(A), IS(B), Q-Score, and S-Score), the IS(B) analysis method appeared to be the most accurate. Using the IS(B) analysis method, accuracy statistics versus GHS, EPA, and EU were provided. Dr. Choksi stated that an additional analysis had been conducted the week prior to the Expert Panel meeting, which was given to the Panel as well as was provided to the public. Data in the third phase of the ZEBET evaluation of HET-CAM was evaluated based solely on the irritation score of substances evaluated. The concentrations tested in this evaluation were 10% and 100 %.

Limitations of the accuracy analysis were then discussed; these included:

- the impact of the differences in test method protocols between the studies reviewed was unknown
- the impact of the different end points evaluated in the studies reviewed was unknown
- the lack of severe irritant test substances evaluated
- the lack of solids evaluated
- the limited chemical classes and product classes evaluated

Dr. Choksi then reviewed the reliability described in the BRD. Only one interlaboratory data set with the IS(B) analysis method was available for evaluation. Qualitative and quantitative evaluations of these data were described. Limitations of the reliability analysis in the draft HET-CAM BRD included lack of intralaboratory reliability information and the interlaboratory data was based on a very small number of substances.

Dr. Choksi stated that the standardized version of the HET-CAM test method proposed in the BRD is similar to the one that was provided by ZEBET (INVITTOX 1992), which uses the IS(B) analysis method. The proposed test method protocol requires the use of positive and negative controls. Some proposed additional optimization studies in the HET-CAM BRD included retrospective analysis of the decision criteria used to identify corrosives and the evaluation of additional endpoints, such as trypan blue, which might provide some quantitative information compared to the qualitative currently obtained from the test method.

After the conclusion of the presentation, Dr. Scala requested questions from the Panel. Dr. Green questioned whether there was any data provided to support the states that the hyperemia endpoint was not reliable for use in the test method. Dr. Choksi replied that the information was provided by personal communication and she did not have data that supported that statement.

Proposed Panel Recommendations for the HET-CAM Test Method

1.0 HET-CAM Test Method Rationale

Dr. Green presented the draft recommendations developed by the primary reviewing group ("the Group") for the HET-CAM test method for consideration and concurrence by the full Expert Panel. Dr. Green stated that the Group felt that the CAM was most similar to the conjunctiva and that a deficiency of the method was that the CAM has no structures similar to the iris or cornea. He noted that the method currently was most useful for assessing short-term effects of substances. Dr. Green stated that the ability of HET-CAM to provide improved prediction of adverse health effects in humans would be difficult to achieve unless comparative data for the *in vitro* test method, animals, and humans was generated using substances from a repository. The Group proposed that the method may have the potential to complement others in a tiered testing approach.

The Panel discussed the ability of the HET-CAM test method to assess substances that induce delayed effects in rabbits or produce reversible effects *in vivo*. The Panel proposed that even though the method cannot directly evaluate those things, that a recommendation be included that suggested exploring the relationship between the short-term effects observed in the HET-CAM test method and the long-term effects in the eye. The Panel agreed to include the recommendation into the Panel report.

The Panel also proposed to revise text in the report that stated that the CAM had no structure similar to the iris or cornea. It was discussed that that the CAM did not contain structures similar to the cornea. It was originally proposed to remove the term "cornea". Panel discussion followed on the topic that there were two issues related to the similarity of the CAM to the structures of the eye: vasculature and avascular tissue. The Panel then proposed to revise the statement so that the phrase read that the CAM tissue structure was not similar to the cornea.

2.0 HET-CAM Test Method Protocol Components

Dr. Green stated that the Group believed that the recommendations in the draft BRD appeared to best standardize test method procedures among the various published literature sources and developed a consistent scoring and calculation of irritation indices. The Group agreed with the BRD recommendations on (1) the strain, stock, and age of recommended eggs, (2) the number of replicate and/or repeat experiments, (3) development of a definition of a positive result, and (4) development of a recommended protocol. The Group also agreed that there were no proprietary components of the method. Dr. Green stated that the response to the basis for any modification to the original HET-CAM test method protocol needed to be developed by the Panel.

