HTML

NIH Publication No: 06-4515

BACKGROUND REVIEW DOCUMENT

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

Hen's Egg Test - Chorioallantoic Membrane Test Method

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

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

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

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

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

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

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

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

The following Federal regulatory and research agencies are ICCVAM members:

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

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

Department of Transportation
Environmental Protection Agency

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

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

Hen's Egg Test - Chorioallantoic Membrane Test Method

Background Review Document

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

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

March 2006
NIH Publication No. 06-4515

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

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

Printed: 08/08/2006

TABLE OF CONTENTS

Page Number

LIST OF TABLES .... **xi**

LIST OF FIGURES .... **xv**

LIST OF ACRONYMS AND ABBREVIATIONS .... **xxii**

ACKNOWLEDGEMENTS .... **xxv**

PREFACE .... **xxiii**

EXECUTIVE SUMMARY .... **xxvii**

1.0 Introduction and Rationale for the Proposed Use of *In Vitro* Test Methods to Identify Ocular Corrosives and Severe Irritants **1-1**

1.1 Introduction 1-1

1.1.1 Historical Background of *In Vitro* Ocular Irritation/Corrosion Test Methods and Rationale for Their Development 1-1

1.1.2 Peer Reviews of the HET-CAM Test Method 1-6

1.2 Scientific Basis for the HET-CAM Test Method 1-6

1.2.1 Purpose and Mechanistic Basis of the HET-CAM Test Method 1-6

1.2.2 Similarities and Differences of Modes and Mechanisms of Action Between the HET-CAM Test Method and Human Ocular Irritancy 1-7

1.2.2.1 The Mammalian Eye: Common Anatomy of the Human and Rabbit Eye 1-7

1.2.2.2 Differences Between Human and Rabbit Eyes 1-9

1.2.2.3 The *In Vivo* Rabbit Eye Test Method 1-9

1.2.2.4 The Chorioallantoic Membrane (CAM) 1-13

1.2.2.5 Comparison of the HET-CAM Test Method with the *In Vivo* Rabbit Eye Test Method 1-13

1.2.3 Intended Range of Substances Amenable to the HET-CAM Test Method and/or Limits of the HET-CAM Test Method 1-14

1.3 Regulatory Rationale and Applicability 1-14

1.3.1 Current Regulatory Testing Requirements and ICCVAM Prioritization Criteria 1-14

1.3.2 Intended Uses of the Proposed HET-CAM Test Method 1-16

1.3.3 Similarities and Differences in the Endpoints Measured in the Proposed Test Method and the *In Vivo* Reference Test Method 1-16

1.3.4 Use of Proposed Test Method in Overall Strategy of Hazard for Safety Assessment 1-17

1.4 Validation of the *In Vitro* HET-CAM Test Method 1-17

1.5 Search Strategies and Selection of Citations for the HET-CAM BRD 1-19

2.0 HET-CAM Test Method Protocol Components 2-1
2.1 Overview of How the HET-CAM Test Method is Conducted 2-1
2.2 Description and Rationale for the Test Method Components 2-1
2.2.1 Materials, Equipment, and Supplies Needed 2-2
2.2.1.1 Hen Breed 2-2
2.2.1.2 Criteria for Egg Use 2-2
2.2.1.3 Egg Incubation Conditions (Temperature, Relative Humidity, Egg Rotation During Incubation, and Length of Incubation) 2-2
2.2.1.4 Rotary Saw 2-4
2.2.2 Dose-Selection Procedures, Including the Need for Any Dose Range-Finding Studies or Acute Toxicity Data Prior to Conducting a Study 2-4
2.2.3 Endpoints Measured 2-5
2.2.4 Exposure Parameters 2-5
2.2.4.1 Test Substance Exposure Amount or Volume 2-5
2.2.4.2 Concentration Tested 2-6
2.2.4.3 Application of Test Substance to CAM 2-6
2.2.4.4 Test Substance Exposure Duration 2-7
2.2.5 Known Limits of Use 2-7
2.2.6 Nature of the Response Assessed 2-7
2.2.7 Appropriate Controls and the Basis for their Selection 2-8
2.2.7.1 Negative Controls 2-8
2.2.7.2 Positive Controls 2-8
2.2.7.3 Solvent Controls 2-9
2.2.7.4 Benchmark Substances 2-9
2.2.8 Acceptable Range of Control Responses and the Basis for the Acceptable Ranges 2-10
2.2.8.1 Negative/Solvent Controls 2-10
2.2.8.2 Positive Controls 2-10
2.2.8.3 Benchmark Controls 2-12
2.2.9 Nature of the Data to be Collected and the Methods Used for Data Collection 2-12
2.2.9.1 Time to Development of Observed Endpoints After Exposure to the Test Substance 2-12
2.2.9.2 Lowest Test Substance Concentration Needed to Produce a Minimal Response on the CAM After Exposure to the Test Substance 2-13
2.2.9.3 Severity of Response Observed After Exposure to the Test Substance 2-13
2.2.10 Type of Media in Which Data Are Stored 2-13
2.2.11 Measures of Variability 2-14

2.2.12 Statistical or Non-Statistical Methods Used to Analyze the Resulting Data 2-14
2.2.12.1 Irritation Scores 2-14
2.2.12.2 Q-Score 2-15
2.2.12.3 Mean Detection Time 2-16
2.2.12.4 ITC 2-16
2.2.12.5 S-Score 2-16
2.2.12.6 SIS 2-16
2.2.13 Decision Criteria and the Basis for the Prediction Model Used to Classify a Test Chemical as a Severe Eye Irritant 2-16
2.2.13.1 Irritation Classification Based on IS 2-24
2.2.13.2 Irritation Classification Based on Q-Score 2-18
2.2.13.3 Irritation Classification Based on Mean Coagulation Detection Time 2-18
2.2.13.4 Irritation Classification Based on IS and ITC 2-18
2.2.13.5 Irritation Classification Based on S-Score 2-20
2.2.14 Information and Data that Will Be Included in the Study Report and Availability of Standard Forms for Data Collection and Submission 2-20
2.3 Basis for Selection of the HET-CAM Test Method System 2-22
2.4 Proprietary Components 2-22
2.5 Basis for Number of Replicate and Repeat Experiments 2-22
2.5.1 Sample Replicates 2-22
2.5.2 Experimental Replicates 2-22
2.6 Compliance with Good Laboratory Practices 2-23
2.7 Study Acceptance Criteria 2-23

3.0 Substances Used For Validation of the HET-CAM Test Method 3-1
3.1 Rationale for the Substances Selected for Use 3-1
3.1.1 CEC (1991) 3-1
3.1.2 Gettings et al. (1991, 1994, 1996) 3-1
3.1.3 Bagley et al. (1992) 3-1
3.1.4 Balls et al. (1995) 3-2
3.1.5 Vinardell and Macián (1994) 3-2
3.1.6 Kojima et al. (1995) 3-2
3.1.7 Gilleron et al. (1996, 1997) 3-2
3.1.8 Spielmann et al. (1996) 3-2
3.1.9 Hagino et al. (1999) 3-3
3.2 Rationale for the Number of Substances Tested 3-3
3.3 Chemicals or Products Evaluated 3-3
3.3.1 Substances Evaluated in Reviewed Studies 3-4
3.3.1.1 CEC (1991) 3-4
3.3.1.2 Gettings et al. (1991) 3-5
3.3.1.3 Gettings et al. (1994) 3-6
3.3.1.4 Gettings et al. (1996) 3-6
3.3.1.5 Bagley et al. (1992) 3-6

3.3.1.6 Vinardell and Macián (1994) 3-7
3.3.1.7 Balls et al. (1995).. 3-7
3.3.1.8 Kojima et al. (1995) 3-7
3.3.1.9 Gilleron et al. (1996)... 3-7
3.3.1.10 Spielmann et al. (1996) 3-8
3.3.1.11 Gilleron et al. (1997)... 3-8
3.3.1.12 Hagino et al. (1999) 3-8
3.4 Coding Procedures Used in the Validation Studies............................... 3-9
3.4.1 Gettings et al. (1991, 1994, 1996).. 3-9
3.4.2 Bagley et al. (1992) 3-9
3.4.3 Balls et al. (1995) 3-9
3.4.4 Spielmann et al. (1996).. 3-9
3.4.5 Hagino et al. (1999)... 3-9

4.0 *In Vivo* Reference Data Used for an Assessment of Test Method Accuracy 4-1
4.1 Description of Protocol Used to Generate *In Vivo* Data 4-1
4.1.1 Draize Rabbit Eye Test... 4-1
4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols......... 4-2
4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems 4-7
4.2 Detailed Reference Data Used to Assess *In Vitro* Test Method Accuracy... 4-8
4.3 *In Vivo* Classification Criteria Used for BRD Analysis........................ 4-10
4.3.1 GHS Classification Rules Used for BRD Analysis............... 4-11
4.3.2 EPA Classification Rules Used for BRD Analysis 4-11
4.3.3 EU Classification Rules Used for BRD Analysis 4-13
4.4 Availability of Original Records for the *In Vivo* Reference Data 4-14
4.5 *In Vivo* Data Quality 4-14
4.6 Availability and Use of Toxicity Information from the Species of Interest.. 4-16
4.7 Information About Accuracy and Reliability of the *In Vivo* Test Method 4-16
4.7.1 Information About the Accuracy of the *In Vivo* Test Method.. 4-16
4.7.2 Information About the Reliability of the *In Vivo* Test Method.. 4-16

5.0 HET-CAM Test Method Data And Results.. 5-1
5.1 Description of the HET-CAM Test Method Protocols Used to Generate Data 5-1
5.2 Availability of Copies of Original Data Used to Evaluate the Accuracy and Reliability... 5-1
5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data 5-2
5.3.1 IS 5-3
5.3.1.1 IS Analysis Method 5-3

5.3.1.2 IS Classification Scheme....5-4
5.3.2 Q-Score....5-4
5.3.2.1 Q-Score Analysis Method....5-4
5.3.2.2 Q-Score Classification Scheme....5-4
5.3.3 mtc10....5-5
5.3.3.1 mtc10 Analysis Method....5-5
5.3.3.2 mtc10 Classification Scheme....5-5
5.3.4 IS and ITC....5-5
5.3.4.1 IS and ITC Analysis Method....5-5
5.3.4.2 IS and ITC Classification Scheme....5-5
5.3.5 S-Score....5-6
5.3.5.1 S-Score Analysis Method....5-6
5.3.5.2 S-Score Classification Scheme....5-6
5.4 Summary of Results....5-6
5.4.1 CEC (1991)....5-7
5.4.2 Gettings et al. (1991)....5-7
5.4.3 Gettings et al. (1994)....5-7
5.4.4 Gettings et al. (1996)....5-7
5.4.5 Bagley et al. (1992)....5-8
5.4.6 Vinardell and Macián (1994)....5-8
5.4.7 Balls et al. (1995)....5-8
5.4.8 Kojima et al. (1995)....5-9
5.4.9 Gilleron et al. (1996)....5-9
5.4.10 Spielmann et al. (1996)....5-9
5.4.11 Gilleron et al. (1997)....5-10
5.4.12 Hagino et al. (1999)....5-10
5.5 Use of Coded Chemicals and Compliance with GLP Guidelines....5-10
5.6 Lot-to-lot Consistency of Test Substances....5-10
5.7 Availability of Data for External Audit....5-11

6.0 HET-CAM Test Method Accuracy....6-1
6.1 Accuracy of the HET-CAM Test Method....6-3
6.1.1 GHS Classification System: HET-CAM Test Method Accuracy....6-3
6.1.1.1 Overall Test Method Accuracy....6-4
6.1.1.2 Discordant Results According to the GHS Classification System....6-4
6.1.2 EPA Classification System: HET-CAM Test Method Accuracy....6-12
6.1.2.1 Overall Test Method Accuracy....6-12
6.1.2.2 Discordant Results According to the EPA Classification System....6-16
6.1.3 EU Classification System: HET-CAM Test Method Accuracy....6-20
6.1.3.1 Overall Test Method Accuracy....6-20

6.1.3.2 Discordant Results According to the EU Classification System 6-25
6.2 Accuracy of the HET-CAM Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Results 6-30
6.2.1 Discordance Among Chemical Classes 6-30
6.2.1.1 IS(A)-10 and IS(A)-100 Analysis Methods 6-30
6.2.1.2 IS(B)-10 and IS(B)-100 Analysis Methods 6-30
6.2.1.3 Q-Score Analysis Method 6-31
6.2.1.4 S-Score Analysis Method 6-31
6.2.2 Discordance Among Physical or Chemical Properties of Interest 6-31
6.2.2.1 IS(A)-10 and IS(A)-100 Analysis Methods 6-31
6.2.2.2 IS(B)-10 and IS(B)-100 Analysis Methods 6-32
6.2.2.3 Q-Score Analysis Method 6-32
6.2.2.4 S-Score Analysis Method 6-32

7.0 HET-CAM Test Method Reliability 7-1
7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the HET-CAM Test Method 7-1
7.2 Analyses of Repeatability and Reproducibility 7-2
7.2.1 Quantitative Assessment of Intralaboratory Repeatability 7-2
7.2.1.1 Gilleron et al. (1996) 7-2
7.2.1.2 Gilleron et al. (1997) 7-4
7.2.2 Quantitative Assessment of Intralaboratory Reproducibility 7-5
7.2.2.1 Gilleron et al. (1996) 7-5
7.2.2.2 Gilleron et al. (1997) 7-7
7.2.3 Quantitative and Qualitative Assessments of Assessment of Interlaboratory Reproducibility 7-10
7.2.3.1 Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System 7-11
7.2.3.2 Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System 7-15
7.2.3.3 Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System 7-20
7.2.3.4 Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results Based on Qualitative Analyses 7-24

7.2.3.5 Quantitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category 7-25
7.2.4 Additional Analyses of Interlaboratory Reproducibility 7-35
7.2.4.1 Balls et al. (1995) 7-35
7.2.4.2 Blein et al. (1991) 7-36
7.2.4.3 Doucet et al. (1999) 7-36
7.2.4.4 Spielmann et al. (1991) 7-36
7.2.4.5 Spielmann et al. (1993) 7-36
7.3 Historical Positive and Negative Control Data 7-38
7.3.1 Historical Positive Control Data 7-38
7.3.1.1 Positive Control Studies Using DMF and Imidazole 7-38
7.3.1.2 Positive Control Studies Using SDS and NaOH 7-39
7.3.2 Historical Negative Control Data 7-49
7.4 Summary 7-40

8.0 Test Method Data Quality 8-1
8.1 Adherence to National and International GLP Guidelines 8-1
8.2 Data Quality Audits 8-1
8.3 Impact of Deviations from GLP Guidelines 8-2
8.4 Availability of Laboratory Notebooks or Other Records 8-2
8.5 Need for Data Quality 8-2

9.0 Other Scientific Reports And Reviews 9-1
9.1 Reports in the Peer Reviewed Literature 9-1
9.1.1 Bagley et al. (1992) 9-1
9.1.2 Balls et al. (1995) 9-1
9.1.3 Blein et al. (1991) 9-2
9.1.4 Brantner et al. (2002) 9-3
9.1.5 Brantom et al. (1997) and Steiling et al. (1999) 9-4
9.1.6 Budai et al. (1997) and Budai and Várnagy (2000) 9-5
9.1.7 CEC (1991) 9-5
9.1.8 Dannhardt et al. (1996) 9-5
9.1.9 Demirci et al. (2003) 9-6
9.1.10 Demirci et al. (2004) 9-6
9.1.11 de Silva et al. (1992) 9-6
9.1.12 Djabari et al. (2002) 9-7
9.1.13 Doucet et al. (1999) 9-8
9.1.14 Gettings et al. (1991, 1994, 1996) and Lordo et al. (1999) 9-8
9.1.15 Gilleron et al. (1996) 9-10
9.1.16 Gilleron et al. (1997) 9-11
9.1.17 Hagino et al. (1991) 9-11
9.1.18 Hagino et al. (1993) 9-12
9.1.19 Hagino et al. (1999) and Ohno et al. (1999) 9-12

9.1.20 Kalweit et al. (1987) ... 9-13
9.1.21 Kalweit et al. (1990) ... 9-13
9.1.22 Kojima et al. (1995) ... 9-14
9.1.23 Lawrence et al. (1990) ... 9-14
9.1.24 Lönnroth et al. (1999) ... 9-14
9.1.25 Luepke (1985) ... 9-14
9.1.26 Luepke and Kemper (1986) ... 9-15
9.1.27 Macián et al. (1996) ... 9-15
9.1.28 Reinhardt et al. (1987) ... 9-15
9.1.29 Rougier et al. (1992) ... 9-16
9.1.30 Schlage et al. (1999) ... 9-16
9.1.31 Spielmann et al. (1991, 1993, 1996) ... 9-16
9.1.32 Spielmann et al. (1997) ... 9-18
9.1.33 Sterzel et al. (1990) ... 9-20
9.1.34 van Erp et al. (1990) ... 9-20
9.1.35 Vinardell and García (2000) ... 9-20
9.1.36 Vinardell and Macián (1994) ... 9-20
9.1.37 Vives et al. (1997) ... 9-21
9.1.38 Wilson and Steck (2000) ... 9-21
9.1.39 Worth and Cronin (2001) ... 9-21
9.2 Data Received in Response to the ICCVAM *Federal Register* Notice or from Study Authors ... 9-22

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement) ... 10-1
10.1 How the HET-CAM Test Method Will Refine, Reduce, or Replace Animal Use ... 10-1
10.2 Requirement for the Use of Animals ... 10-1

11.0 Practical Considerations ... 11-1
11.1 Transferability of the HET-CAM Test Method ... 11-1
11.1.1 Facilities and Major Fixed Equipment Required to Conduct a Study ... 11-1
11.1.2 General Availability of Other Necessary Equipment and Supplies ... 11-1
11.2 Training Considerations ... 11-1
11.2.1 Required Level of Training and Expertise Needed to Conduct the HET-CAM Test Method ... 11-2
11.2.2 Training Requirements Needed to Demonstrate Proficiency ... 11-2
11.3 Cost Considerations ... 11-2
11.4 Time Considerations ... 11-2