3.0 Substances Used for Previous Validation Studies of the HET-CAM Test Method

Dr. Green continued that the Group felt that the type and number of substances tested the studies evaluated in the HET-CAM BRD was adequate. The group also stated that it was difficult to determine if coding procedures were appropriate. They noted that not enough information provided was in all studies to make a full assessment. The Group then proposed that as long as the quality and multiplicity of data sources was sufficient to draw conclusions, coding did not matter.

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

Dr. Green then discussed the *in vivo* data that used in the HET-CAM BRD analysis. The Group proposed that the *in vivo* test method protocols used to generate data used in BRD were appropriate and that interpretation of *in vivo* rabbit eye test results was correct. There was concern expressed by the Panel that the regulatory classification methods may be less than adequate for use in evaluating or making distinctions between *in vitro* methods and their suitability for chemical or product class evaluations. The Group agreed that since original study records were not available for any of the reports evaluated, data quality could not be determined. However, the Group felt that an evaluation of the results could be made and the

quality of the studies otherwise appeared to be adequate. The Group stated that not all studies evaluated for the HET-CAM test method identified the studies as being conducted in compliance with GLP guidelines. The Group noted that there was not a large database on human ocular injury data and that most of the available information originates from accidental exposure. The Group recommended that COLIPA and the International Life Sciences Institute (ILSI) be consulted for information on human eye irritation databases to assess current ocular hazard classification scheme adequacy in protecting human health. The Group stated that the potential variability of the *in vivo* rabbit data had not been adequately discussed in the HET-CAM BRD.

The Panel discussed the use of the phrase "spirit of GLP". The appropriate use of the phrase and the definition of such phrase were discussed. Several Panel members indicated that the use of the phrase generally indicated that most, but not all, GLP guidelines were complied with during the course of conducting a study. The items within the guidelines that may not have been complied with each time may not be the same. However, the lack of compliance with specific items should be identified.

5.0 HET-CAM Test Method Data and Results

Dr. Green stated that the Group noted that the test method protocols used to generate data were adequately described in the BRD. The Group recommended including a description of which type of irritation score (IS(A) or IS(B)) was evaluated by each study in BRD Section 5.4 Additionally, the Group stated that data generated by Gautheron et al. (1994) may be useful in the BRD development. The Group recognized that the lack of consistent evaluation methods complicated BRD evaluations, but that the IS(B) appeared to be the optimal approach. It was proposed that data censoring could be an issue for the method. The Group stated that the BRD documents use of coding and GLP guidelines were adequately discussed and that there was insufficient information on lot-to-lot consistency in studies reviewed.

The Panel discussed a proposal in Section 5.3 that indicated that censoring was an issue for the HET-CAM test method. It was originally proposed that censoring methods would permit use of partial data obtained from studies not fully completed and that development of methods to use partial data should be considered. Several Panel members were unclear as to the intent of the statement. Dr. Lovell indicated that the statement could either refer to making use of incomplete studies or using a "meta-analysis" to combine across a number of different studies to bring them all into sort of a similar framework. There was concern by Panel members that the language was vague and the meaning of the statement was unclear. Panel members attempted to reword the statement to increase clarity. Panel members then proposed to remove the text from the document. A show of hands of Panel members indicated support to remove the text of concern.

The Panel then discussed the statement in Section 5.3 that stated that the approaches used in the evaluation of data were appropriate. Panel members expressed concern that the statement indicated that the Panel members accepted the statistical analysis and non-statistical approaches of evaluation of the data. Members indicated that there were alternative ways to

analyze the data. Initially it was proposed to delete the statement from the document. Upon further consideration, Panel members proposed to add text that indicated:

> *That the approaches used to evaluate the HET-CAM data appear to adequately describe the accuracy and reliability of the test method. However, given the unavailability of original HET-CAM data, a definitive statement regarding the adequacy of these approaches is not feasible.*

There was unanimous agreement on the modification.