12.0 References ... 12-1

13.0 Glossary ... 13-1

Appendix A Protocols for the HET-CAM Test Method....................................... A-1
A1 INVITTOX Protocol 47. The HET-CAM Test – Method of Spielmann and Liebsch A-3
A2 Table of HET-CAM Protocols from the Reviewed Literature Used to Assess Ocular Irritancy A-13

Appendix B Chemical and Product Class Information for Substances Tested in the HET-CAM Test Method B-1
B1 Table of Chemical and Product Class Information................................ B-3
B2 Table of Formulation Components in Gettings et al. (1991)................ B-19
B3 Table of Formulation Components in Gettings et al. (1994)................ B-25
B4 Table of Formulation Components in Gettings et al. (1996)................ B-35

Appendix C *In Vitro* Data for Substances Tested in the HET-CAM Test Method.. C-1
C1 Test Method Protocols used to Generate *In Vitro* Data C-3
C2 Data Sorted by Reference.. C-15
C3 Data Sorted by Substance for IS(A) Analysis Method C-53
C4 Data Sorted by Substance for IS(A)-10 Analysis Method C-63
C5 Data Sorted by Substance for IS(A)-100 Analysis Method C-69
C6 Data Sorted by Substance for IS(B) Analysis Method......................... C-75
C7 Data Sorted by Substance for IS(A)-10 Analysis Method C-89
C8 Data Sorted by Substance for IS(A)-100 Analysis Method C-111
C9 Data Sorted by Substance for Q-Score Analysis Method C-139
C10 Data Sorted by Substance for S-Score Analysis Method... ....C-149

Appendix D *In Vivo* and *In Vitro* Comparison of Ocular Irritancy Classification D-1
D1 Data Sorted by Reference.. D-3
D2 Data Sorted by Substance for IS(A) Analysis Method D-17
D3 Data Sorted by Substance for IS(A)-10 Analysis Method D-23
D4 Data Sorted by Substance for IS(A)-100 Analysis Method D-27
D5 Data Sorted by Substance for IS(B) Analysis Method......................... D-31
D6 Data Sorted by Substance for IS(B)-10 Analysis Method D-45
D7 Data Sorted by Substance for IS(B)-100 Analysis Method D-57
D8 Data Sorted by Substance for Q-Score Analysis Method D-75
D9 Data Sorted by Substance for S-Score Analysis Method..................... D-81

Appendix E Intralaboratory Repeatability and Intralaboratory Reproducibility Data and Results for Gilleron et al. (1996, 1997) E-1
E1 Individual Egg HET-CAM Scores (Endpoints Evaluated and Overall Irritancy) for Data of Gilleron et al. (1996) E-3
E2 Individual Egg HET-CAM Scores (Endpoints Evaluated and Overall Irritancy) for Data of Gilleron et al. (1997) E-11

E3 Intralaboratory Repeatability Results (Expressed as %CV) for Data of Gilleron et al. (1996)....E-21
E4 Intralaboratory Repeatability Results (Expressed as %CV) for Data of Gilleron et al. (1997)....E-29
E5 Intralaboratory Reproducibility Results (Expressed as %CV) for Data of Gilleron et al. (1996) E-37
E6 Intralaboratory Reproducibility Results (Expressed as %CV) for Data of Gilleron et al. (1997) E-41

Appendix F Interlaboratory Reproducibility Data....F-1
F1 Individual HET-CAM Test Laboratory Data for CEC (1991)....F-3
F2 Average HET-CAM Scores and Coefficient of Variations for Substances Evaluated in Hagino et al. (1999) and Ohno et al. (1999)....F-13

Appendix G Interlaboratory Correlation Coefficients from the EC/HO Validation Study (Balls et al. 1995) G-1

Appendix H Positive Control Data....H-1
H1 Individual HET-CAM Egg Data for Dimethylformamide (provided by Drs. Vanparys and van Goethem)....H-3
H2 Individual HET-CAM Egg Data for Imidazole (provided by Drs. Vanparys and van Goethem) H-9
H3 Individual HET-CAM Egg Data for Sodium Dodecyl Sulfate and Sodium Hydroxide (provided by Drs. Spielmann and Liebsch) H-11

Appendix I Individual HET-CAM Scores for Eggs Treated with 0.9% Sodium Chloride I-1

LIST OF TABLES

		Page Number
Table ES-1	Ranges of Performance Statistics for Evaluated Analysis Methods for GHS, EPA, and EU Classification System	xxxvii
Table 1-1	Summary of Current U.S. Legislation Related to Ocular Health	1-1
Table 1-2	*In Vivo* Ocular Irritancy Classification Systems	1-4
Table 1-3	Terms and Phrases Used for Online Literature Searches	1-19
Table 2-1	Means and Standard Deviations of Negative Control Test Substances Evaluated With and Without Use of the Test Substance Applicator	2-10
Table 2-2	Comparison of Means and Standard Deviations of Positive Control Test Substances Evaluated With and Without Use of a Test Substance Applicator	2-11
Table 2-3	Means and Standard Deviations of Positive Control Test Substances	2-11
Table 2-4	HET-CAM Scoring System	2-13
Table 2-5	Summary of HET-CAM Score Ranges Used in Irritancy Classification	2-17
Table 2-6	Decision Criteria Used by Gettings and Colleagues to Define Test Substances as Irritants According to the FHSA Classification System	2-18
Table 2-7	Summary of Q-Scores Used in Irritancy Classification	2-18
Table 2-8	Summary of mtc10 Values Used in Irritancy Classification	2-19
Table 2-9	Summary of ITC and IS Combinations Used in Irritancy Classification	2-19
Table 2-10	Summary of S-Scores Used in Irritancy Classification	2-20
Table 3-1	Chemical Classes Tested in the HET-CAM Test Method	3-4
Table 3-2	Product Classes Tested in the HET-CAM Test Method	3-5
Table 4-1	Scale of Weighted Scores for Grading the Severity of Ocular Lesions	4-2
Table 4-2	Test Guidelines for *In Vivo* Ocular Irritation Test Methods	4-3
Table 4-3	Criteria for Classification of Animals According to the GHS Classification System (Modified from UN 2003)	4-12
Table 4-4	Criteria for Classification of Substances According to the GHS Classification System (Modified from UN 2003)	4-12
Table 4-5	Criteria for Classification of Animals According to the EPA Classification System	4-13
Table 4-6	Criteria for Classification of Substances According to the EU Classification System	4-13
Table 5-1	IS Classification Scheme Used to Classify Substances for Accuracy Analysis	5-4
Table 5-2	Q-Score Classification Scheme Used to Classify Substances for Accuracy Analysis	5-5
Table 5-3	mtc10 Classification Scheme Used to Classify Substances for Accuracy Analysis	5-5

Table 5-4 IS and ITC Classification Scheme Used to Classify Substances for Accuracy Analysis 5-6
Table 5-5 S-Score Classification Scheme Used to Classify Substances for Accuracy Analysis 5-6
Table 6-1 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by Study 6-5
Table 6-2 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by HET-CAM Analysis Method 6-6
Table 6-3 False Positive and False Negative Rates for the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the GHS Classification System 6-7
Table 6-4 False Positive and False Negative Rates for the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the GHS Classification System........ 6-9
Table 6-5 False Positive and False Negative Rates for the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the GHS Classification System 6-11
Table 6-6 False Positive and False Negative Rates for the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the GHS Classification System 6-12
Table 6-7 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by Study 6-14
Table 6-8 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by HET-CAM Analysis Method 6-15
Table 6-9 False Positive and False Negative Rates for the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EPA Classification System 6-17
Table 6-10 False Positive and False Negative Rates for the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EPA Classification System 6-18
Table 6-11 False Positive and False Negative Rates for the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the EPA Classification System 6-20
Table 6-12 False Positive and False Negative Rates for the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the EPA Classification System 6-20
Table 6-13 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the

In Vivo Rabbit Eye Test Method, as Defined by the EU Classification System, by Study 6-22
Table 6-14 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System, by HET-CAM Analysis Method 6-24
Table 6-15 False Positive and False Negative Rates for the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EU Classification System 6-26
Table 6-16 False Positive and False Negative Rates for the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EU Classification System 6-27
Table 6-17 False Positive and False Negative Rates for the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the EU Classification System 6-28
Table 6-18 False Positive and False Negative Rates for the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the EU Classification System 6-29
Table 6-19 Ranges of Performance Statistics for Evaluated Analysis Methods for GHS, EPA, and EU Classification Systems.................................. 6-30
Table 7-1 Intralaboratory Repeatability Results for HET-CAM Studies of Gilleron et al. (1996)... 7-4
Table 7-2 Intralaboratory Repeatability Results for HET-CAM Studies of Gilleron et al. (1997)... 7-6
Table 7-3 Intralaboratory Reproducibility Results for HET-CAM Studies of Gilleron et al. (1996)... 7-8
Table 7-4 Intralaboratory Reproducibility Results for HET-CAM Studies of Gilleron et al. (1997)... 7-9

Table 7-5 Evaluation of the Reliability of the HET-CAM Test Method in Predicting Ocular Corrosives and Severe Irritants as Defined by the GHS Classification System, by Study... 7-12
Table 7-6 Evaluation of the Reliability of the HET-CAM Test Method in Predicting Ocular Corrosives and Severe Irritants as Defined by the EPA Classification System, by Study 7-16
Table 7-7 Evaluation of the Reliability of the HET-CAM Test Method in Predicting Ocular Corrosives and Severe Irritants as Defined by the EU Classification System, by Study... 7-21
Table 7-8 %CV Values for Substances Evaluated at 100% Concentration *In Vitro* Using the IS(B) Analysis Method (from CEC 1991).................. 7-26
Table 7-9 %CV Values for Substances Evaluated at 10% Concentration *In Vitro* Using the IS(B) Analysis Method (from CEC 1991).................. 7-27
Table 7-10 %CV Values for Substances Evaluated Using the Q-Score Analysis Method (from Balls et al. 1995) 7-28
Table 7-11 %CV Values for Substances Evaluated Using the S-Score Analysis Method (from Balls et al. 1995)... 7-29

Table 7-12 %CV Values for Substances Evaluated Using IS(B) Analysis Method (from Spielmann et al. 1996) 7-30
Table 7-13 %CV Values for Substances Evaluated Using the IS(A) Analysis Method (from Hagino et al. 1999) 7-34
Table 7-14 Interlaboratory Correlation Coefficients in Balls et al. (1995) 7-35
Table 7-15 Comparison of Means and Standard Deviations for Positive Controls Tested With and Without Test Substance Applicator 7-38
Table 7-16 Intralaboratory Reproducibility Results for Evaluation of Imidazole as a Positive Control 7-39
Table 7-17 Intralaboratory Reproducibility Results for Evaluation of DMF as a Positive Control 7-39
Table 7-18 Means and Standard Deviations of Positive Control Test Substances 7-40
Table 7-19 Comparison of Means and Standard Deviations of 0.9% NaCl With and Without Use of the Test Substance Applicator 7-40
Table 9-1 *In Vitro/In Vivo* Range of Correlations Reported in Balls et al. (1995) 9-3
Table 9-2 Accuracy Statistics for Test Substances Evaluated in Gilleron et al. (1996) 9-10
Table 9-3 Accuracy Statistics for Test Substances Evaluated in Gilleron et al. (1997) 9-11

LIST OF FIGURES

Page Number

Figure 1-1 Anatomy of the Human Eye 1-8
Figure 1-2 GHS Testing Strategy for Serious Eye Damage and Eye Irritation....... 1-18

LIST OF ACRONYMS AND ABBREVIATIONS

^{o}C	Degrees Centigrade
ASTM	American Society for Testing and Materials
BCOP	Bovine Corneal Opacity and Permeability
BLS	U.S. Bureau of Labor Statistics
BRD	Background Review Document
CAM	Chorioallantoic Membrane
CAS	Chemical Abstract Service
CASRN	Chemical Abstracts Service Registry Number
COLIPA	European Cosmetic, Toiletry, and Perfumery Association
CPSC	U.S. Consumer Product Safety Commission
CTFA	Cosmetic, Toiletries and Fragrance Association
CV	Coefficient of Variation
DOD	U.S. Department of Defense
DOT	U.S. Department of Transportation
EC	European Commission
EC/HO	European Commission/British Home Office
ECETOC	European Center for Ecotoxicology and Toxicology Of Chemicals
ECVAM	European Center for the Validation of Alternative Methods
EEC	European Economic Council
EPA	U.S. Environmental Protection Agency
EU	European Union
FDA	U.S. Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FHSA	Federal Hazardous Substance Act
FR	*Federal Register*
g	Gram
GHS	Globally Harmonized System
GLP	Good Laboratory Practice
HET-CAM	Hen's Egg Test-Chorioallantoic Membrane

ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye
INVITOXX	*In Vitro* Techniques in Toxicology (ERGATT FRAME ECVAM Data bank)
IRAG	Interagency Regulatory Alternatives Group
IRE	Isolated Rabbit Eye
IS	Irritation Score
ITC	Irritation Threshold Concentration
IU	International Units
mg	Milligram
min	Minutes
mL	Milliliter
MAS	Maximum Average Score
MDES	Maximal Draize Eye Score
mM	Millimolar
MMAS	Modified Maximum Average Score
MMMIS	Mean (nonweighted) of Modified Maximum Individual Score
Mtc	Mean time of coagulation
N	Normality
N/A	Not Available or Not Applicable
NS	Not Specified
NaOH	Sodium Hydroxide
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIOSH	National Institute of Occupational Safety and Health
NTP	U.S. National Toxicology Program
OECD	Organization for Economic Cooperation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances

OPAL	Oeuvre Pour l'Assistance aux Animaux de Laboratories
OSHA	U.S. Occupational Safety & Hazards Administration
OTWG	Ocular Toxicity Working Group
P.L.	Public Law
QA	Quality Assurance
RN	Registry Number
SCNM	Study Criteria Not Met
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SIS	Severity Irritation Score
SLS	Sodium Lauryl Sulfate
TG	Test Guideline
TSA	Test Substance Applicator
TSCA	Toxic Substances Control Act
UN	United Nations
v/v	Volume to Volume Ratio
w/v	Weight to Volume Ratio
w/w	Weight to Weight Ratio
WHO	World Health Organization
ZEBET	German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments

ACKNOWLEDGEMENTS

The following individuals are acknowledged for their contributions to the In Vitro Ocular Test Method review process:

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

Consumer Product Safety Commission (CPSC)

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

Department of Defense (DOD)

Harry Salem, Ph.D.

Department of Transportation (DOT)

Steve Hwang, Ph.D.

Environmental Protection Agency (EPA)

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

Food and Drug Administration (FDA)

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

National Institute for Occupational Safety & Health (NIOSH)

Surender Ahir, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)

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

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

David Allen, Ph.D.
ILS, Inc.

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

Sue Brenzel
ILS, Inc.

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

Patricia Ceger, M.S.
ILS, Inc.

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

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

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

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

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

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

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

James Truax, M.A.
ILS, Inc.

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

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

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

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

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

Penny Jones
Unilever Research
Sharnbrook, United Kingdom

Menk Prinsen
TNO Nutrition & Food Research Institute
The Netherlands

Horst Spielmann, Dr. med.
ZEBET
Berlin, Germany

Companies and Individuals that Provided *In Vitro* and/or *In Vivo* Data for the HET-CAM Background Review Document

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

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

ECVAM
Chantra Eskes, Eng, Ph.D.

Johnson & Johnson Pharmaceutical, R&D
Philippe Vanparys, Ph.D.
Freddy Van Goethem, Ph.D.

National Institute of Health Sciences
Yasuo Ohno, Ph.D.

U.S. Food and Drug Administration
Donnie Lowther

ZEBET
Horst Spielmann, Dr. med.
Manfred Liebsch, Ph.D

PREFACE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

EXECUTIVE SUMMARY

This Background Review Document (BRD) reviews available data and information regarding the validation status of the Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) test method for identifying ocular corrosives and severe irritants. The test method was reviewed for its ability to predict ocular corrosives and severe/irreversible effects as defined by the U.S. Environmental Protection Agency (EPA) (EPA 1996), the European Union (EU) (EU 2001), and the United Nations (UN) Globally Harmonized System (GHS) of Classification and Labeling of Chemicals (UN 2003). The objectives of this BRD is to describe the current validation status of the HET-CAM test method, including what is known about its accuracy and reliability, the scope of the substances tested, and the availability of a standardized test method protocol.

The information summarized in this BRD is based on publications obtained from the peer-reviewed literature, as well as unpublished information submitted to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in response to two *Federal Register* (*FR*) Notices requesting high quality *in vivo* rabbit eye test and *in vitro* ocular irritation data for HET-CAM, the Isolated Chicken Eye (ICE), the Isolated Rabbit Eye (IRE), and the Bovine Corneal Opacity and Permeability (BCOP) test methods. An online literature search identified 214 publications that contained HET-CAM test method results and protocol information; of these publications, detailed *in vivo* and *in vitro* data were available for 12 studies[1] that allowed for an evaluation of test method accuracy[2] and reliability[3].

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

The HET-CAM test method uses the chorioallantoic membrane (CAM), which is a vascular fetal membrane composed of the fused chorion and allantois. The method is proposed to provide information on the effects that may occur in the conjunctiva following exposure to a test substance. Published reviews note that chicken-embryo models have long been used as models by embryotoxicologists and virologists. (Parish 1985; Luepke and Kemper 1986). Extending the use of chicken embryos, the HET-CAM test method was proposed by Luepke (1985) and Luepke and Kemper (1986). It was assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane are similar to

---

[1] Sufficient information was available for 10 of these publications to assess test method accuracy when compared to the GHS (UN 2003), EPA (1996), and EU (2001) classification systems. For two publications, sufficient information was only available to assess test method accuracy when compared to the EU (2001) classification system.