6.0 HET-CAM Test Method Accuracy

Dr. Green continued to discuss the accuracy calculations described in the BRD. The Group stated that it was essential to identify structural classes the test method works well for and poorly for. The Group also recommended replacing the term “accuracy” with “concordance” or “agreement” since HET-CAM accuracy was not being evaluated directly against human data but against the *in vivo* test. The Group stated that the limitations of the analysis were discussed in the BRD. The Group also noted that there was *in vivo* data variability and in cases where false positives and negatives were noted, variability of the *in vivo* responses should be reviewed. The Group agreed that data interpretations were sufficiently described in the BRD and the organization of the document identified the IS(B) method as being the best in identifying most ocular corrosives and severe irritants. The Group noted that the use of a standardized test method protocol was needed to produce more interpretable data. Additionally, the Group recommended development of a table with non-accepted studies (HET-CAM BRD Section 9.0) to evaluate outcomes of these studies.

The Panel discussed the issue of replacing the term "accuracy" with “concordance” or “agreement”. The Panel noted that this issue was relevant to all the test method BRDs. Dr. Stokes stated that the ICCVAM submission guidelines state:

> Accuracy is defined as the closeness of agreement between a test method result and an accepted reference value. The term is often used interchangeably with concordance.

The term “concordance” is defined, in the same document as,:

> The proportion of all chemicals tested that are correctly classified as positive or negative. The term is often used interchangeably with accuracy

Dr. Stokes stated that to ensure consistency in the documents, the term accuracy was used in these reports. Dr. Scala requested discussion on whether the recommendation of replacing the term “accuracy” with another term should be carried to all the BRDs or deleted. A topic of Panel discussion was to whether the term concordance was more descriptive of the analyses presented in the BRDs. Dr. Stokes pointed out that the definition of accuracy states that the closeness of agreement can be between a test method result and a separate reference value. Additional discussion on the topic followed. A vote was taken as to whether the text

should be changed. The proposal to change the text was voted down by a show of eight hands to twelve.

7.0 HET-CAM Test Method Reliability (Repeatability/Reproducibility)

The Group noted that the rationale for substance selection was primarily based on data availability and the quality of the *in vivo* data was a limitation of all studies used in the BRD. The Group noted that even though there were no intralaboratory repeatability and reproducibility evaluations due to lack of data, it should not be a roadblock for use. The Group indicated that the topic of interlaboratory variability was well addressed in the BRD. However, the use of %CV was not an optimal approach and that non-parametric evaluations would be preferred. Additionally, there were several general items for revision noted. The Group agreed that there was an absence of historical data. The Group recommended the use of data from non-accepted studies (BRD Section 9.0) as a source of control data. The Group agreed with the BRD that appropriate recommendations were made for the selection and use of positive and negative controls and that the effect of protocol changes is unknown without having more standardized studies with measures of variability. The Group recommended that a video on the method and scoring be developed to increase test method transferability.

The Panel discussed the potential use of non-parametric analyses to evaluate interlaboratory variability. The Panel noted that the %CV assessment has limitations when evaluating a narrow range of scores. It was recommended to revise the text to indicate that alternative approaches for measuring agreement could be used for evaluating reproducibility. The remainder of the Panel agreed to the revision.

8.0 HET-CAM Test Method Data Quality

Dr. Green stated that the Group agreed with all the items in this section of the BRD. They stated that failure to use GLP guidelines by studies may have had qualitative impact only on borderline classifications. The Panel recommended including coding information (BRD Section 3.4) in Appendix B2.

9.0 Other Scientific Reports and Reviews

Dr. Green stated that the Group agreed that the BRD provided a comprehensive review of available publications made in BRD and adequately and completely provided the conclusions of the publications. The Group noted that it was useful to have information on why the studies were excluded from the evaluation in BRD Sections 6.0 and 7.0. The Group proposed that criteria for data acceptance could be relaxed to allow more studies to be evaluated in the BRD and that information on whom was contacted for additional data and who did and did not respond be included in the document. The Group also proposed that a preferred list of compounds be generated for distribution to companies and a request for additional data on these compounds be provided.

The Panel questioned what types of data were being requested. Dr. Tice (NICEATM) responded that the request was for any kind of data (e.g., *in vivo* human or rabbit data, *in vitro* data) that could be used for evaluation in the BRDs.

10.0 Animal Welfare Considerations

The Group stated that additional discussion on some issues was needed. The Group proposed to define the test as "*ex vivo*" and not "*in vitro*". The Group also proposed that this section should discuss the tiered testing strategy that animals would only be used to confirm negative response, and that HET-CAM should be considered a "partial replacement". The Group also proposed that additional discussion should state: (1) no new animal testing should be conducted, (2) the low false negative rate has the advantage of reducing the exposure of animals to severe irritants and corrosives, and (3) any additional optimization should focus on decreasing the false negative rate.

The Panel discussed the terms "*ex ovo*" and "*in vitro*". It was indicated that the term "*in vitro*" was used broadly to encompass "*ex ovo*". Dr. Stokes stated that subcategories of "*in vitro*" could be made, like "*ex ovo*" and perhaps the subcategories could be included into the glossary. There was no dissent by the Panel on that proposal. The Panel proposed to revise the first proposal to state that in the interest of consistency between the BRDs that the term "*in vitro*" should be used.

11.0 Practical Considerations

Dr. Green stated that the Group agreed with all the statements made in the BRD.

The Panel discussed that use of a training video on the technical aspects of the assay could be applicable to all the test methods being discussed. It was recommended that the language be modified to state that the Panel recommended the use of a training video and development of training approaches in the application of this test method. The Panel agreed that similar language should be included in all the test methods being evaluated at the meeting.

12.0 Proposed HET-CAM Test Method Recommendations

The Group discussed the proposal of using a modified HET-CAM test method protocol to confirm positive results obtained in the proposed, standardized HET-CAM test method protocol provided in the BRD. Discussion as to what type of modifications should be made to the test method protocol to identify potential false positives followed. Inclusion of *in vitro* test method data with various test concentrations into the BRD was proposed. It also was proposed that the report be modified to state that HET-CAM data exists to evaluate the use of a lower concentration and such information and analysis of the data should be included into the BRD.

The Panel then discussed revising or including language similar to the ICE test method report to indicate that even though the Panel recommended the current version of the HET-CAM test method could be used to classify substances as corrosives and/or severe irritants that the

Panel also recommends that the test method could be optimized and validated. The Panel discussed the proposal but determined that the current language in the HET-CAM report was clear since it indicated that the test method, except for test method procedures, were appropriate. The consensus agreement was to leave the language as it currently read.

Panel Vote on the HET-CAM Report

Dr. Scala asked if any individuals on the Panel had a conflict of interest with regard to this method. No hands were raised.

Dr. Scala asked for a vote on whether the Panel concurred with the HET-CAM information and results as presented. There was a show of 19 hands concurring with the presentation with three abstentions.

Dr. Scala asked if any of the three abstentions were for reasons of conflict of interest or for minority opinions. Dr. Stephens stated that his abstention was because of a minority opinion. His minority opinion was the same as for the ICE report regarding whether or not additional animal testing is discouraged or encouraged. Dr. Theran stated that his abstention was because of a minority opinion and was related to the use of "accuracy" in the document. Dr. Theran's opinion was that an *in vitro* test could be more accurate in representing the human experience and yet be not in agreement with the *in vivo* rabbit eye test and therefore did not agree with the use of the term "accuracy" when comparing it to the rabbit eye test. Dr. Yasuo Ohno, stated that the text for Section 8.1 was unclear. He felt that the sentence should state that the lack of GLP compliance would be compensated by using coded samples and appropriate data handling.