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

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

effects induced by the same test substance in the eye of a treated rabbit. The CAM has been proposed as a model for a living membrane (such as the conjunctiva) since it comprises a functional vasculature. Additionally, evaluation of coagulation (i.e., protein denaturation) may reflect corneal damage that may be produced by the test substance. The CAM is evaluated for the development of irritant endpoints (hyperemia, hemorrhage, and coagulation). Depending on the method used to collect data on the endpoints (time to development, severity of observed effect) qualitative assessments of the irritation potential of test substances are made.

U.S. Federal regulatory agencies were surveyed to determine whether HET-CAM test method data have been considered for regulatory use where submission of testing data is required. Responses indicated that such data have not been provided to surveyed regulatory agencies.

The HET-CAM test method is currently used by some companies for the identification of ocular corrosives and severe irritants in a tiered testing strategy on a case-by-case basis. In this strategy, positive *in vitro* test results are considered in a weight-of-evidence decision as to whether to classify the substance as an ocular corrosive or severe irritant. Negative results and suspected false positive *in vitro* results proceed to standard *in vivo* testing or to validated *in vitro* test methods that are capable of detecting false negative corrosives and severe irritants.

The HET-CAM test method protocols used in the various studies considered in this BRD are similar, but not identical. Examples of some of the test method components that differed among the HET-CAM protocols used to generate data include:

- relative humidity during egg incubation ranged from 52.5 to 62.5%,
- volume or quantity of the test substance applied to the CAM (when reported) was either 0.1 or 0.3 mL for liquids and 0.3 g for solids,
- number of replicate eggs per test substance ranged from 3 to 6, and
- some studies included concurrent positive control substances, while others did not.

In addition to the various test method protocol permutations in the published literature, there were several HET-CAM analysis methods utilized to assess acute eye irritation. The analysis methods that are described in the literature include: Irritation Score (defined as IS(A) and IS(B)), Q-Score, S-Score, mtc value, and the IS and ITC method. All of these analysis methods are reviewed and evaluated in the BRD. Furthermore, the data available allowed for additional assessments based on the concentration tested *in vivo* and *in vitro*.

A total of 260 substances and formulations were evaluated in the studies. A variety of chemical and product classes have been tested in the HET-CAM assay. The chemical classes with the greatest number of substances tested are alcohols, carboxylic acids, and organic salts. For some of the test substances that were identified as formulations, components of the formulation and the relative concentrations of the components were available. The most common product classes tested are solvent, shampoo, surfactants, and cosmetics.

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

None of the studies provided original test result data. However, summary *in vitro* data was available for all of the test substances evaluated such that they could be assigned *in vitro* irritancy classifications for comparison to the available *in vivo* reference data.

The accuracy evaluation of the HET-CAM test method was limited to the substances evaluated in 10 to 12 *in vitro-in vivo* comparative studies. The ability of the HET-CAM test method to correctly identify ocular corrosives and severe irritants, as defined by the EPA (1996), the EU (2001), and the GHS (UN 2003) was evaluated using two approaches. In the first approach, the accuracy of HET-CAM was assessed separately for each *in vitro-in vivo* comparative study. In the second approach, the accuracy of HET-CAM was assessed after pooling data across *in vitro-in vivo* comparative studies that used the same method of data collection and analysis. While there were some differences in results among the three hazard classification systems evaluated (i.e., EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that HET-CAM test method performance was comparable among the three hazard classification systems (see **Table ES-1**).

Table ES-1 Ranges of Performance Statistics for Evaluated Analysis Methods for GHS, EPA, and EU Classification Systems

Analysis Methods	Accuracy	Sensitivity	Specificity	False Positive Rates	False Negative Rates
IS(A)-10	48-50%	24-25%	100%	0%	75-76%
IS(A)-100	85%	100%	83%	17%	0%
IS(B)-10	65-68%	68-70%	64-67%	33-36%	30-32%
IS(B)-100	51-57%	87-93%	40-47%	52-59%	6-13%
Q-Score	61-64%	100%	43-46%	54-57%	0%
S-Score	44-50%	36-44%	60-67%	33-40%	56-64%

Abbreviations: EPA = U.S. Environmental Protection Agency, EU = European Union, GHS = Globally Harmonized System.
A single value indicates the same percentage results for all three hazard classification systems.

Most of the substances evaluated by the IS(A)-10 and IS(A)-100 analysis methods were formulations. For the IS(A)-10 analysis method, which evaluated mostly surfactant-based formulations, the false negative rates ranged from 75% to 76%, while the false positive rate

was 0% for all classification systems. Comparatively, the IS(A)-100 analysis method, which evaluated primarily oil-water formulations, had a higher false positive rate than false negative rate.

With regard to physical form of the substances tested by these analysis methods, a majority was tested as liquids/solutions *in vitro* and *in vivo*. Therefore, the false negative and false positive rates for these analysis methods were consistent or the same as to the overall false positive and false negative rates. No solids were evaluated using the IS(A)-10 analysis method, while the false negative and false positive rates were 0% for the IS(A)-100 analysis method. For the GHS classification scheme, the evaluation indicated that substances were more likely to be underpredicted if (a) the *in vivo* lesion was based on persistence of effect and (b) if the *in vitro* test concentration was 100%.

The chemical class of substances that was consistently overpredicted according to the GHS classification system (i.e., were false positives) by the IS(B)-10 and IS(B)-100 analysis methods is alcohols (89% to 90% for the IS(B)-10 analysis method and 79% to 88% for the IS(B)-100 analysis method). Additional chemical classes that were overpredicted by both analysis methods were ethers, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% for IS(B)-10 and 23% to 26% for IS(B)-100). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% and 37% to 38% (7/18), respectively for liquids and 58% to 65% and 0% to 13% for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 61% to 65% and 0%, respectively for liquids and 48% to 67% and 8% to 24% for solids. The physical form of many of the tested substances was unknown based on the available information.

Information regarding the pH of test substances was available for a subset of the substances tested (29 to 35 substances). Overall, substances were observed to have a higher false positive rate when (a) tested at a 100% concentration (IS(B)-100) and (b) had a pH greater than 7.0. For the GHS classification scheme, the evaluation indicated that substances were more likely to be underpredicted if (a) the *in vivo* lesion was based on persistence of effect and, (b) if the *in vitro* test concentration was 10%.

The accuracy analysis indicated that alcohols and esters are often overpredicted (43 to 50% and 43% false positive rate, depending on the classification system used) in the Q-score analysis method. The numbers of substances among the remaining chemical classes were too few to resolve any definitive trends in overprediction by the Q-Score analysis method. The false negative rate for all chemical classes with a sufficient number of substances ($n \geq 5$) was 0%.

With regard to physical form of the substances overpredicted by the Q-Score analysis method, 14 to 17 were liquids and none were solids. The ranges of false positive and false negative rates for liquids were 56% to 61% and 0%, respectively. The false positive and false negative rates for solids were 0% for both parameters. There was insufficient information for the other evaluated categories (e.g., surfactant-based formulations) to conduct an analysis.

Due to the limited database for the S-Score analysis method, a chemical class evaluation could only be conducted for carboxylic acids/carboxylic acid salts for the GHS classification system. For this chemical class and classification system, the false negative rate was 75% (3/4) and the false positive rate was 0% (0/1).

With regard to physical form of the substances overpredicted by the S-Score analysis method, 14 to 16 were solids. There were no liquids evaluated with analysis method. The false negative rates for solids ranged from 56%-64% (5/9 to 7/11) and the false positive rates ranged from 33% to 40% (2/6 to 2/5). There was insufficient information for the other evaluated categories (e.g., surfactant-based formulations) to conduct an analysis.

The analysis of intralaboratory repeatability was evaluated using data from two different publications (Gilleron et al. 1996, 1997) for the IS(B) analysis method. In both studies, the hemorrhage endpoint had a high %CV value (104 to 117). Additionally, the %CV values for the coagulation endpoint were the lowest of the three endpoints evaluated in the HET-CAM test method. However, the actual values were quite disparate between the two studies (e.g., Gilleron et al. 1996 coagulation %CV = 95.69; Gilleron et al. 1997 coagulation %CV = 41.78). The difference in the numbers may be due to several factors including test substances evaluated and differences in the test method protocols used between the two studies. The calculated variability for the endpoints and the overall test method may be exaggerated because of the relatively small values that are obtained from each of the endpoints (5 for hemorrhage, 7 for lysis, and 9 for coagulation). Similar results were obtained from the analysis of intralaboratory reproducibility. The overall irritation score was generally reproducible (%CV values of 53 and 17.5 for the two studies evaluated).

A qualitative assessment of the data provided for multiple laboratories in three to four studies indicates the extent of interlaboratory reproducibility. Given the relatively homogeneous performance of the HET-CAM test method among the three classification systems, the discussions for the individual studies and analysis methods encompasses all three hazard classification systems, unless otherwise indicated. The two to four participating laboratories that used the Q-Score analysis method were in 100% agreement in regard to the ocular irritancy classification for 21 (45%) of the 47 substances analyzed. Comparatively, participating laboratories were in 100% agreement for 12 to 13 (66% to 68%) of the 18 to 19 substances analyzed using the S-Score analysis method, depending on the classification system used. For the IS(B)-10 analysis method, the participating laboratories were in 100% agreement for 84 to 85 (79% to 81%) of 104 to 107 substances evaluated. For the IS(B)-100 analysis method, the participating laboratories in Spielmann et al. (1996) were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated. There was 100% agreement in regard to the GHS ocular irritancy classification for 11 (64% to 69%) of the 16

to 17 substances evaluated in five laboratories using the IS(A) analysis method in Hagino et al. (1999).

The overall reliability statistics, arranged by HET-CAM data analysis method, for the IS(B), IS(B)-10, S-Score and Q-Score are identical to what was discussed previously. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For both of these analysis methods, the addition of the results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Hagino et al. (199) and Spielmann et al, (1996).

A quantitative evaluation of interlaboratory reproducibility was conducted for four studies (CEC 1991; Balls et al. 1995; Spielmann et al. 1996; Hagino et al. 1999) by performing a %CV analysis of *in vitro* scores obtained for substances tested in multiple laboratories. For CEC (1991), two different evaluations were conducted based on the concentration tested *in vitro*. For 14 substances evaluated at 100% concentration, the mean and median %CV values were 31.86 and 33.04, respectively. For 12 substances evaluated at 10% concentration, the mean and median %CV values were 34.6 and 33.1, respectively. For the Balls et al. (1995) study, the average and median %CV values for substances evaluated with the Q-Score were 49.83 and 42.50, respectively. The average and median %CV values for the substances evaluated with the S-Score were 84.42 and 71.90, respectively. For the substances evaluated in Spielmann et al. (1996), the average and median %CV values for substances tested at 10% concentration were 60.17 and 42.65, respectively. For substances tested at 100% concentration in Spielmann et al. (1996), the average and median %CV values were lower: 35.21 and 26.22, respectively. When substances that were tested in three different testing laboratories were removed from the assessment, little change was seen in the mean and median %CV values for both concentrations tested. For Hagino et al. (1999), the average and median %CV for substances classified as GHS Category 1 (UN 2003) were 24.4 and 27.0, respectively. The average and median %CV for substances classified as EPA Category I (EPA [1996]) were 23.86 and 26.0, respectively.

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

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

1.1 Introduction

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

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

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

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

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

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

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

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

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

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

persistence as Category 2A ("irritating to eyes") (reverses within 21 days) and Category 2B ("mildly irritating to eyes") (reverses within 7 days). The GHS (UN 2003) categories are based on severity of the lesions and/or the duration of persistence. The GHS, the US, and the EU *in vivo* ocular irritancy classification systems are described in greater detail in **Section 4.1.3**.

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

This BRD evaluates the ability of the HET-CAM test method to identify severe ocular irritants and corrosives. The HET-CAM test method was developed by Luepke (1985) and Luepke and Kemper (1986). The chorioallantoic membrane (CAM) is a vascularized respiratory membrane that surrounds the embryonic bird within an egg. The test method is based on the observation that the CAM of an embryonated hen's egg is similar to the vascularized mucosal tissues of the eye. The test method developers assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane would be similar to effects induced by the same test substance in the eye of a treated rabbit. Thus, it was proposed that adverse effects on the CAM induced by a test substance would correlate to irritation and/or corrosion in human eyes.

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

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

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

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

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

The HET-CAM test method is currently used in some U.S. and European companies (e.g., pharmaceutical, cosmetic, and personal care product companies) as an in-house screen to assess the ocular irritation potential of a wide range of substances or products. Substances are tested either individually, as mixtures, or in product formulations. The test method is used in the following ways: (1) for classification of industrial chemicals as severe eye irritants for labeling purposes, and (2) for safety assessment of raw materials, new ingredients, and formulations (Spielmann H, personal communication).
Although the HET-CAM test method is not yet validated, the EU national regulatory authorities accept positive outcomes from this test method for classifying and labeling severe eye irritants (R41). Where a negative result is obtained, an *in vivo* test is subsequently required, as the HET-CAM test method has not been shown to adequately discriminate between eye irritants and nonirritants (Liebsch and Spielmann 2002; European Communities 2004).

1.1.2 Peer Reviews of the HET-CAM Test Method

Studies have been conducted in recent years to assess the validity of the HET-CAM test method as a complete replacement for the *in vivo* ocular irritation and corrosion test method (e.g., Balls et al. 1995). Additionally, Spielmann et al. (1996) assessed the ability of the HET-CAM test method to identify severe ocular irritants as classified by the EU classification system (EU 1992). Previous validation efforts for the HET-CAM test method may have failed because: 1) they attempted to support the utility of an *in vitro* alternative as a full replacement for the *in vivo* rabbit test, rather than as a component in a tiered testing strategy; and/or 2) data generated with the *in vitro* test method(s) have typically been compared to *in vivo* Maximum Average Scores (MAS).

However, there have been no formal evaluations of the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003) and the EPA (1996). This BRD was prepared for use by an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) expert panel review of HET-CAM as a method to identify ocular corrosives and severe irritants. Parallel reviews of the ICE, IRE, and BCOP test methods were conducted. Results of the Expert Panel Report, combined with the analyses presented in the BRDs, were used to support ICCVAM recommendations on the proposed standardized test method protocols, proposed list of recommended reference substances, and additional optimization and/or validation studies that may be necessary to further develop and characterize the usefulness and limitations of these methods.

1.2 Scientific Basis for the HET-CAM Test Method

1.2.1 Purpose and Mechanistic Basis of the HET-CAM Test Method

The HET-CAM is proposed to provide information on the effects that may occur in the conjunctiva following exposure to a test substance. Chicken-embryo models have long been used as models for embryotoxicity by virologists (Parish 1985; Luepke and Kemper 1986). Extending the use of chicken-embryos, the HET-CAM test method was proposed by Luepke (1985) and Luepke and Kemper (1986).

The CAM is a vascularized respiratory membrane that surrounds the developing bird embryo. The CAM is composed of an ectodermal layer that consists of epithelium that is two to three cells thick; a mesodermal layer that consists of connective tissue, ground substance, and blood vessels, and an endodermal layer (Parish 1985; Bruner 1992). The blood vessels that are present in the mesodermal layer of the CAM are branches from the embryo-allantoic arteries and veins. These vessels contain erythrocytes and leukocytes that are believed to be involved in the inflammatory response following exposure to external stimuli (Parish 1985). It was assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane are similar to effects induced by the same test substance in the eye of a treated rabbit (Luepke 1985; Luepke and Kemper 1986). The denaturation of proteins (observed as coagulation) is proposed to be an indicator of effects on epithelial cells in the CAM. Such effects are proposed to relate to adverse effects on the cornea of the eye. Alterations on the CAM blood vessels are a proposed predictor of overall toxicity and conjunctival damage in the eye.

1.2.2 Similarities and Differences of Modes and Mechanisms of Action Between the HET-CAM Test Method and Human Ocular Irritancy

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

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

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

Figure 1-1 Anatomy of the Human Eye

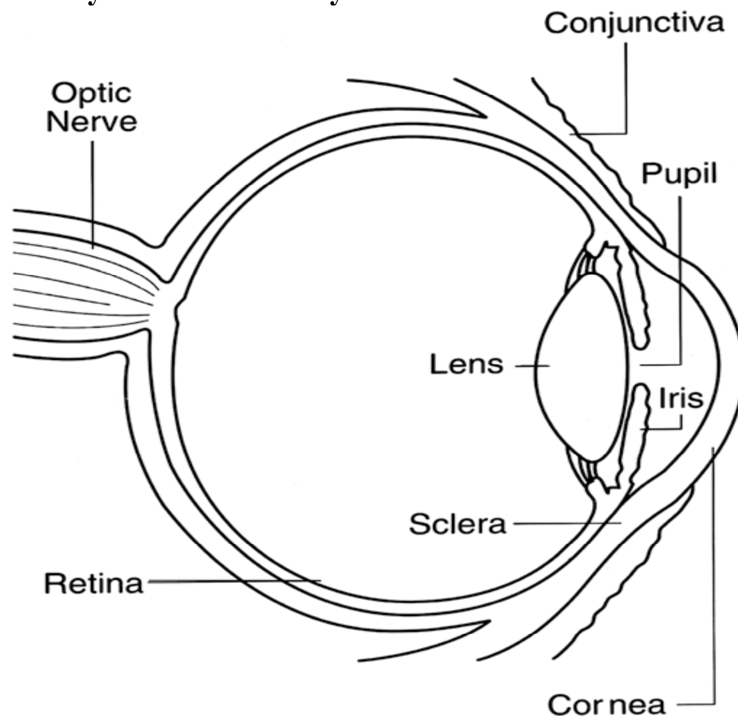


Figure obtained at http://www.nei.nih.gov/photo/eyean/index.asp

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

The eye is critically dependent on the highly vascularized middle coat (uvea) for regulation of blood and ocular permeability barriers, maintenance of intraocular pressure in the aqueous humor, and drainage of ocular fluid (Unger 1992). The uveal tract is richly innervated by somatic sensory neurons, derived from the ophthalmic division of the trigeminal nerve. Importantly, alterations to any of these features (e.g., edema, cell destruction, vascularization, cell proliferation) can cause corneal opacity and concomitant loss of function (Parish 1985; Wilkie and Wyman 1991; Bruner 1992).