Dr. Stokes asked for clarification as to whether those with dissenting views agreed with the rest of the recommendations and conclusions except for the specific item that they expressed a minority opinion on. Dr. Stephens noted that he also dissented on the use of the term "accuracy" and that would be part of his minority opinion. However, there were no other issues that he dissented on. Dr. Theran stated that he did not dissent with any other issues related to the document. Dr. Ohno stated that he did not dissent with any other issues related to the document.

Dr. Fox stated that he believed that they should discuss the topic that Dr. Ohno presented in his minority opinion and potentially modify the statement. Dr. Scala stated that the topic would re-opened for discussion for no more than three minutes. Dr. Spielmann stated that he agreed with the statement, as proposed to be revised by Dr. Ohno. Dr. Scala requested that a Panel member propose that even though the vote was closed that it be re-opened and use Dr. Ohno's language as replacement for the language was present. Dr. Fox so moved. Dr. Scala asked for a show of hands to indicate concurrence with the text as changed. There were no dissenting votes.

Public Comment Session

1. Dr. Dan Marsman (Procter & Gamble)

Dr. Marsman discussed that the *in vivo* data are always requested alongside the *in vitro* methods but they weren't directly called for as part of this assay review and as such there wasn't a complete submission of *in vivo* data. Dr. Marsman stated that he believed that this hampered part of the discussion of these *ex vivo* methods and that submission and inclusion of that *in vivo* data would alter the interpretation of some of the *in vitro* results.

Dr. Marsman stated he believed it was critical that the *in vivo* methodology, its human equivalent, and its quality be reviewed to credibly evaluate the *in vivo* method and then the *ex vivo* method. Since the methods were being evaluated for their utility and classification and labeling for regulatory purposes all relevant *in vivo* data, specifically data from the LVET, should be included. Dr. Marsman stated that the LVET test method represented a minor modification of the traditional Draize method and that it has been accepted for classification and labeling purposes in some regulatory contexts. He stated that the minor modifications of the method yielded improvements. These improvements included refinement of pain and distress in the *in vivo* assay and the relevance to the human experience.

Dr. Marsman stated that all methods should be evaluated in the context for which they are being proposed and the historical LVET data and its associated *in vitro* data sets should be included in the evaluation. Inclusion of such data would likely alter the sensitivity and specificity scores of some of the *ex vivo* methods evaluated. Dr. Marsman concluded that the individual raw historical data for the LVET and the compiled data on the LVET as well as some of the mechanistic information on the pathophysiology, histopathology of the ocular toxicity is data that could further be submitted.

2. Dr. Rodger Curren (IIVS)

Dr. Curren requested that the Panel consider, when drafting the final Expert Panel report, strive for comparability in language between the BRDs for each test method of these as well as clarity of the final conclusions for each evaluated test method. Dr. Curren stated that there were a number of statements made in the Panel presentation on HET-CAM that dealt with many of the same problems or issues that were discussed in the ICE test method. However, Dr. Curren stated that the discussion of them was in many cases was more positive for HET-CAM than for ICE. Therefore, Dr. Curren requested that similar terms be used for similar ideas between the different test methods.

Dr. Curren than noted that Dr. Stokes previously stated that institutional animal care and use committees would have a significant role in how these tests are used. Dr. Curren stated that if at the end of the Panel deliberations there was a positive conclusion, such as stating that the test method could be used as the first stage of a tier-testing process and if it finds the positive results the material can be so labeled, that a statement should be included earlier in the document.

V. REFERENCE SUBSTANCES FOR USE IN VALIDATION STUDIES

Primary Reviewers: Drs. Ih Chu, Sidney Green, Yasuo Ohno, Robert Peiffer

Summary of the Recommended Reference Substances in each BRD

Dr. Allen (NICEATM) presented a summary of the list of reference substances included in each BRD. He summarized the selection criteria for reference substances as outlined in the ICCVAM Submission Guidelines (ICCVAM 2003), which states that to the extent possible, reference substances should:

- Be representative of the range of responses for which the proposed test method is expected to cover (i.e., nonirritants and mild, moderate, or severe eye irritants)
- Represent the classes of chemicals for which the proposed test method is expected to be used.
- Have produced high quality results (i.e., produced in a GLP compliant study) in the reference test method and/or the species of interest
- Have well-defined chemical composition
- Be readily available
- Not be associated with excessive hazard or prohibitive disposal costs

Dr. Allen proceeded to describe that the reference list covers the range of anticipated responses, based on irritancy classification according to the GHS. He highlighted the limitation of the available database that only a fraction of the substances for which *in v*ivo data had been obtained were also commercially available. Also described were the GHS Category I subcategories that have been developed by NICEATM. These subcategories were established to delineate among severe irritants/corrosives based on the type and severity of lesion upon which an irritancy classification was assigned and are as follows:

- Subcategory 4: substances that induce a corneal opacity of 4 at any time in at least one out of three animals.
- Subcategory 3: substances with a positive response, based on mean, within the first three days in two out of three animals, and a persistent lesion (i.e., a response that is severe, as well as persistent).
- Subcategory 2: substances with a positive response, based on mean, within the first three days in two out of three animals that was reversible (i.e., it was not persistent).
- Subcategory 1: substances classified as severe based only on a positive response at day 21.

Dr. Allen then detailed the five different *in vivo* data sources for the substances included in the reference list (the CTFA, the ECETOC, the EPA – TSCA, the FDA, and the Japanese National Institute of Health Sciences).

Finally, Dr. Allen summarized the 25 different chemical classes and 30 different product classes represented among the list. He also indicated the range of responses represented by each chemical and product class, based on GHS ocular irritancy classification.

Proposed Panel Recommendations on Reference Substances

12.4 Recommended Reference Substances for Validation Studies

Dr. Scala summarized the conclusions and recommendations of the primary reviewers of the list of reference substances. He indicated that one person from each of the test groups was included as a primary reviewer of the reference substances.

Dr. Scala then stated that the list of recommended substances is comprehensive in that the three major groups of products to which the eye is exposed (i.e., industrial chemicals, pharmaceuticals, cosmetics) are represented. He also stated that the substances appear to be readily available and in acceptably pure form, and the range of possible ocular toxicity responses in terms of severity and types of lesions appears to be adequately represented. Dr. Scala recognized that the selection of reference substances is in part limited by the availability of *in vivo* reference data. He then detailed the following comments and recommendations for the list:

- The current list has entirely too many substances and is unwieldy.
- Surfactants are over-represented and correspond to an area where the panel can make selective recommendations.
- The list appears to have too few inorganic substances; more should be added to the list if feasible.
- Classification data for each *in vitro* test should not be included in a list of test substances that are proposed for validating *in vitro* tests; this information should be removed from the list.
- Colored substances that might interfere with the observation of the endpoints should not be included.

Dr. Scala then described an approach to determine the most appropriate numbers and types of substances that should be included in the reference list. He described a two-staged study design to validation studies. During the first stage, a small number of substances from a wide range of chemical classes and spanning the range of severe irritancy should be tested among several laboratories to assess reliability. He stated that substances selected for this stage should:

- have an applicable pre-existing *in vivo* database
- cover a broad range of chemical classes that are representative of substances that are most likely to come in contact with the eye (e.g., acids - organic and mineral; alkalis; amines, imines, and amides; alcohols (including polyols); ethers; esters; thiols; halides; quaternary ammonium compounds; N- and S-heterocyclics; and hydrocarbons)
- encompass the range of GHS Category 1 responses (i.e., GHS Category 1 subcategories described above)
- include a reasonable range of molecular weights, but no formulations, prototypes or products should be included