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

part of the inflammatory response in the conjunctiva, dilation of the blood vessels, fluid leakage, and cellular leakage occurs (Bruner 1992).

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

1.2.2.2 *Differences Between Human and Rabbit Eyes*

There are several anatomical and physiological differences between the rabbit eye and the human eye. One difference is the presence of a nictitating membrane, or third eyelid, in the rabbit. As this membrane slides horizontally across the eye, it is proposed that it aids removing and/or excluding irritating substances from the corneal surface (Calabrese 1983). It also is proposed that the kinetic removal of a substance from a rabbit eye may occur at a rate different than in humans, due to the presence of the nictitating membrane, although this has not been documented in comparative studies (Curren and Harbell 1998). Another difference is the larger conjunctival sac in the rabbit, which allows for larger test volumes to be instilled, perhaps more than could be accounted for on accidental exposure (Curren and Harbell 1998).

The rabbit cornea is thinner than that found in humans, and rabbits tend to have less tear production (Curren and Harbell 1998; Cooper et al. 2001). The thicknesses of structural components of the cornea also are different between the two species. For example, Descemet's membrane is proposed to be about 5 to 10 μm in humans and 7 to 8 μm in rabbits (Calabrese 1983). Furthermore, the area of the cornea in relation to the total surface of the globe varies significantly between species; in humans the relationship is 7%, while in rabbits the relationship is 25% (Swanston 1985). Finally, young rabbits have the ability to regenerate damaged corneal endothelium, while humans do not (Chambers W, personal communication).

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

1.2.2.3 *The In Vivo Rabbit Eye Test Method*

The current *in vivo* rabbit eye irritation test method evaluates the cornea, the iris, and the conjunctiva for adverse effects after exposure to a potential irritant (see **Section 4.0** for a discussion of the *in vivo* scoring system for lesions at these sites). The cornea is visually observed both for the degree of corneal opacity and the area of the cornea in which opacity is involved. The iris is assessed for inflammation, iridal folds, congestion, swelling,

circumcorneal injection, reaction to light, hemorrhage, and gross destruction. The conjunctiva is evaluated for the degree of redness, chemosis (swelling), and discharge (Draize et al. 1944). Draize and colleagues (1944) developed an analysis method where the severities of the effects are weighted differently, with corneal effect being weighted the most. The effects of a test substance on the cornea, conjunctiva, and iris play a role in severe ocular irritant and corrosive labeling and classification in classification systems used by some regulatory agencies (CPSC 1995; EPA 1998; EU 2001; UN 2003).

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

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

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

nourish dependent corneal tissue. Changes in the distribution, physical structure, or secretion rate of the tear film by an applied test substance might have significant nutritional, refractory, chemical and physical impacts on corneal tissue (Mishima and Hedbys 1968a, 1968b).

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

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

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

Corneal damage also may be characterized by morphological changes (e.g., described as stippling, ulceration, mottling, pannus, neovascularization).

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

While not in the direct optical path, the Palisades of Vogt, located in the sclero-corneal limbus, are thought to house corneal stem cells and serve as a generative organ for normal replacement of dead corneal epithelial cells for re-epithelialization during repair of corneal

injury. Depletion or partial loss of the limbal stem cell population may result in corneal vascularization due to loss of the barrier function of the limbus, which serves to prevent conjunctival epithelial cells from migrating to the corneal surface (Dua and Azuara-Blanco 2000).

Neutrophils are recruited in response to acid and alkali injury as well as in response to other ocular toxicants (Pfister 2005). Neutrophil migration is stimulated by the release of chemotatic factors (e.g., interleukins, growth factors, etc.) from damaged or chemically activated local resident epithelial cells or stromal keratocytes (Wilson et al. 2001). Loss of keratocytes following either chemical or mechanical epithelial injury may be mediated by apoptosis, perhaps by release of IL-1 and TNFα (Wilson et al. 2001). Resident mast cells may release biogenic amines that perturb the hydrostatic balance and permit inflammatory or edemagenic mediators into the locally inflamed area. Migrated neutrophils release additional cytokines (e.g., IL-1 and TNF-α) and enzymes such as proteases, collagenases, kinases, and phospholipaseA_2 (PLA_2). PLA_2 produces edemagenic and vasoactive mediators such as prostaglandins and leukotrienes from arachidonic acid in cellular membranes.

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

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

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

1.2.2.4 *The Chorioallantoic Membrane (CAM)*

The HET-CAM test method uses the CAM, which is a vascular fetal membrane, composed of the fused chorion and adjacent wall of the allantois. The chorion is the outermost sac that contains the embryo. It is found in most high-level vertebrates, and in the chicken it serves to contain the amnion and yolk sac. The CAM is composed of three layers. The layer first seen when the eggshell is opened is the ectodermal layer, which is two to three cells thick. The next layer, a mesodermal layer, is comprised of blood vessels, connective tissues, and ground substance. The inner layer is referred to as the endodermal layer and is composed of squamous cells (Parish 1985).

The allantois develops appears from the hindgut, as an outgrowth, starting at about 60 hours of incubation (Tufan and Satiroglu-Tufan 2005). The allantois then pushes out from the hindgut of the chick embryo on incubation day 4 or 5 (Tufan and Satiroglu-Tufan 2005). It is composed of endoderm and splanchnic mesoderm (Sinn-Harlon 1998a). The allantois has four major functions in maintaining chick embryo viability: 1) serve as an embryonic respiratory organ; 2) store kidney excretions; 3) absorb albumen for the embryo; and 4) absorb calcium from the eggshell for the embryo (Clauer 2002). As the allantois increases in size, between incubation days four and 10, it wraps around the embryo and fuses with the chorion to form the CAM (Tufan and Satiroglu-Tufan 2005). The fusion of the two membranes allows for a free exchange of gases between the embryo and the outside environment (Sinn-Harlon 1998a). After formation of the CAM, there is rapid growth in the surface area until incubation day nine (Tufan and Satiroglu-Tufan 2005).

Irritation responses in the CAM are limited, likely due to the immaturity of the immune system in the embryo (Bruner 1992). Studies indicate that there are few heterophils (neutrophils in chickens) and macrophages in the chick embryo. Additionally, the macrophages that are present in the embryo do not accumulate in damaged tissue as is seen in mammals (Lawrence et al. 1986). Lesions on the CAM appear to be due to necrosis in the area of application of the test substance (Parish 1985).

1.2.2.5 *Comparison of the HET-CAM Test Method with the In Vivo Rabbit Eye Test Method*

Comparison of the HET-CAM and *in vivo* rabbit eye test methods focuses on a comparison of the CAM to the mammalian eye. Comparison of the CAM to the structures of the eye indicates that it is most similar to the conjunctiva. Both structures are mucous membranes that contain a functional vascular system. However, the CAM is much thinner than the conjunctiva and contains an ectodermal layer that is more primitive than the conjunctiva (Parish 1985). Unlike organotypic test methods (e.g., IRE, ICE, and BCOP), corneal responses such as opacification and swelling are not evaluated in the HET-CAM test method.

Irritation responses in the CAM and conjunctiva are shown to occur upon exposure to irritants. However, the actual responses of the CAM and conjunctiva to irritants are significantly different. Conjunctival irritation typically leads to neutrophil infiltration and macrophage accumulation. Comparatively, CAM irritation leads to cell death in the area of the insult (i.e., location of test substance application). Anatomical differences and relative

immaturity of the immune system in the egg (and thus the CAM) are proposed to contribute to these different responses.

In addition to subjectively evaluating corneal opacity and effects on the iris and conjunctiva, the *in vivo* rabbit eye test evaluates the delayed onset and/or reversibility of any ocular effects detected. The HET-CAM assay does not take into account effects on these other structures in the eye, assess reversibility, or attempt to identify slow-acting irritants.

Finally, HET-CAM does not account for systemic effects following ocular instillation that may be noted with the *in vivo* rabbit eye test (e.g., toxicity or lethality as in the case of certain pesticides).

1.2.3 Intended Range of Substances Amenable to the HET-CAM Test Method and/or Limits of the HET-CAM Test Method

Studies indicate that the HET-CAM test method is amenable for use with a broad range of solid and liquid substances with few limitations. Substances amenable to testing include, but are not limited to, inorganic chemicals; aliphatic, aromatic, and heterocyclic chemicals; surfactants; polymers; and mixtures/formulations.

One limitation of the test method is that test substances that are colored, turbid, or adhere to the CAM may inhibit visualization of the CAM. In some currently used HET-CAM protocols, the CAM is exposed to test substances and the CAM is observed during that entire exposure period. However, colored test substances may not allow for clear and complete evaluation of an adverse effect. To allow for a clear assessment, such substances may be rinsed off the CAM or diluted to a concentration that allows for clear and complete evaluation of the CAM. The rinsing procedure would therefore not allow for a continuous exposure and observation, as is performed for non-colored test substances.

Another potential limitation of the test method is that it can be used only for short-term assessments of the irritancy of a test substance. The currently accepted *in vivo* test method usually observes the rabbits for up to 21 days after treatment to assess reversibility of any of the observed endpoints and to evaluate test substances that produce eye effects over an extended time period. Comparatively, the observation period for evaluating effects in the HET-CAM test method post-treatment is up to five minutes. Therefore, potential reversibility of the affected endpoint beyond five minutes or an effect with a delayed onset (e.g., slow-acting irritants) cannot be adequately evaluated with this test method.

1.3 Regulatory Rationale and Applicability

1.3.1 Current Regulatory Testing Requirements and ICCVAM Prioritization Criteria

The following section reviews and summarizes the extent to which HET-CAM addresses the five ICCVAM prioritization criteria apply to the HET-CAM test method (ICCVAM 2003).

Criteria 1. The extent to which the proposed test method is (a) applicable to regulatory testing needs, and (b) applicable to multiple agencies/programs.
The HET-CAM assay has been proposed as a method to identify ocular corrosives or severe irritants, as is required by several U.S. laws. **Table 1-1** identifies the U.S. agencies and programs that classify and label substances for eye irritation and corrosion. These agencies include the FDA, EPA, Department of Agriculture, Department of Labor, the Consumer Products Safety Commission (CPSC), and the Chemical Safety and Hazard Investigation Board. Therefore, the HET-CAM test method is applicable to the regulatory testing needs of multiple U.S. Federal agencies and programs.

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

Criteria 3. The potential for the proposed test method, compared to current test methods accepted by regulatory agencies, to (a) refine animal use (decreases or eliminates pain and distress), (b) reduce animal use, or (c) replace animal use.[2]
The HET-CAM test method has the potential to refine or reduce animal use in eye irritation testing. Substances that are identified as ocular corrosives or severe irritants would be excluded from testing *in vivo*, which would reduce the number of rabbits used for ocular testing. The HET-CAM method also would spare animals the pain and distress of exposure to severe eye irritants.

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

---

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

Criteria 5. The extent to which the test method provides other advantages (e.g., reduced cost and time to perform) compared to current methods.
The HET-CAM test method could reduce the time needed to assess a substance, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days and can be extended for up to 21 days. It is noted that for some substances (i.e., severes) the test may be completed within an hour after application of a test substance. Completion of the HET-CAM test method requires a nine-day pre-treatment incubation period, followed by approximately one hour for the treatment and observation/measurement period. The current cost of a GLP compliant EPA OPPTS Series 870 Acute Eye Irritation (EPA 1998) or Organization for Economic Cooperation and Development (OECD) Test Guideline (TG) 405 test (OECD 2002) at MB Research Laboratories (Spinnerstown, PA) ranges from $765 for a three day/three animal study up to $1,665 for a 21 day/three animal study (MB Research Laboratories, personal communication). The current costs of performing a GLP-compliant HET-CAM test have not yet been identified but are expected to be equivalent to or lower than the cost of an *in vivo* rabbit eye test.

1.3.2 Intended Uses of the Proposed HET-CAM Test Method

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

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

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

As detailed in **Section 1.2.2**, the CAM used in the HET-CAM test method is used as a model for a living membrane (such as the eye conjunctiva), since it comprises a functional vasculature and can be evaluated for other endpoints that are associated with ocular injuries. The HET-CAM test method evaluates the development of adverse effects on blood vessels (e.g., hemorrhage, coagulation, hyperemia, injection, and/or vessel lysis). The endpoints evaluated in the HET-CAM test method are not similar to those evaluated in the *in vivo* test method (redness, chemosis and discharge from the conjunctiva), but are proposed to represent mechanisms of toxicity that could elicit these *in vivo* endpoints.

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

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

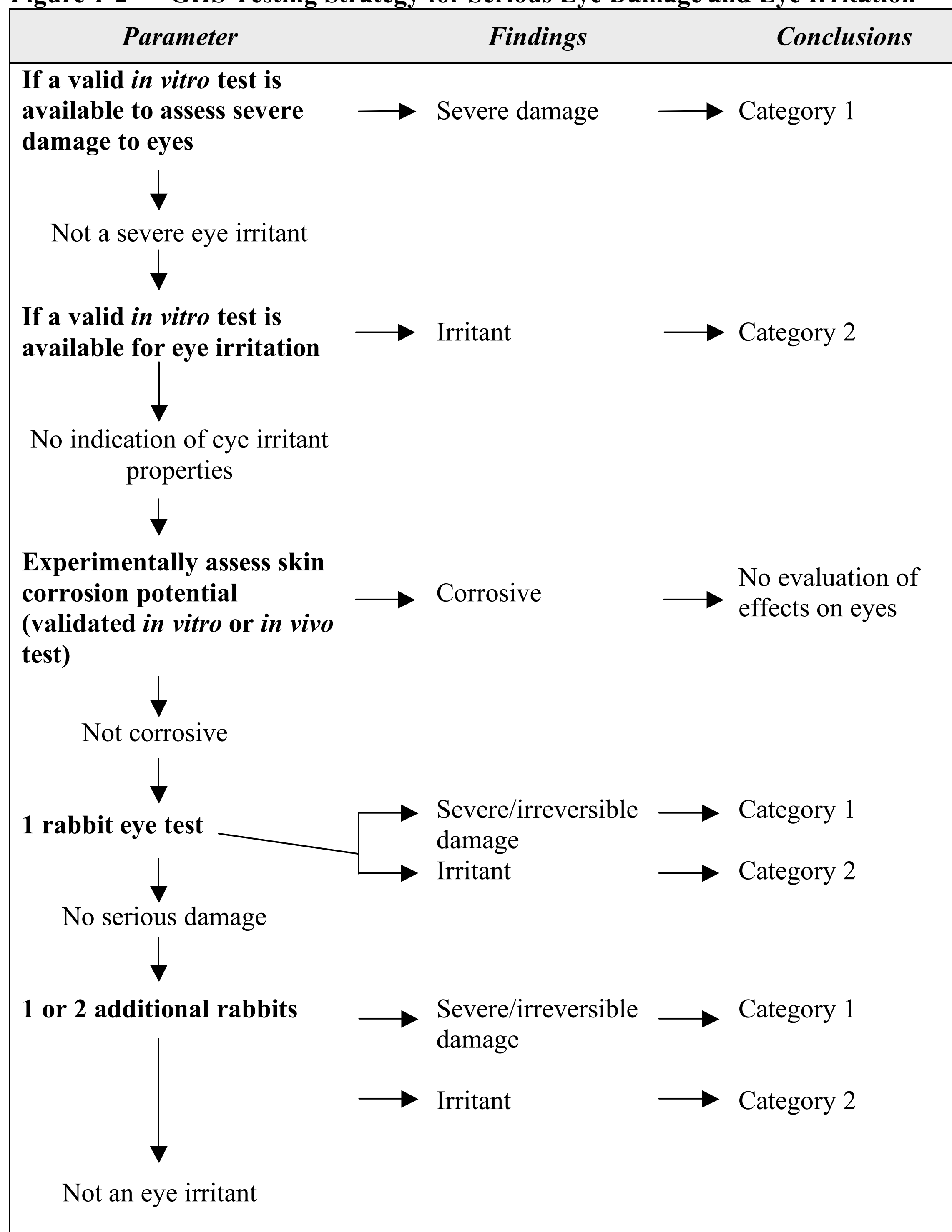
1.4 Validation of *In Vitro* HET-CAM Test Method

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

Validation is the process by which the reliability and relevance of an assay for a specific purpose are established (ICCVAM 1997). Relevance is defined as the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 1997). For the HET-CAM test method described in this BRD, relevance is restricted to how well the assay identifies substances that are capable of producing corrosive or severe irritant effects to the eye. Reliability is defined as the reproducibility of a test method within and among laboratories and should be based on performance with a diverse set of substances that are representative of the types of chemical and product classes that are expected to be tested and the range of response that needs to be identified. The validation process will provide data and information that will allow U.S. Federal agencies to develop guidance on the development and use of the HET-CAM test method as part of a tiered-testing approach to evaluating the eye irritation potential of substances.

The first stage in this evaluation is the preparation of a BRD that presents and evaluates the relevant data and information about the assay, including its mechanistic basis, proposed uses, reliability, and performance characteristics (ICCVAM 1997). This BRD summarizes the available information on the various versions of the HET-CAM test method that have been published. Where adequate data are available, the qualitative and quantitative performances of the assay are evaluated and the reliability of each version of HET-CAM is compared with the reliability of the other versions. If there are insufficient data to support the recommendation of a standardized protocol for HET-CAM, this BRD will aid in identifying essential test method components that should be considered during its development and validation.

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



Adapted from UN (2003).

1.5 Search Strategies and Selection of Citations for the HET-CAM BRD

The HET-CAM test method data summarized in this BRD are based on information found in the peer-reviewed scientific literature. An online literature search of entries in MEDLINE, ALTBIB, Web of Science, and STN International was conducted to retrieve database records on publications reporting on *in vitro* testing of substances using the HET-CAM test method. The search was conducted in the database basic index, which includes words in the title and abstract, and indexing words. Search terms used in various database indexes are shown in **Table 1-3**.

Table 1-3 Terms and Phrases Used for Online Literature Searches

Database Searched	Search Term
ALTBIB	hen's egg
PubMed and Web of Science	HET-CAM
PubMed and Web of Science	hen's AND egg* AND membrane
STN International	hen's AND egg*
STN International	hen's AND egg* AND membrane
STN International	(chorioallantoic OR (chorion AND allantoic)) AND test
STN International	HET-CAM

*represents wildcard character used in the search term.

Each database record included authors, bibliographic citation, and indexing terms. Most records also included abstracts.

Of the 128 records obtained from the search in ALTBIB, MEDLINE, and Web of Science completed in November 2003 (and updated in October 2004), 38 records contained relevant information on HET-CAM test method protocols and/or contained data obtained using the HET-CAM method. Of the 86 records obtained from the search in STN International completed in February 2004, 13 records contained relevant information on additional HET-CAM test method protocols and/or contained data obtained using the HET-CAM method. Abstracts of selected titles were reviewed, and the relevant articles were selected and retrieved from the literature for analysis. A database of the literature citations was established using bibliographic or reference database software. Subsequent to the initial search, additional articles with relevant information were identified and retrieved; many of these were identified from the bibliographies of the articles that were selected initially.

[This Page Intentionally Left Blank]

2.0 HET-CAM TEST METHOD PROTOCOL COMPONENTS

2.1 Overview of How the HET-CAM Test Method is Conducted

The HET-CAM protocol was first described by Luepke (1985). According to the original test method, fertilized hen's eggs are incubated, under optimized conditions, for nine days. On the 10th day, the eggs are opened and the CAM exposed. Then, 0.3 mL of the test substance is applied to the surface of the CAM. After a 20-second exposure period, the CAM is rinsed with 5 mL of water. The CAM is evaluated for development of irritant endpoints (hyperemia, hemorrhage, and coagulation) at 0.5, 2, and 5 minutes after rinsing off the test substance. Irritant effects in the CAM blood vessels and albumen are subjectively assessed and a score is assigned based on the time required for development of each endpoint. The scores are totaled to yield a total irritation score for the tested substance (maximum score of 21).

Since the initial description of the HET-CAM test method, several studies have been conducted to evaluate the feasibility of using HET-CAM as a complete replacement for the *in vivo* rabbit ocular test. Most of these reports describe a HET-CAM test method protocol that is similar, but not identical, to the original protocol. These differences include the breed of hen from which eggs are obtained, the endpoints evaluated, data collection procedures, and methods used to analyze the data. To date, no single HET-CAM test method protocol has gained wide acceptance as a standard protocol.

2.2 Description and Rationale for the Test Method Components

Currently, there is no widely accepted, single HET-CAM test method protocol for detecting ocular corrosives and severe irritants. The essential principles of the test method protocol include exposing the CAM, treating the CAM with a test substance, observing the development of endpoints, and evaluating the data in relation to a prediction model. However, given the various uses and applications of the HET-CAM test method by different investigators and laboratories, and the evolution of the assay over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method. Variations in the publicly available protocols include different prediction models or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations. These test method protocol differences are described in detail in **Section 2.2.1**, where variations in specific test method components for the HET-CAM test method are discussed.

The following sections describe in detail the major components of the HET-CAM test method protocol. Similarities and differences in the test method components of available HET-CAM protocols are discussed. For many of these components, no rationale for inclusion in the HET-CAM was provided in the published literature; in such cases, historical use is considered the rationale. **Appendix A** provides a comparative summary of test method components for all protocols reviewed.

2.2.1 Materials, Equipment, and Supplies Needed

2.2.1.1 *Hen Breed*

A limited number of hen breeds have been used for obtaining eggs for HET-CAM studies. Most studies used the White Leghorn breed of hens for egg production. Specific strains of White Leghorn hens that were used included Shaver Starcross 288A, Lohmanns Select LSL, and Leghorn SA31. Additional breeds of chickens used to supply eggs were White Essex and Lohmann Brown.

No rationale was provided in the reports for the selection of a specific hen breed or strain. Furthermore, no information was provided in the reports that indicated that there were differences in the results obtained from studies using eggs from different hen breeds. Therefore, while a formal study to determine the optimal hen breed to use for these studies was not found in the published scientific literature, use of the White Leghorn breed appears to produce consistent results. It also is noted that White Leghorns and their descendants are the most numerous today and are prodigious egg layers (Anonymous 1996).

2.2.1.2 *Criteria for Egg Use*

Published HET-CAM protocols are inconsistent in describing the criteria for selecting eggs for use. Most reports fail to note specific criteria (such as age of eggs, weight of eggs) that should be met prior to use in the test method. Those reports that do note criteria for egg selection and use typically state that the eggs weighed between 50 and 80 grams (Luepke 1985; CEC 1991; Bagley et al. 1992; de Silva et al. 1992; Spielmann 1995). Additional criteria noted for egg selection include (1) viability, as determined by candling the eggs, and (2) age (eggs were not older than seven days) (Gilleron et al. 1996, 1997; Budai et al. 1997).

While no rationale was provided in the scientific literature for the use of these criteria, studies indicate that excessively large or small eggs may have less than optimal embryonic growth and development while eggs with cracked or thin shells may affect retention of the necessary moisture for proper embryonic development (Lowther D, personal communication).

2.2.1.3 *Egg Incubation Conditions (Temperature, Relative Humidity, Egg Rotation During Incubation, and Length of Incubation)*

Temperature: There is little variation in the incubation temperature used for the eggs. A majority of incubation temperatures ranged from 37 to 38ºC (e.g., Luepke 1985; Kalweit et al. 1987; de Silva et al. 1992; Hagino et al. 1993; Gilleron et al. 1997; Spielmann et al. 1997; Schlage et al. 1999). The most common temperature used for incubation was 37.5 ± 0.5ºC (Luepke 1985; Blein et al. 1991; CEC 1991; Bagley et al. 1992; Rougier et al. 1992; INVITTOX 1992; Spielmann 1995).

Relative Humidity: The relative humidity used during the incubation period either was not reported or differed among the studies. A majority of reports indicate that the average relative humidity used during incubation ranged from 40% to 65%. The most commonly reported average relative humidity was 62.5%, with standard deviations of 1.5% or 7.5% (Luepke 1985; Blein et al. 1991; CEC 1991; Bagley et al. 1992; Rougier et al. 1992; INVITTOX 1992; Spielmann 1995; Gilleron et al. 1996). Other reported relative humidity

values include about 70%, 60% to 70%, 80%, and 80% to 90% (Hagino et al. 1991,1993, 1999; Kojima et al. 1995; Budai et al. 1997; Budai and Várnagy 2000; Brantner et al. 2002; Demirci et al. 2003, 2004).

Egg Rotation During Incubation: Some of the published reports discussed rotation of the eggs during incubation. Information about rotation frequency, however, was infrequently provided. When reported, eggs were rotated either once an hour or twice an hour (van Erp et al. 1990; Hagino et al. 1993, 1999; Kojima et al. 1995; Brantner et al. 2002). Dr. med Horst Spielmann indicated that eggs incubated for the HET-CAM test method were rotated five times per day (Spielmann H, personal communication). Generally, where reported, the eggs were rotated up to the day prior to application of the test substance (Spielmann 1995; Gilleron et al. 1996; Budai et al. 1997; Budai and Várnagy 2000).

Length of Incubation: In most studies, the eggs were incubated for either 9 or 10 days[1] (e.g., Luepke 1985; Kalweit et al. 1987; Sterzel et al. 1990; CEC 1991; Spielmann 1995). Several of these studies reported that the eggs were evaluated for viability by a candling technique on the day prior to use in the test method. Those eggs that were considered defective or nonviable were excluded from the study. Viable eggs were returned to the incubator, with the larger end placed upwards, and not rotated for the remainder of the incubation period. In contrast to the typical 9- or 10-day incubation period, three references reported an incubation period of approximately three days (Brantner et al. 2002; Demirci et al. 2003, 2004). Interestingly, these same investigators also used high relative humidity conditions (80% to 90%) compared to the majority of test method protocols reviewed.

Published sources indicate that proper development of the chicken embryo occurs when the incubation temperatures range from 35.6 to 38.9ºC for still-air incubators and 37.5 to 38.1ºC for forced-air incubators. The optimal relative humidity ranges, from 50% to 60%, were not dependent on the temperature of the incubator or the incubator type (Clauer 2000; Smith 2000; Anonymous 2002, 2004a, 2004b). The core range of relative humidity used was from 55% to 60%.

Hand rotation of the eggs on the eight days prior to use in the test method prevents attachment of the CAM to the shell (Spielmann 1995). Candling of the eggs prior to the last day ensures that the eggs used in the test method are viable and have developed normally during the incubation period. After establishing the viability of the eggs, placing the eggs back into the incubator, with the large end facing upwards, ensures accessibility to the CAM (Spielmann 1995).

In the embryo, the allantois begins to emerge from the hindgut of the chicken on incubation/gestation days four or five, but does not fuse with the chorion until incubation/gestation day six to form the chorioallantoic membrane (Sinn-Harlon 1998b; Clauer 2002). Therefore, based on the development of the embryo, the earliest incubation/gestation day that could be used in the HET-CAM test method is incubation day seven. The use of a nine-day incubation period is based on chicken embryo development and

---

[1] The difference in the days of incubation usually depended upon whether investigators defined the first day of incubation as Day 0 or Day 1.

international regulations that define when a chicken-based test is no longer considered an *in vitro* test method. International regulations have provisions for the protection of animals used for experimental or other scientific purposes. Some provisions indicate the time in which a test method using an animal embryo or fetus is considered an animal and therefore protected by the regulations. According to some of these regulations, a bird is considered a protected animal (and therefore the test is considered an *in vivo*, and not *in vitro* test) when greater than half of the gestation or incubation period has elapsed (day 10.5 for the 21 day incubation period of the chicken embryo) (Animals [Scientific Procedures] Act 1986; EU 1986). In the United States, the Public Health Service Policy, with which all National Institutes of Health (NIH)-funded research projects must comply, covers all live vertebrate species. The NIH Office of Laboratory Animal Welfare has provided written guidance in this area, interpreting "live vertebrate animal" to apply to avians only after hatching (Kulpa-Eddy J, personal communication; NIH 2000).

It has been proposed that at incubation day nine, the embryonic differentiation of the chicken central nervous system is sufficiently incomplete that suffering from pain perception is unlikely to occur (MSPCA 2005; Liebsch M, personal communication). Evaluations suggest that there are few sensory fibers present at day nine in the avian embryo and that there is significant development of the sensory nerve ending between incubation days 11 and 14 (Romanoff 1960). Studies also have suggested that the extraembryonal vascular systems (e.g., yolk sac, CAM) are not sensitive to pain (Rosenbruch 1997; Spielmann H, personal communication). Combined, these studies suggest that at incubation day nine there is little to no pain perceived by the developing embryo during the conduct of the HET-CAM test method.

2.2.1.4 *Rotary Saw*

Many of the test method protocols reviewed discussed a procedure for opening the eggshell prior to application of test substances. These protocols described using a dentist rotary saw, small rotary saw, or scissors to scratch the eggshell around the air space or cut away at the eggshell. The shell was then removed to expose the inner membrane. Only one literature reference provided information on the type of saw used to scratch the eggshell. Wilson and Steck (2000) described using a Dremel™ tool (Multi-Pro model 225T2) with a coarse 25 mm cutting disk to cut away the eggshell around the marked air space. A rationale for using this specific tool was not provided in the report.

The use of a rotary saw or tool with a cutting disk appears to be the most efficient way to remove a portion of the eggshell and allow access to the inner membrane and CAM.

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

Dose-selection procedures are not relevant to this *in vitro* test method. As described in **Section 2.2.4.2**, test substances are applied as neat chemicals (solids or liquids) if possible. Substances should be diluted with a preferred solvent, if technical limitations require the evaluation of a diluted test substance.

2.2.3 Endpoints Measured

A variety of endpoints were evaluated in the HET-CAM test method protocols reviewed. The original protocol described by Luepke (1985), as well as many additional protocols, evaluated the CAM for development of hemorrhage (blood from a ruptured vessel), hyperemia (increased blood flow), and coagulation (presence of blood clots) after application of the test substance (Blein et al. 1991; Hagino et al. 1991, 1993, 1999; Bagley et al. 1992; Rougier et al. 1992; de Silva et al. 1992; Kojima et al. 1995; Doucet et al. 1999). Other endpoints that were evaluated to assess the irritancy potential of test substances include injection (mild hemorrhage), vasoconstriction (narrowing of the vessels), dilation (expansion of the vessels), and lysis (disintegration of the vessels) (e.g., Luepke and Kemper 1986; Sterzel et al. 1990; CEC 1991; Spielmann 1995; Macián et al. 1996; Gettings et al. 1996; Budai and Várnagy 2000). The rationale for evaluation of these specific endpoints was not provided.

A combination of three endpoints was typically evaluated by each test method protocol. However, the combination of evaluated endpoints was not consistent. Most test method protocols evaluated the development of hemorrhage and coagulation in combination with one of the other endpoints noted. No rationale was provided for the combination of endpoints used in the various HET-CAM protocols.

One test method protocol called for the evaluation of the development of an "anti-angiogenic" effect (Demirci et al. 2003, 2004). This protocol was used to evaluate the anti-inflammatory and anti-angiogenic properties of various test substances. Therefore, the endpoint selected for this test method protocol was likely unique for these effects.

Prevalidation studies conducted in Germany evaluated the reproducibility of various HET-CAM endpoints. These studies indicated that some of the endpoints (e.g., injection, hyperemia) were not reproducible within or between laboratories (Spielmann H, personal communication); the most reproducible endpoints were hemorrhage, lysis, and coagulation. Therefore, the endpoints that appear to be the most reproducible (i.e., lysis, coagulation, and hemorrhage) are the endpoints that appear to be evaluated the most often during the performance of the HET-CAM test method.

2.2.4 Exposure Parameters

2.2.4.1 *Test Substance Exposure Amount or Volume*

The quantities of the test substance used in the HET-CAM test method were provided for liquids and solids. However, there was no consensus on the amount or volume of the test substance used in the protocols and no rationale was provided for the amounts used. For liquids, many of the studies used 0.2 mL or 0.3 mL of test substance (e.g., Luepke 1985; Kalweit et al. 1987; CEC 1991; Hagino et al. 1991; Gettings et al. 1994). Additional volumes used in the reviewed protocols were 0.2 to 0.3 mL, 0.1 to 0.2 mL, and 0.1 mL (Luepke and Kemper 1986; Gettings et al. 1996; Lönnroth et al. 1999). For solid forms of test substances, amounts between 0.1 g and 0.3 g were used (e.g., Luepke 1985; CEC 1991; Hagino et al. 1993; Gilleron et al. 1996, 1997).

No rationale was provided in the reports for the selection of specific test volumes. Furthermore, no information was provided in the reports that indicated that there were differences in the results obtained from studies using different volumes and/or amounts. Therefore, while a formal study to determine the optimal test volume and/or amount was not found in the published scientific literature, use of amounts and volumes that sufficiently cover the CAM appear to produce consistent results.

2.2.4.2 *Concentration Tested*

A variety of test substance concentrations have been evaluated in the HET-CAM test method. Concentrations tested for liquid and solid test substances have ranged from 1% to neat. Several test method protocols indicated that colored or turbid test substances, in particular, were diluted to allow for evaluation of the CAM after application of the test substance.

Historical use generally supports testing substances at a variety of concentrations. However, it is recognized that concentrations may require adjustment for certain chemical or product classes depending on the physical form and/or color of the test substance.

2.2.4.3 *Application of Test Substance to CAM*

Application of test substances to the CAM depended on the physical form of the test substance. Several HET-CAM test method protocols did not provide information on the quantity of solids applied to the CAM, suggesting that all test substances were tested at a standard amount or were solubilized prior to application (e.g., Kalweit et al. 1987; Vives et al. 1997). For liquids or solutions, most protocols indicated that the substances were placed directly onto the CAM (e.g., Kalweit et al. 1987; Vives et al. 1997). For those studies where solid forms of the test substances were used, different application techniques were noted. Most of the protocols indicated that the solid test substances were placed directly onto the CAM. However, a few studies identified the use of a rubber ring or similar device to contain the test substances to a specific area of the CAM (Gilleron et al. 1996, 1997; Hagino et al. 1999). The test substance applicator (TSA) described in Gilleron et al. (1996, 1997) consisted of a perlon mesh locked between two Teflon rings. The substance was placed in the ring and after treatment, it was removed to allow for evaluation of the CAM.

Two protocols required producing agarose pellets containing the test substance, which was then applied to the CAM surface (Demirci et al. 2003, 2004).

The only rationale located in the reports on the methodology used during the application of the test substance was for the use of a device to confine a test substance to a specific location on the CAM. The rationale for such a device was to reduce variability in the area covered by different test substances. Additionally, there was some discussion in the reports that at least 50% of the CAM surface should be covered with solid, paste, or particulate test substances to ensure high quality results. While a formal study to determine the optimal test volume and/or amount was not found in the published scientific literature, direct application of substances to the CAM (with or without a device that would contain the substances to a specific location) appears to produce consistent results.

2.2.4.4 *Test Substance Exposure Duration*

The exposure duration of the test substance differed among published reports. Several protocols noted that after a 20-second exposure of the CAM, all test substances were rinsed (typically with water) from the CAM (e.g., Luepke 1985; Luepke and Kemper 1986; Lawrence et al. 1990; Hagino et al. 1999). One protocol indicated that test substances were rinsed from the CAM after a three-minute exposure (Balls et al. 1995). Some test method protocols indicated that only insoluble or colored, opaque, or turbid substances were rinsed after a user defined time period (Spielmann et al. 1993, 1996; Spielmann 1995). This distinction suggests that the exposure duration for clear and/or soluble test substances was greater.

Rationales for the various exposure durations and discussion of optimal exposure duration were not located. Historical use of these exposure periods and criteria for rinsing suggest that these procedures are appropriate to assess the irritancy potential of test substances (Spielmann H, personal communication).