- include only liquid substances as these represent the majority of chemicals in the "real world" that will come in contact with the eye (using only liquids minimizes the inclusion of additional variables in the first stage of validation)

If deemed adequately reliable, an expanded set of substances would be tested in a second stage that would include multiple representatives of each chemical classes, diverse physicochemical characteristics, and the full range of irritancy responses to assess accuracy. Substances included in this stage should include:

- multiple representatives from each chemical class
- multiple representatives from each GHS Category 1 subcategory
- within each chemical class, compounds of different physical properties (solubility, molecular weight, pH) where feasible

Dr. Scala noted that for all validation studies, Material Safety Data Sheets (MSDS) for the recommended substances should be provided (e.g., a coded MSDS), in concert with a prestudy safety briefing

During the Panel discussion, Dr. Stitzel voiced concern over including a large proportion of Category 1 chemicals that were classified based only on a persistent response, given the fact that these "less severe" substances represent the area for which the Draize test is most variable. Dr. Spielmann concurred and stated that no substances classified based on a single animal should be included. Dr. Allen noted that some of these substances may have been classified based on all three animals showing the same response, which could potentially alleviate some of these concerns of variability.

Dr. Spielmann also voiced concern over the fact that the function of the list was not clear. He questioned if all substances were intended for use, or only a subset. He stated that the list should be preceded by a statement of precisely what the list was to be used for. Dr. Stokes responded by citing Section 12.4 of each BRD and the fact that a statement is included that the list has multiple purposes. He stated that for a full validation study, a large set of substances would be used. However, for performance standards, a smaller subset of the list could be used, and for proficiency chemicals, which a laboratory can use to demonstrate its proficiency with a test method, even a further subset of the reference chemicals in performance standards could be used.

Dr. Stephens stated that with regard to the BCOP, this validation effort might be satisfied by the submission of additional historical data. Therefore, he believed that with respect to BCOP, there may be a way to avoid a complete validation study.

In reference to the proposed two-staged validation study, Dr. Stitzel recalled earlier discussions among the Panel in which they concluded that validation against a grading system (i.e., classification) was not appropriate. She stated that a new method should be evaluated against a ranking of the severity of each chemical, rather than trying to evaluate new methods based on grading systems, which are very difficult even for the *in vivo* test to get right every time. Therefore she questioned why the substances were apparently still being selected based on a grading system. Dr. Scala responded by summarizing the selection

criteria used to generate the outline of the types of substances recommended by the Panel. He stated that the initial intention was to select from the list of the 50 most common industrial chemicals (which would serve as a surrogate for exposure). However, *in vivo* data were available for only five of these chemicals and therefore a list of the most common chemical classes was constructed from which a list of substances could be derived.

Dr. Stokes raised the issue of whether substances known to induce severe lesions, *in vivo*, in the eyes of humans should be included, even in the absence of rabbit data. The Panel agreed that such substances should be included.

Dr. Spielmann noted that in order to compare the different tests, it would be nice if a few standard positive controls or reference chemicals were available that could be used across the tests, so that a comparison of the efficiency of the tests could be conducted. He noted that there has been no agreement on what is acceptable as a positive control. Dr. Tice responded that each BRD proposed the type of positive control that would be useful. He noted that the positive control should be a severe, but on the very borderline or just slightly above the borderline between severe and nonsevere, and that it be either liquid or solid, based on what is being tested. However, because historical positive-control data were not available, reproducibility of such substances were not known, and therefore it would be premature to identify a specific substance that should be used all the time.

Dr. Robert Peiffer queried if it was ICCVAM's intention to use the Panel's recommendations to compile a final list of substances. Dr. Stokes responded that the intention is indeed to recommend a list, taking into consideration the Panel's advice, as well as advice received in public comments and will also receive in public comments on the Panel's report. Dr. Scala clarified that the Panel was not voting on approval of the list, rather on the outline of how to revise the list.