2.2.5 Known Limits of Use

A review of the available literature indicates that there are some test substance characteristics that could have an impact on the accuracy and/or reliability of the HET-CAM test method. The accuracy of irritancy prediction of test substances that are: (a) colored or turbid, or (b) that adhere to the CAM may be compromised. Such substances may impair clear visualization of the CAM during the study and thus could lead misclassification of the substance. Therefore, the usefulness of the test method for testing such substances undiluted may be limited.

2.2.6 Nature of the Response Assessed

The endpoints evaluated in the HET-CAM test method were evaluated by visual inspection of the CAM surface after test substance application. Depending on the type of data compiled from the test method, the response could be considered qualitative or semi-quantitative.

Several test methods evaluated if the endpoint of interest had developed at a specified time point and/or the severity of the response elicited by the test substance and recorded a qualitative score.

Other test methods observed the CAM after application of the test substance and recorded the time (in seconds) when an effect occurred. These data are considered semi-quantitative since the development of an endpoint is subjective and dependent upon interpretation of the investigator, but the data collected (the time after test substance application when an endpoint developed) is a quantitative value.

No rationale was provided in the reports for the selection of a specific response assessed. Therefore, while evaluating either type of response (qualitative or semi-quantitative) may be appropriate for evaluating ocular toxicity it appears that historical evaluation of semi-quantitative data is more accurate.

2.2.7 Appropriate Controls and the Basis for their Selection

2.2.7.1 *Negative Controls*

Of those test method protocols that identified an appropriate negative control, the most commonly identified substance was 0.9% NaCl (Vinardell and Macián 1994; Budai et al. 1997; Spielmann et al. 1997; Lönnroth et al. 1999; Budai and Várnagy 2000). An agarose pellet also was identified by one set of studies as a negative control (Demrici et al. 2003, 2004). No rationale was provided for the choice of negative control substance used. It appears that the commonly used negative controls for the HET-CAM test method offer no distinct advantages or disadvantages, except for their use for substances with differing solubilities.

However, it is clear that a negative control is useful in the HET-CAM test, so that nonspecific changes in the test system can be detected. This type of control also provides a baseline for the assay endpoints, and ensures that the assay conditions do not inappropriately result in an irritant response. Any of the commonly used negative controls is acceptable as long as the same negative control is used consistently within a laboratory.

2.2.7.2 *Positive Controls*

Of those test method protocols that identified an appropriate concurrent positive control, the most commonly identified substances were sodium dodecyl sulfate (SDS), laureth-8-sulfate, and sodium hydroxide (NaOH) (e.g., Reinhardt et al. 1987; Sterzel et al. 1990; INVITTOX 1992; Spielmann 1995; Macián et al. 1996; Budai et al. 1997; Spielmann et al. 1997; Vives et al. 1997). When used as positive controls, these substances were used at relatively low concentrations. SDS was used at a concentration ranging from 0.1% to 1%, with 1% being the most commonly used concentration. NaOH was typically used at a concentration of 0.1M. Additional positive controls that were identified included benzalkonium chloride (0.01% to 100%), *N,N*-dimethylformamide (DMF; 0.1% to 100%), and imidazole (0.1% to 100%) (Gilleron et al. 1996). No rationale was provided in the literature reference for the choice of positive control substances used.

Review of *in vivo* data provides some conflicting information about some of the test substances used as positive controls for the *in vitro* test method. Evaluation of data from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) database (ECETOC 1998) indicates that 3% sodium lauryl sulfate (SLS; also referred to as SDS) is a nonirritant according to the EU (2001) and GHS (UN 2003) classification systems and a Category III irritant according to the EPA (1996) classification system (see **Sections 1.0** and **4.0** for descriptions of various regulatory classification systems). Based on this data, it is likely that a 1% SDS solution (which is the concentration most typically used as a positive control) also would be considered a nonirritant according to the EU (2001) and GHS (UN 2003) classification systems and at most a Category III irritant according to the EPA (1996) classification system. Evaluation of the *in vivo* NaOH data indicates that a 1% solution is classified as an R36 irritant according to the EU classification system (EU 2001), a Category 2B irritant according to the GHS classification system (UN 2003), and a Category III irritant according to the EPA classification system (EPA 1996). Based on this data, it is likely that a 0.1 N NaOH solution (which is approximately a 0.4% solution) would be classified as a nonirritant or mild irritant by these different classification systems.

The rationale for the use of these positive controls, which are likely nonirritants or mild irritants *in vivo*, was not provided in the various study reports. Furthermore, there is no discussion in the reviewed literature of the use of different positive controls in the HET-CAM test method depending on the purpose of the test method.

Based on historical use in the HET-CAM test method, 1% SDS or 0.1 M NaOH are the most commonly used positive controls. Inclusion of a known severe ocular irritant substance in each experiment as a positive control demonstrates the functional adequacy of the test method and the consistency of laboratory operations in accurately identifying ocular corrosives and severe irritants. A positive control not only ensures the integrity of the test system and its proper execution, but also provides a measure of test method performance over time.

2.2.7.3 *Solvent Controls*

Although testing substances directly or undiluted is preferred, some HET-CAM test method protocols used a solvent to dilute the test substance. Of those studies that used solvents, there was little variability in which ones were used. The most commonly identified solvents were distilled or sterilized water, 0.9% sodium chloride (NaCl), and olive oil (INVITTOX 1992; Vinardell and Macián 1994; Spielmann 1995; Gilleron et al. 1996; Macián et al. 1996; Doucet et al. 1999). Carboxymethyl cellulose was used in one study (Macián et al. 1996) for water insoluble test substances. Two additional solvents described were cell culture media (MEM, 2 mM L-glutamine, 100 IU/mL penicillin, 100 mg/ml streptomycin, 5% fetal bovine serum) and 2.5% agarose (Lönnroth et al. 1999; Demrici et al. 2003, 2004). Special solubilization methods were described only for the study that listed cell culture media as the preferred solvent (Lönnroth et al. 1999). No rationale was provided for the choice of solvent used in the various studies.

It is clear that a solvent control is useful in the HET-CAM test, so that changes in the test system due to a solvent can be detected. This type of control also provides a baseline for the assay endpoints, and ensures that the assay conditions do not inappropriately result in an irritant response. Any of the commonly used solvent controls is acceptable.

2.2.7.4 *Benchmark Substances*

None of the study reports reviewed indicated the use of the benchmark controls. In the European Commission (EC)/British Home Office (HO) study (Balls et al. 1995), a reference standard (a 5% solution of Texapon ASV, an anionic surfactant) was used to evaluate the relative irritancy potential of the substances tested.

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

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

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

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

2.2.8.1 *Negative/Solvent Controls*

HET-CAM studies using 0.9% NaCl as a negative control, were conducted with and without the use of a TSA (Vanparys and Van Goethem 2005). The use of a TSA was described in Gilleron et al. (1996, 1997) (see **Section 2.2.4.3**). Over 90 tests using the TSA and three tests without using TSA were provided. As shown in **Table 2-1**, time to development of endpoints and the overall irritation scores calculated were consistent and classified as nonirritants for all tests (for additional information see **Section 7.0**).

Table 2-1 Means and Standard Deviations of Negative Control Test Substances Evaluated With and Without Use of the Test Substance Applicator

0.9% NaCl	Hemorrhage[1] (mean ± SD)	Lysis[1] (mean ± SD)	Coagulation[1] (mean ± SD)	*In Vitro* Score (mean ± SD)[2]
With TSA (n=92)	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Without TSA (n=3)	0 ± 0	0 ± 0	0 ± 0	0 ± 0

[1]Mean values of time until development of identified endpoint.
[2]*In Vitro* irritation score calculated as described in Kalweit et al. (1987, 1990). See **Section 2.2.12.1** for further details on this analysis method.

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

2.2.8.2 *Positive Controls*

HET-CAM studies using DMF as a positive control were conducted with and without the use of a TSA (for further information see **Section 2.2.4.3** and **Section 7.0**; Vanparys and Van Goethem 2005).

With the studies that were conducted with the TSA, the hemorrhage endpoint was evaluated inside the TSA and outside the TSA. Of note, the time of development of the hemorrhage endpoint inside the TSA was significantly lower than the time to development of the hemorrhage endpoint outside the TSA (see **Table 2-2**). The reason for the difference is not clear. Two proposed reasons for the difference in time to development are (1) the vessels outside the TSA may open more easily than those under the TSA, or (2) once the liquid is applied it the liquid accumulates around the edge of the TSA rather than between the TSA and CAM.

Table 2-2 Comparison of Means and Standard Deviations of Positive Control Test Substances Evaluated With and Without Use of the Test Substance Applicator

Positive Control	Hemorrhage[1] (mean ± SD)	Lysis[1] (mean ± SD)	Coagulation[1] (mean ± SD)	*In Vitro* Score (mean ± SD)[2]
With TSA[3] (n=69)	0.02 ± 0.17	6.93 ± 0.03	8.82 ± 15.77	15.77 ± 0.19
With TSA[3] (n = 10)	3.36 ± 0.32	6.54 ± 0.19	8.81 ± 0.04	18.71 ± 0.38
Without TSA (n = 2)	4.00 ± 0.13	6.84 ± 0.05	8.76 ± 0.08	19.60 ± 0.15
Imidazole (n=15)	4.50 ± 0.39	6.84 ± 0.08	8.66 ± 0.17	20.00 ± 0.45

[1]Mean values of time until development of identified endpoint.
[2]*In Vitro* irritation score calculated as described in Kalweit et al. (1987, 1990). See **Section 2.2.12.1** for further details on this analysis method.
[3]Hemorrhage endpoint in studies described in the first row were evaluated inside the TSA while hemorrhage endpoint in studies described in the second row was evaluated outside the TSA.

In addition to DMF, 100% imidazole also was evaluated as a positive control. In all studies with imidazole a TSA was used. As shown in **Table 2-2**, time to development of each endpoint evaluated and the overall irritation score calculated were consistent and classified as irritants for all tests with the substance.

HET-CAM studies using 1% SDS and 0.1 N NaOH also were provided upon NICEATM request (Spielmann and Liebsch 2005). Using the mean values determined for these studies, the overall irritation score calculated (according to the method of Kalweit et al. 1987, 1990) for these substances classified them as irritants (see **Table 2-3** and **Section 7.0**).

Table 2-3 Means and Standard Deviations of Positive Control Test Substances

Positive Control	Hemorrhage[1] (mean ± SD)	Lysis[1] (mean ± SD)	Coagulation[1] (mean ± SD)
1% SDS (n=377)	14.69 ± 5.36	35.18 ± 17.15	---[2]
0.1 N NaOH (n=336)	8.96 ± 4.96	35.60 ± 24.71	48.04 ± 34.56

[1]Mean values of time until development of identified endpoint.
[2]It was indicated that 1% SDS does not produce coagulation in the CAM after application. However, in the studies conducted coagulation was identified in a single study. In these evaluations, the non-existing data was calculated with an arbitrary value of "0". Therefore, the calculation of a mean value for the coagulation endpoint was not meaningful.

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

2.2.8.3 *Benchmark Controls*

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

2.2.9 <u>Nature of the Data to be Collected and the Methods Used for Data Collection</u>

Review of the literature indicates that the nature of the response assessed from the HET-CAM test method varied. The nature of the data collected depends on whether a qualitative response or semi-quantitative response was evaluated (see **Section 2.2.6**). In general, the HET-CAM test method protocols evaluated the CAM for (1) the time to development of observed endpoints after exposure to the test substance, (2) the highest or lowest test substance concentration needed to produce a minimal response on the CAM after exposure to the test substance, or (3) the maximum severity of response observed after exposure to the test substance. Some test method protocols described evaluating the CAM for two of the three of these responses. The following sections describe the three responses and the data collected by each, as described in the literature.

Visual inspection of the CAM is the only procedure described for collecting data.

2.2.9.1 *Time to Development of Observed Endpoints After Exposure to the Test Substance*

In studies that monitored the development of endpoints, two different data collection procedures were followed:

(A) One procedure evaluated the development of the response (i.e., endpoint) at specific time points after the test substance was applied to the CAM. In these reports, the blood vessels on the CAM were examined and scored for the presence of, for example, vessel lysis, coagulation, and hemorrhage at 0.5, 2, and 5 minutes after exposure to the test substance. The assigned score was time dependent (see **Table 2-4**); the earlier the endpoint developed in the 5-minute observation period, the higher the resulting score assigned to that endpoint (e.g., hyperemia developing before 0.5 minutes was assigned a score of 5 while hyperemia developing between two and five minutes was assigned a score of 1). The assigned scores were weighted with the highest scores being assigned to the coagulation endpoints. There was no rationale provided in the study reports on the scoring system or for the weighting in the scores used. Individual values for the observed endpoints were then used to assign the irritation potential of the test substance.

(B) Another procedure used to evaluate the time to development of endpoints after exposure to the test substance was to continually observe the CAM during the 5-minute observation period and record (typically in seconds) the time at which each of the endpoints developed (e.g., Kalweit et al. 1987, 1990; Sterzel et al. 1990; CEC 1991; Spielmann et al. 1991; Macián et al. 1996; Spielmann et al. 1996; Gilleron et al. 1997; Schlage et al. 1999). Therefore, three separate time values were obtained and recorded for each egg (one time value for each endpoint). Individual values for the observed endpoints were then used to determine the irritation potential of the test substance.

Table 2-4 HET-CAM Scoring System

Endpoint	Score at Different Observation Times		
	0.5 min	2 min	5 min
Hyperemia	5	3	1
Hemorrhage	7	5	3
Coagulation	9	7	5

Adapted from Luepke (1985).

None of the evaluated test method protocols provided guidance and/or information on the identification of any endpoints or what constituted development of each of the endpoints.

2.2.9.2 *Lowest Test Substance Concentration Needed to Produce a Minimal Response on the CAM After Exposure to the Test Substance*

Several test method protocols identified the use of a threshold concentration in evaluating the irritation potential of the test substance (e.g., Spielmann et al. 1993; Spielmann 1995). The lowest test substance concentration that produced a slight to weak irritant response on the CAM was recorded. The test method protocols that evaluated the threshold concentration did not provide guidance on what constituted a weak or slight response on the CAM. The identified test substance concentration was then used in combination with other evaluated responses to evaluate the irritation potential of the test substance.

2.2.9.3 *Severity of Response Observed After Exposure to the Test Substance*

One set of test method protocols evaluated the severity of each of the endpoints at a single time point. The severity of each of the endpoints was evaluated and scored based on a user-defined scoring scale; the scores ranged from 0 (no effect) to 3 (severe irritant effect) (Balls et al. 1995; Steiling et al. 1999). There was no guidance in either set of test method protocols as to the decision criteria for the different scores.

Historical use of all types of methods of collection appear useful in identification of ocular corrosives and severe ocular irritants. It is noted that collection of the time to development of each endpoint (see **Section 2.2.9.1 (B)**) allows investigators to use a variety of analysis methods to assess irritancy classification (see **Section 2.2.12**).

2.2.10 Type of Media in Which Data are Stored

One of the reviewed test method protocols noted that data were entered into and stored in electronic format on a computer (Spielmann et al. 1991). However, no additional information was provided. No other information on the type of media in which data are stored was noted in the published literature, but it can be assumed the data were written on data sheets and stored in a study book.

It would seem appropriate that data from the HET-CAM be stored and archived in a manner consistent with international Good Laboratory Practices (GLP) guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines are nationally and internationally recognized

rules designed to produce high-quality laboratory records. These guidelines provide a standardized approach to report and archive laboratory data and records, and information about the test protocol, to ensure the integrity, reliability, and accountability of a study (EPA 2003a, 2003b; FDA 2003).

2.2.11 Measures of Variability

There is little to no discussion in the published literature on how variability of response among replicate eggs within an experiment or among experiments conducted in the same laboratory is evaluated. Several of the studies indicated that the mean of the scores are taken as the final score. However, whether this is the mean of replicate eggs or replicate experiments is not clear.

Calculation of the mean score and SD provides the user with information on the performance of the test method. These values allow for an assessment of the performance of the test conducted and whether the observed variability between replicates is greater than would be considered acceptable.

2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data

As shown in **Section 2.2.9**, several types of data can be obtained from the HET-CAM test method. Therefore, depending on the data collected, different analysis methods were used to evaluate the irritancy potential of tested substances in the test method protocols. For data that evaluated the time to development of observed endpoints after application of the test substance to the CAM (**Section 2.2.9.1**), an Irritation Score (IS) or a “Q-Score” was determined. Alternatively, a mean detection time for the appearance of coagulation (mtc) was determined. For data that evaluated the lowest test substance concentration needed to produce a minimal response on the CAM after application of the test substance (**Section 2.2.9.2**), the Irritation Threshold Concentration (ITC) was determined. For data that evaluated the severity of response observed after application of the test substance (**Section 2.2.9.3**), an “S-Score” or Severity Irritation Score (SIS) was calculated.

2.2.12.1 *Irritation Scores*

A majority of the test method protocols calculated a score (referred to as irritation score, irritation index, or irritation potential) that represented the irritation potential of the test substance based on endpoint development. This score (referred to in this document as the IS value) could be determined by a variety of mathematical models. Several of the test method protocols did not provide the model used to determine the IS. However, the two main mathematical models are discussed below.

For those test method protocols that assigned a score to each of the endpoints evaluated, depending on the time the endpoint develops (see **Section 2.2.9.1 [A]**), the values assigned to each endpoint were totaled to give an overall IS value for the test substance (referred to in this document as the “IS[A]) analysis method”). The final IS value ranged from 0 (for test substances that do not induce development of any of the observed endpoints) to 21 (for test substances that induce development of all three endpoints within 30 seconds of application of the test substance).

For those test method protocols that noted the time that a specific endpoint was observed (see **Section 2.2.9.1 [B]**), an IS score could be calculated using the general formula (referred to in this document as the “IS[B] analysis method”):

$$\left(\left(\frac{(301 - \textit{Hemorrhage time})}{300}\right) \times 5\right) + \left(\left(\frac{(301 - \textit{Lysis time})}{300}\right) \times 7\right) + \left(\left(\frac{(301 - \textit{Coagulation time})}{300}\right) \times 9\right)$$

where:
Hemorrhage time = time (in seconds) of the first appearance of blood hemorrhages
Lysis time = time (in seconds) of the first appearance of vessel lysis
Coagulation time = time (in seconds) of first appearance of protein coagulation

The IS, when calculated using this formula, has a maximal value of 21.

There is no explanation of the source for the multiplication factors of 5, 7, and 9 in the above formula (Kalweit et al. 1987). However, it is likely that these values were selected to correspond to the highest score value possible for each of the endpoints in the IS(A) analysis method (see **Section 2.2.9**).

When development of hyperemia, injection, or another endpoint was evaluated in place of vessel lysis, the time point for the alternative endpoint replaced the lysis time point. There was no specific description in any of the test method protocols of statistical or nonstatistical methods used to analyze dose-response relationships described for either of the analysis methods.

Despite the fact that both analytical methods yield IS values that range from 0 to 21, care should be taken if values from these two methods are compared. The time to endpoint development ranges required to obtain IS values are different for each analytical method. For example, a maximum IS value of 21 can be obtained for the IS(A) analysis method when all three endpoints develop within 30 seconds after exposure of the CAM to the test substance. Comparatively, the same IS value can only be obtained by the IS(B) analysis method when all three endpoints develop within one second after exposure of the CAM to the test substance.

2.2.12.2 *Q-Score*

The Q-Score represents a comparison of the irritation potential of the test substance with that of a reference substance. To determine the Q-Score, the irritation potential of both the test substance and reference substance are evaluated. The irritation potential could be determined using any method, but typically are expressed as IS values. Of the test method protocols that evaluated the Q-Score, one did not provide the mathematical model used to determine the IS and the other used a weighted differential model. The IS of the test substance was then compared to the IS of the reference standard to calculate a ratio for the values, which was then used to assess the irritation potential of the test substance.

2.2.12.3 *Mean Detection Time*
For those test method protocols that noted the time that a specific endpoint was observed (see **Section 2.2.9.1 [B]**), the mean detection time for the appearance of an endpoint was determined. Mean detection times for the development of hemorrhage, lysis, and coagulation were evaluated, based on the times for three tested eggs (Spielmann et al. 1996).

2.2.12.4 *ITC*
Several test method protocols described using a combination of IS value and ITC to evaluate the irritation potential of a tested substance. The ITC was defined as the lowest concentration required to produce a slight or weak response on the CAM after application of the test substance. No definition was provided in the test method protocols for the terms "slight response" or "weak response." No data analysis or manipulation was required for this value, since the ITC was strictly the test substance defined as a percentage of the volume tested.

2.2.12.5 *S-Score*
Another set of test method protocols calculated a term defined as the S-Score. This score represents the highest total score for any endpoint evaluated for a test substance. For this score, the severity scores assigned for each endpoint (which ranged from 0 to 3 and were assigned at a single user-defined time point) were totaled for all of the replicate eggs evaluated per test substance to produce an endpoint total score. Therefore, there were three endpoint total scores per test substance. The endpoint that yielded the highest score was used as the final test substance S-Score. Many of the test method protocols that evaluated the irritation potential of test substances using this analysis method advocated the use of six eggs per test substance. In such cases, the maximal S-Score is 18.

According to this analysis method, the endpoints used to develop the S-Score could be different for different test substances. Thus, the S-Score could be defined by the development of coagulation for one test substance and hyperemia for another test substance.

2.2.12.6 *SIS*
A single test method protocol used this method to assess the anti-irritation potential of test substances (Demirci et al. 2003, 2004). The SIS is based on the potency of the anti-angiogenic effect produced by the test substance. Using the scoring range of 0 to 2, the ability of the test substance to produce an anti-irritant effect was noted. The number of eggs scored with a 0, 1, or 2 was inserted into the following formula to calculate the SIS:

$$\text{Average Score} = \frac{\text{Number of Eggs (Score 2)} * 2 + \text{Number of Eggs (Score 1)} * 1}{\text{Total Number of Eggs (Score 0,1,2)}}$$

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

A review of the test method protocols indicates that there is not a single set of decision criteria that has been used to classify test substances as producing a positive (i.e., irritant) or negative (i.e., nonirritant) response. Depending on the type of data collected and the method used to analyze the data, various irritation classification schemes and decision criteria have

been developed. Most of these schemes and criteria were developed by individuals and do not typically correlate *in vitro* scores with *in vivo* irritation classification schemes, such as those used by the EPA (1996), EU (2001), and GHS (UN 2003). However, as detailed below, some correlations were noted.

2.2.13.1 *Irritation Classification Based on IS*

Many of the test method protocols that calculated IS values (either by the IS[A] or IS[B] analysis methods) used a similar classification scheme. Ranges of *in vitro* scores were assigned to one of four irritation categories. The major ranges of scores for each category are shown in **Table 2-5**.

Table 2-5 Summary of HET-CAM Score Ranges Used in Irritancy Classification

HET-CAM Score Range	Irritation Category
0-0.9	Nonirritant or Practically None
1-4.9	Weak or Slight Irritation
5-8.9 or 5-9.9	Moderate Irritation
9-21 or 10-21	Strong or Severe Irritation

From Luepke (1985) and Kalweit et al. (1987).

In addition to the IS ranges listed above, several other irritation classification schemes have been described. For example, van Erp and colleagues (1990) used the following classification scheme for IS values: Nonirritant ≤ 0.5; Slight Irritant = 0.5 to 3.4; Moderate Irritant = 3.5 to 4.9; Severe Irritant ≥ 5. Gilleron et al. (1996, 1997) classified test substances as either nonirritant or irritant. In these studies, substances inducing HET-CAM IS values from 0 to 4.9 were defined as nonirritant while those inducing scores of 5.0 and above were defined as irritants.

One of the test method protocols correlated *in vitro* HET-CAM test method scores with a classification system used by a regulatory agency. The studies conducted by Gettings and colleagues (1991, 1994, 1996) used different criteria to classify test formulations according to the FHSA classification system (CPSC 1988). Using this system, in which substances are defined as irritants or nonirritants, a range of IS values were used. The range depended on the test method protocol and the data analysis method. **Table 2-6** provides the decision criteria used by the various test method protocol analysis methods to classify a test substance IS value as an irritant according to the FHSA classification system.

According to the published reports, the *in vitro* values used to classify a test substance as an irritant or nonirritant were determined *post hoc* and were deliberately chosen to minimize the number of false positives or false negatives (Gettings et al. 1994, 1996).

Table 2-6 Decision Criteria used by Gettings and Colleagues to Define Test Substances as Irritants According to the FHSA Classification System

Literature Source	IS Value for Irritants
Gettings et al. 1991 (HET-CAM I)	> 10
Gettings et al. 1991 (HET-CAM II)	> 300[1]
Gettings et al. 1994 (HET-CAM)	≥ 5
Gettings et al. 1994 (Chorioallantoic Membrane Assay)	≥ 4.8
Gettings et al. 1996 (HET-CAM I)	≥ 5.1
Gettings et al. 1996 (HET-CAM III)	≥ 4.83

[1]The analysis method used for this evaluation was different than previously described (see **Section 2.2.12**). The mathematical model of Bartnik et al. (1987) was used for this evaluation. However, information could not be located on this model to assess the range of scores that could be obtained.

2.2.13.2 *Irritation Classification Based on Q-Score*

Ranges of Q-Scores were assigned to one of three or four irritation categories. A summary of the ranges of scores used in the classifying the irritancy potential of a test substance is shown in **Table 2-7**.

Table 2-7 Summary of Q-Scores Used in Irritancy Classification

Q-Score (Range 1)[1]	Q-Score (Range 2)[2]	Irritation Category
< 1.5	-	Nonirritant
-	≤ 0.8	Slight
1.5 ≤ Q < 2	0.8 < Q < 1.2	Moderate
-	1.2 ≤ Q < 2	Irritant
≥ 2	≥ 2	Severe

[1]From Balls et al. (1995).
[2]From Brantom et al. (1997).

No information was provided on the correlation of Q-Scores to irritation categories defined by the GHS (UN 2003), EPA (1996), or EU (2001) ocular irritation classification systems.

2.2.13.3 *Irritation Classification Based on Mean Coagulation Detection Time*

A range of mean coagulation times when using a 10% solution (mtc10) values were assigned to one of two EU irritation categories. The studies conducted by Spielmann and colleagues (1996) used different criteria to classify test substances according to the EU classification system (EU 1992). Using this system, in which substances are defined as R41 or Remainder (R36 and nonirritants), a range of mtc10 values were used. The ranges provided in **Table 2-8** were based on different analyses and sets of data used in the evaluation. **Table 2-8** provides the decision criteria used to classify a test substance mtc10 value as an irritant according to the EU classification system.

Table 2-8 Summary of mtc10 Values Used in Irritancy Classification

mtc10 (Range 1)	mtc10 (Range 2)	mtc10 (Range 3)	Irritation Category
<174 seconds	< 139 seconds	< 50 seconds	R41
≥ 174 seconds	≥ 139 seconds	≥ 50 seconds	Remainder

From Spielmann et al. (1996).

2.2.13.4 *Irritation Classification Based on IS and ITC*

Several studies based irritation classifications on combinations of IS and ITC values. **Table 2-9** provides an example of a combination of scores defined in one study for various irritation categories (Spielmann et al. 1996).

Table 2-9 Summary of ITC and IS combinations Used in Irritancy Classification

ITC (% Concentration)	Irritation Score (10% Concentration)	Irritation Category (EU Category)[1]
> 10%	< 16	None/slight (Nonirritant)
> 10%	> 16	Moderate (Nonirritant)
< 10%	< 16	Moderate (Nonirritant)
< 10%	> 16	Irritant (R36)
< 2.5%	< 16	Irritant (R36)
1% < ITC ≤ 2.5%	≥ 16	Severe (R41)
≤ 1%	-	Severe (R41)

From Spielmann et al. (1996).
[1]EU (1992, 2001).

The combination of IS and ITC values used to define various irritation categories were similar, but not identical, between test method protocols. For example, **Table 2-9** shows that when a test substance had an ITC greater than 10% and the IS was less than 16, the test substance was classified as none/slight (nonirritant). In contrast, another test method protocol classified the same combination as a moderate irritant (Spielmann 1995).

Two of the reviewed test method protocols utilized IS and ITC scores to assign a classification that corresponded to an *in vivo* irritation classification system used by a regulatory agency. The study conducted by Spielmann and colleagues (1996) classified test substances according the classification system used by the EU (1992). The correlation of IS and ITC values to this classification system is provided in **Table 2-9**. Spielmann and colleagues (1996) reported that the prediction model was developed empirically and was not supported by statistical methods.

The study conducted by Gettings et al. (1996) classified test formulations according to the classification system defined in the FHSA (CPSC 1988). For this analysis, the ratio of the IS value to the ITC value was determined. A final value equal to or greater than three was

defined as an irritant, as defined by the FHSA (CPSC 1988). According to the published reports, the *in vitro* values used to classify as an irritant or nonirritant were determined *post hoc* and were deliberately chosen to minimize the number of false positives and false negatives (Gettings et al. 1996).

2.2.13.5 *Irritation Classification Based on S-Score*

Ranges of S-Scores were assigned to one of three or four irritation categories. A summary of the classification system using this value is shown in **Table 2-10**.

Table 2-10 Summary of S-Scores Used in Irritancy Classification

S-Score (Range 1)[1]	S-Score (Range 2)[2]	Irritation Category
< 6	-	Nonirritant
-	< 6	Slight
6 ≤ S < 15	6 ≤ S ≤12	Moderate
-	12 < S < 16	Irritating
≥ 15	≥ 16	Severe

[1]From Balls et al. (1995).
[2]From Brantom et al. (1997).

No information was provided on the correlation of S-Scores to irritation categories defined by the GHS (UN 2003), EPA (1996), or EU (2001) ocular irritation classification systems.

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

There was no description provided in the studies reviewed about the information and data that was included in the study reports for this test method. There are no known standard forms for data collection and submission.

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

Test and Control Substances

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known
- The CAS Registry Number (RN), if known
- Purity and composition of the substance or preparation (in percentage(s) by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- Stability, if known

Information Concerning the Sponsor and the Test Facility

- Name and address of the Sponsor
- Name and address of any test facilities involved
- Name and address of the Study Director

Justification of the Test Method and Protocol Used

Test Method Integrity

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

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data
- If applicable, acceptable concurrent benchmark control ranges based on historical data

Test Conditions

- Experimental starting and completion dates
- Details of test procedure used
- Test concentration(s) used
- Description of any modifications of the test procedure
- Reference to historical data of the model (e.g., negative and positive controls, proficiency substances, benchmark substances)
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., irritancy scores for the test substance and the positive, negative, and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means and the standard deviation for each experiment)

Description of Other Effects Observed

Discussion of the Results

Conclusion

A Quality Assurance Statement for GLP-Compliant Studies

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

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

2.3 Basis for Selection of the Test Method System

The HET-CAM is proposed to provide information on the effects that may occur in the conjunctiva following exposure to a test substance. Published reviews note that chicken-embryo models have long been used as models by embryotoxicologists and virologists. (Parish 1985; Luepke and Kemper 1986). Extending the use of chicken embryos, the HET-CAM test method was proposed by Luepke (1985) and Luepke and Kemper (1986). It was assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane are similar to effects induced by the same test substance in the eye of a treated rabbit. The CAM has been proposed as a model for a living membrane (such as the conjunctiva) since it comprises a functional vasculature.

Additionally, evaluation of coagulation (i.e., protein denaturation) may reflect corneal damage that may be produced by the test substance. It has been inferred that only very mildly irritating substances are capable of inducing conjunctival effects in the absence of corneal effects (Prinsen M, personal communication). Therefore, this would not appear to limit the effectiveness of the HET-CAM with respect to predicting severe irritants.

2.4 Proprietary Components

There are no proprietary components used in the HET-CAM test method.

2.5 Basis for the Number of Replicates and Repeat Experiments

2.5.1 Sample Replicates

The number of replicate eggs used per test substance or test concentration varied between protocols. Several protocols did not describe or discuss the number of replicates used in the method. Of those test method protocols that did describe the number of replicate eggs used per test substance or test concentration, most of the methods described using three, four, or six eggs. Three test method protocols described using 10-15 or 20 replicate eggs per test substance (Brantner et al. 2002; Demrici et al. 2003, 2004). Comparison of results from studies conducted with three eggs to those conducted with six eggs previously indicated no significant difference (Spielmann H, personal communication).

2.5.2 Experimental Replicates

Few studies reviewed provided information on the number of replicate experiments conducted for each test substance. A sample of the reviewed studies indicates that one conducted a single experiment, five repeated the experiment two times, four repeated the experiment three times, and three repeated the experiment four times. Additionally, one investigator stated that a total of four eggs were used per test substance and that the experiment could either be conducted a single time with four eggs, or could be conducted two times with two eggs used in each experiment. There was no discussion provided in any of the sources on the optimal number of repeat experiments and no rationale was provided for selecting a specific number of repeats. However, based on sound scientific judgment, it

would seem reasonable to expect that equivocal responses or divergent results among test cornea would mandate repeating the experiment.

2.6 Compliance with Good Laboratory Practices

Compliance with GLP guidelines only can be evaluated by the information that was provided in the published reports. No attempt was made to review original records to assess the quality of the data presented. Based on the available information, the only reports that were identified to have followed GLP guidelines or used data obtained according to GLP guidelines were Gettings et al. (1991, 1994, 1996), Spielmann et al. (1993, 1996), Balls et al. (1995), Brantom et al. (1997), Hagino et al. (1999), and Steiling et al. (1999).

Conducting studies under GLP guidelines increases confidence in the quality and reliability of test data. Furthermore, if data using this test method is to be submitted to the EPA or another agency in response to Federal testing requirements, then compliance with appropriate GLP guidelines will be required.

2.7 Study Acceptance Criteria

There was no description provided in the test method protocols reviewed about the study acceptance criteria.

A test should be considered acceptable if the negative/solvent and positive controls each give an values that falls within the classification of nonirritating or irritating, respectively. A test also should be considered acceptable if the benchmark control (if used) produces an irritation response that is within acceptable limits of historical data.

[This Page Intentionally Left Blank]

3.0 SUBSTANCES USED FOR VALIDATION OF THE HET-CAM TEST METHOD

3.1 Rationale for the Substances Selected for Use

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

In general, both criteria were used to select the substances used in the studies considered in this BRD: CEC (1991); Gettings et al. (1991, 1994, 1996); Bagley et al. (1992); Vinardell and Macián, (1994); Balls et al. (1995); Kojima et al. (1995); Gilleron et al. (1996, 1997); Spielmann et al. (1996); and Hagino et al. (1999).

3.1.1 CEC (1991)

The selection of substances used in this evaluation was based on the following criteria:

- The substances should be representative of currently used industrial chemicals and should represent a range of chemical structures.
- The substances should cover the range of eye effects from nonirritant to severe irritant.
- The *in vivo* rabbit eye studies should have been conducted in accordance with European Economic Commission (EEC) criteria and the animal data should be sufficient to allow an irritancy classification to be definitively assigned to the test substance.
- Whenever possible, the substances should have been used in previous validation studies.

3.1.2 Gettings et al. (1991, 1994, 1996)

The studies described in this set of papers focused evaluating the ability of alternative test methods to identify ocular corrosives and severe irritants that are developed by the cosmetic, toiletry, and fragrance industries. Therefore, for this evaluation a set of formulations were developed that were representative of cosmetic, toiletry, and fragrance formulations used at the time of the study.

3.1.3 Bagley et al. (1992)

The studies described in paper focused evaluating the ability of alternative test methods to identify ocular corrosives and severe irritants. Therefore, substances that were (1) raw materials commonly used in the cosmetics/toiletries and household cleaning product industries, and (2) formulations representing products from these industries were evaluated.