Panel Vote on the Recommended Reference Substances Report

A vote of concurrence on this section was taken and all members except Dr. Stephens agreed. Dr. Stephens's dissention was that additional validation studies involving prospective testing of chemicals in rabbits was not necessary.

Public Comment Session

1. Sadhana Dhruvakumar (PETA)

Ms. Dhruvakumar began by expressing her opinion that the discussion over the validation status of the methods (during the BCOP test method discussion) highlighted the apparent confusion among the Panel as to their ultimate charge in reviewing these methods. She believed that a statement that the validation status has been met should have been used with regard to the BCOP test method.

She then made reference to the reported data gaps for each of the test methods, and that these gaps may have been the result of inclusion that were too restrictive. She also pointed out that

other data sources had been mentioned during the course of the meeting, and therefore a sufficiently diligent search for the data had not been conducted.

Finally, Ms. Dhruvakumar commented on the evaluation of the underprediction rate of the Draize test conducted by Dr. Haseman. She stated that there was a lot of apparent bias in the study design. For example, assuming homogeneity within chemical classes was an assumption she stated would bias the test towards a more favorable outcome. She also stated that the evaluation only looked at intraexperimental variability. She closed by stating that due to the variability in the *in vivo* test, the calculated underprediction rate was a minimum at best, and PETA does not agree with the analysis.

2. Rodger Current (IIVS)

Dr. Curren opened by recognizing the amount of effort that was required of the Panel in reviewing the BRDs. He stated that he hoped that Panel's efforts could be used to yield the greatest potential value of outcome. He asked the practical question of whether it was the Panel's recommendation that these methods were to be used by industry prior to conducting an animal test. He declared if that indeed was the Panel's intention, then they in fact are indicating that the methods are "valid" for that purpose, and they should use this language in their recommendations. He indicated that the term "useful" was ambiguous and only caused confusion about whether the methods should actually be used or not.

Adjournment

Literature Citations

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.

Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 8:307-310.

CEC. 1991. Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test. Doc. XI/632/91/V/E/1/131/91 Part I and II.

CPSC. 1995. Test for eye irritants. 16CFR1500.42. Available: http://www.access.gpo.gov/nara/cfr/waisidx_00/16cfr1500_00.html [accessed 7 December 2005].

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 Pharm Exp Ther 82:377-390.

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

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

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..

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.

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, 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.

Guerriero FJ, Seaman CW, Olsen MJ, Guest R, Whittingham A. 2004. Retrospective assessment of the rabbit enucleated eye test (REET) as a screen to refine worker safety studies. [Abstract No. 1282]. Toxicol Sci (The Toxicologist Supplement). 78(1-S).

Hackett RB, McDonald TO. 1991. Eye irritation. In: Advances in Modern Toxicology: Dermatoxicology. 4th ed. (Marzulli F, Maibach H, eds.). Washington, DC: Hemisphere Publishing Corporation, 749-815.

INVITTOX. 1992. Protocol No. 47: HET-CAM Test. Available: http://ecvam-sis.jrc.it/invittox/published/indexed_47.html. [accessed 18 February 2004].

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

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

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

Swanson JE, 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 14:179-195.

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/officialtext.html. [accessed 26 August 2004].

Weil CS, Scala RA.1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicol Appl Pharmacol. 19:276-360.

January 11-12, 2005
Expert Panel Meeting to Assess the Current Validation Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability (BCOP), Hen's Egg Test – Chorioallantoic Membrane (HET-CAM), Isolated Chicken Eye (ICE) and Isolated Rabbit Eye (IRE)

"These Summary Minutes have been read and approved by the Chair of the Expert Panel Meeting on the Evaluation of the Validation Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants, as certified below."

_____ _____
Dr. Robert Scala Date
Panel Chair

_____ _____
Dr. William Stokes Date
Panel Executive Secretary