3.1.4 Balls et al. (1995)

In the EC/HO validation study (Balls et al. 1995), the test substances were initially selected from the 1992 European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Reference Data Bank for ocular irritation (ECETOC 1992) based on the following criteria:

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

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

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

The final list of test substances included a total of 51 substances, four of which were tested at two different concentrations and two of which were tested at three concentrations, for a total of 59 different tests.

3.1.5 Vinardell and Macián (1994)

There was no specific rationale for the selection of substances used by Vinardell and Macián (1994) provided in the literature reference.

3.1.6 Kojima et al. (1995)

Kojima et al. (1995) evaluated substances that were major ingredients in cosmetic formulations and preparations. These substances included surfactants and solvents.

3.1.7 Gilleron et al. (1996, 1997)

Gilleron et al. (1996, 1997) selected substances that represented a broad spectrum of ocular irritancies, chemical classes, and chemical structures. Substances also were selected on the basis of availability of historical *in vivo* data, to avoid conducting additional tests for the validation study.

3.1.8 Spielmann et al. (1996)

Spielmann et al. (1996) selected substances that represented a broad spectrum of ocular irritancies, chemical classes, and chemical structures. Substances also were selected on the basis of availability of historical *in vivo* data, to avoid conducting additional tests for the validation study.

3.19 Hagino et al. (1999)
Hagino et al. (1999) evaluated substances that were major ingredients in cosmetic formulations and preparations. These substances included surfactants and solvents.

3.2 Rationale for the Number of Substances Tested

No rationale was provided for the number of substances tested in these studies except for Spielmann et al. (1996). Spielmann et al. (1996) noted that the Amden validation workshop recommended that approximately 200 substances should be used to assess the performance of an alternative test method. Therefore, they originally selected a total of 200 substances for their effort to validate HET-CAM (Phase I and Phase II). The number was reduced to a total of 118 substances, after substances were excluded due to unacceptable *in vivo* or *in vitro* data quality.

3.3 Chemicals or Products Evaluated

Physicochemical properties for each of the tested substances was obtained from information provided in the published reports and submitted data. No attempt was made to review original records in order to obtain additional information about each substance. For each substance tested in HET-CAM, **Appendix B** provides information on its CASRN, chemical and/or product class. **Appendix C** provides information on the physicochemical properties (e.g., pH, physical form tested), where available from the published reports or submitted data.

Chemical and product classes were assigned for each test substance based on information found in the study report. If a chemical class was not assigned in the study report, such information, when available, was retrieved from the National Library of Medicine's ChemID Plus database, or assigned based on chemical structure. Chemical classes were assigned to each test substance using a standard classification scheme, based on the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at http//www.nlm.nih.gov/mesh) that ensures consistency in classifying substances among all *in vitro* ocular test methods under consideration. If a product class was not assigned in the study report, such information, when available, was retrieved from publicly available sources that discussed the substance. A substance could be assigned to more than one chemical or product class.

Table 3-1 provides the chemical class information on the test substances evaluated with HET-CAM. The chemical classes with the greatest number of substances tested are alcohols, carboxylic acids, and organic salts. Of the substances included in **Appendix B,** 53 were formulations. For some of the test substances that were identified as formulations, components of the formulation and the relative concentrations of the components were available. Summaries of the relative concentrations of each component in these formulations are provided in **Appendices B-2** to **B-4**.

Table 3-1 Chemical Classes Tested in the HET-CAM Test Method

Chemical Class	# of Substances
Acyl halide	2
Alcohol	75
Aldehyde	9
Alkali	4
Amide	2
Amidine	6
Amine	34
Amino acid	7
Carbohydrate	1
Carboxylic acid	51
Ester	34
Ether	38
Formulation	53
Heterocyclic compound	37
Hydrocarbon, Acyclic	5
Hydrocarbon, Cyclic	5
Inorganic boron compound	2

Chemical Class	# of Substances
Inorganic salt	14
Imide	4
Ketone	15
Lactone	5
Nitrile	3
Nitro compound	3
Onium compound	22
Organic salt	50
Organometallic compound	2
Organophosphorous compound	1
Organosilicon compound	6
Phenol	4
Polycyclic compound	11
Organic sulfur compound	18
Unknown	28
Urea	3

Table 3-2 provides the product class information on the test substances evaluated with HET-CAM. The most common product classes tested are solvent, shampoo, surfactants, and cosmetics. Of the substances included in **Appendix B,** 167 were not be classified within a product class.

3.3.1 Substances Evaluated in Reviewed Studies

3.3.1.1 *CEC (1991)*

This report described the results of a study commissioned by the Division Control of Chemicals, Industrial Risks and Biotechnologies of Directorate General Environment, Nuclear Safety, Civil Protection and the Health and Safety Directorate of Directorate General Employment Industrial Relations and Social Affairs. In this study, 21 substances were evaluated. All substances appear to have been tested as 100% or 10% concentrations. The authors provided purity of the tested substances but not other physicochemical properties. The authors used the IS(B) analysis method to evaluate the irritancy potential of the test substances.

Table 3-2 Product Classes Tested in the HET-CAM Test Method

Product Class	# of Substances
Aerosol formulation ingredient	1
Anti-freezing agent	1
Anti-infective agent, Anti-bacterial agent	2
Anti-perspirant	1
Bactericide, Biocide, Fungicide, Germicide	4
Beverage	1
Cationic surface active agent	1
Chemical intermediate	6
Cleaner	1
Conditioner, Hair	2
Cosmetics	14
Cream	1
Disinfectant	1
Drug vehicle	1
Emollient	2
Fertilizer	1
Flavor ingredient	5
Fragrances	4
Industrial explosive	1
Laboratory reagent	7
Lotion	3
Lubricant	1
Mouthwash	1
Neurotransmitter	2
Pesticide	5
Pharmaceutical agent, Pharmaceutical intermediate, Pharmaceutical metabolite	4
Plasticizer	2
Polymer	1
Preservative	1
Raw material	1
Shampoo, Hair	13
Solvent	13
Sunscreen	3
Surfactant	17
Synthetic flavor ingredient, Flavor ingredient	4
Synthetic intermediate	1
Unknown	167

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for 15 test substances. Chemical classes of the tested substances included, but were not limited to, alcohols, esters, and carboxylic acids. Product classes of the tested substances included, but were not limited to, bactericide and surfactant.

3.3.1.2 *Gettings et al. (1991)*

This report described results from Phase I of the Cosmetic, Toiletry, and Fragrance Association (CTFA) Evaluation of Alternatives Program, a program that evaluated promising *in vitro* alternative test methods for the *in vivo* rabbit eye test. Each phase of the program evaluated a specific product type. Phase I (1991) evaluated 10 hydroalcoholic formulations. Formulations were generic formulations that represented formulations used at the time of the study (e.g., facial cleaner). All formulations in Phase I were tested undiluted. The IS(B) analysis method was used to evaluate the irritancy potential of the test substances. The

product classes of the tested formulations included, but were not limited to, fragrances, mouthwash, and sunscreen.

Information (e.g., formulation components, physiochemical properties) was extracted for all formulations evaluated. Information on the components of the 10 formulations was obtained from the literature; this information is provided in **Appendix B-2**.

3.3.1.3 *Gettings et al. (1994)*
This report described results from Phase II CTFA Evaluation of Alternatives Program. Each phase of the program evaluated a specific product type. Phase II evaluated 18 oil/water formulations. Formulations were generic formulations used at the time of the study (e.g., conditioner). All formulations in Phase II were tested undiluted. The authors used the IS(A) and IS(B) analysis methods to evaluate the irritancy potential of the test substances. The product classes of the tested formulations included, but were not limited to, conditioner, sunscreen, and cream.

Information (e.g., formulation components, physiochemical properties) was extracted for all formulations evaluated. Information on the components of the 18 formulations was obtained from the literature; this information is provided in **Appendix B-3**.

3.3.1.4 *Gettings et al. (1996)*
This report described results from Phase III CTFA Evaluation of Alternatives Program. Each phase of the program evaluated a specific product type. Phase III evaluated 25 surfactant-based personal cleaning formulations. Formulations were generic formulations used at the time of the study (e.g., shampoo). In Phase III, nine of the substances were evaluated at a concentration of 25% (v/v) in distilled water. The IS(A) and IS(B) analysis methods were used to evaluate the irritancy potential of the test substances. The product class of the tested formulations was shampoo.

Information (e.g., formulation components, physiochemical properties) was extracted for all formulations evaluated. Information on the components of the 25 formulations was obtained from the literature; this information is provided in **Appendix B-4**.

3.3.1.5 *Bagley et al. (1992)*
In this study, 32 substances were evaluated; 12 were raw materials commonly used in cosmetics, toiletries, and household products and 20 were prepared formulations. All substances appear to have been tested as neat liquids or solutions. The authors did not provide information on the constituents of the formulations or the physicochemical properties of any of the tested substances in the reviewed study. The range of MAS values for the substances was 0.3 to 44.7. The source of the raw materials and the concentration tested were provided in the report. The authors used the IS(A) analysis method to evaluate the irritancy potential of the test substances.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for two of the raw materials. The chemical classes of these raw materials were ether and alcohol/amine. The range of MAS values of the substances extracted from this

study was 2.7 to 40.0. All substances appear to have been tested as neat liquids or solutions *in vitro*.

3.3.1.6 *Vinardell and Macián (1994*)

The study evaluated six vehicles and six commercial disinfectant products. The substances were tested at concentrations ranging from 0.1% to 100%. Other than pH of the tested solutions, which ranged from 3.3 to 13.02, no other physicochemical properties were provided. All substances appear to have been tested as neat liquids or solutions. In the study report, the authors did not provide any information about the ingredients of the disinfectant products. The IS(B) analysis method was used to evaluate the irritancy potential of the test substances.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for two vehicles. The pH values of these two vehicles were 3.3 and 7.2; both were tested as neat liquids *in vitro*.

3.3.1.7 *Balls et al. (1995)*

The study evaluated 51 substances with the HET-CAM test method. The substances were evaluated at concentrations ranging from 0.1% to 100%. Of these substances, 30 were water-soluble, 18 were water insoluble, and 12 were classified as surfactants by the study authors. Fourteen substances were tested as solutions, 20 were tested as solids, and 26 were tested as liquids *in vitro* and *in vivo*. For each substance, the authors provided in the report the CASRN, chemical class, source, catalog number, purity, and form tested. The S-Score and Q-Score analysis methods were used to evaluate the irritancy potential of the test substances. The chemical classes evaluated included, but were not limited to, amine, carboxylic acid, and organic salt.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for all substances.

3.3.1.8 *Kojima et al. (1995)*

In this study, 24 substances were evaluated with the HET-CAM test method. Solubility and other physicochemical properties of the test substances were not provided in the paper. For each substance, the authors provided in the report information on its source and the concentration tested (10% solution). The authors used the IS(B) analysis method to evaluate the irritancy potential of the test substances.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for five substances. The chemical classes of the extracted test substances were alcohol, organic salt, carboxylic acid salt, onium, and ether. All substances appeared to have been tested as solutions.

3.3.1.9 *Gilleron et al. (1996)*

The 46 evaluated substances were classified by the authors as solids (17), liquids (21), and surfactants (8). Chemical classes of the substances tested included alcohols, carboxylic acid, heterocyclic, and amine. Solubility and other physicochemical properties of the test

substances were not provided in the paper. For each substance, the CASRN and source were provided in Gautheron et al. (1994). The authors used the IS(B) analysis method to evaluate the irritancy potential of the test substances.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for all tested substances.

3.3.1.10 *Spielmann et al. (1996)*

In this study, 200 substances from the pharmaceutical and chemical industries were initially selected for evaluation; 34 substances were evaluated in Phase 1 and 166 substances were evaluated in Phase II of the study. All substances were tested at various concentrations to determine the threshold concentration for inducing an effect (ITC). Chemical classes of the tested substances included alcohol, amine, ester, ether, heterocyclic, and organic salt. Several analysis methods were used to asses the irritancy potential of the tested substances including, but not limited to, the IS and ITC analysis method and the mtc10 analysis method.

Information (e.g., CASRN, physiochemical properties) for all the substances evaluated by this study is provided in Spielmann et al. (1996). Information for 112 substances evaluated in the HET-CAM BRD was extracted for the IS(B)-10 and IS(B)-100 analysis methods.

3.3.1.11 *Gilleron et al. (1997)*

In this study, the 52 different substances were tested *in vitro* and compared to 60 *in vivo* studies. Solubility and other physicochemical properties of the test substances were not provided in the paper. The CASRN and physicochemical properties of the tested substances were detailed in Balls et al. (1995). Chemical classes of the tested substances included alcohol, amine, ester, ether, carboxylic acid, and heterocyclic. The authors used the IS(B) analysis method to evaluate the irritancy potential of the test substances.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for all tested substances.

3.3.1.12 *Hagino et al. (1999)*

In Phase III of a three-part validation study, 14 cosmetic ingredients were evaluated. For each substance, the authors provided in the report its chemical class, the concentration tested, and its physical form. The chemical classes of the tested substances included, but were not limited to, alcohol, carboxylic acid, organic salt, onium, and amine/amidine. Of these 14 substances, 12 were tested as a solution and two were tested as a suspension. The pH of the tested substances ranged from 2.4 to 12.48. The MAS of the tested substances ranged from 0 to 102.7. CASRN were obtained from the National Library of Medicine's ChemID Plus database. In the study, the authors used the IS(A) analysis method to evaluate the irritancy potential of the test substances.

Information (e.g., CASRN, chemical and/or product class, physiochemical properties) was extracted for all 14 cosmetic ingredients discussed above.

3.4 Coding Procedures Used in the Validation Studies

The coding procedures used in the studies considered in this BRD were evaluated by the information provided in the published reports. No attempt was made to obtain original study records to assess these procedures. Based on the available information, the reports that identified using coded chemicals were Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999).

3.4.1 Gettings et al. (1991, 1994, 1996)

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

3.4.2 Bagley et al. (1992)

The samples were transferred from original containers at a central coordinating point and then randomly coded from 1 to 32 prior to shipping to the participating laboratories.

3.4.3 Balls et al. (1995)

Test substances and participating laboratories were each assigned a numeric code in order for subsequent data analysis to be performed without knowledge of the identities of the test substance or laboratory. The total number of aliquots of each test substance required for the full study was determined. Computer software was then used to generate random codes for the total number of samples, so that a unique number could be assigned to each sample.

3.4.4 Spielmann et al. (1996)

The substances were coded prior to distribution to the participating laboratories. No information was provided in the report on how the substances were coded and/or tracked.

3.4.5 Hagino et al. (1999)

The Japanese Cosmetic Industry Association provided the test substances to the Test Substance Control Committee. The substances were then coded and distributed to the participating laboratories. No information was provided in the report on how the substances were coded and/or tracked.

[This Page Intentionally Left Blank]

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

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

4.1.1 Draize Rabbit Eye Test

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

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

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

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

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

[1]From Draize et al. (1944)
[2]Scores of 0 are assigned for each parameter if the corneal opacity, iris, or conjunctiva are normal.

4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols

Since the original description of the *in vivo* rabbit eye test method, regulatory agencies in the U.S., as well as in other countries, have modified the test method protocol to suit their specific needs and goals in protecting human health (**Table 4-2**). Regulatory agencies generally recommend using healthy adult albino rabbits (e.g., White New Zealand). The eyes of each test rabbit are examined within 24 hours prior to test initiation. A quantity of 0.1 mL (for liquids) or 0.1 g (for pulverized solid, granular, or particulate test substances) is placed into the conjunctival sac of one eye of each rabbit, after pulling the lower lid away from the eyeball. The other eye remains untreated. The lids are held together for about one second to decrease loss of test substance from the eye. Although the observation period

varies, the eyes are typically examined at 24-hour intervals for at least 72 hours after application of the test substance for adverse effects to the cornea, conjunctiva, and iris. The length of the observation period should be sufficient to evaluate reversibility of any of the observed effects, but generally does not exceed 21 days. The ocular effects observed are usually those described by Draize et al. (1944) in **Table 4-1**. For current uses, other lesions, such as pannus[1] and herniation of the cornea, also are noted. Corneal, iris, and conjunctival lesions are scored using the individual numerical grades described in **Table 4-1**, but weighted scores and an overall score for irritation are not typically calculated or used for U.S. or European regulatory purposes.

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

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Evaluate existing animal and human eye data	NA	Yes	Yes[1]	NS	Yes
Results from dermal irritation study	NA	Yes	Yes[1]	Yes	Yes
Perform SAR for eye irritation	NA	Yes	Yes[1]	NS	Yes
Screen for pH	NA	Yes	Yes[1]	Yes	Yes
Results from validated alternative ocular methods	NA	Yes	Yes[1]	Yes	Yes
Rabbit model/Number of rabbits					
Rabbit species and strain	Albino rabbit	Healthy young adult albino rabbits.	New Zealand White rabbit	Healthy adult albino rabbits recommended. Other mammalian species may be substituted with justification.	Healthy young adult albino rabbits.
Sex and weight	NS	NS	Sex NS; 2.0-3.0 kg	NS	NS

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

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Screen for severe effects	NS	1 rabbit – further testing not required if substance produces corrosive or severe effects.	NS	1 rabbit – further testing not required if substance produces corrosive or severe effects.	1 rabbit – further testing not required if substance produces corrosive or severe effects.
Main test/confirmatory test	NS	Up to 2 additional rabbits, tested sequentially. if irreversible effects are suspected. Test discontinued, if severe effects occur in 2nd rabbit. Additional rabbits may be needed to confirm weak or moderate responses.	A minimum of 6 rabbits, and up to 18 rabbits for confirmatory tests.	≥ 3 rabbits	Up to 2 additional rabbits, tested sequentially. if irreversible effects are suspected. Test discontinued, if severe effects occur in 2nd rabbit.
Test substance (amount and method of application)					
Liquids	0.1 mL	0.1 mL	0.1 mL	0.1 mL	0.1 mL
Solids, pastes, particulates	NS	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL or 100 mg
Aerosols	NS	Single burst of about 1 second sprayed at 10 cm.	NS	Single burst of about 1 second sprayed at 10 cm.	Single burst of about 1 second sprayed at 10 cm.
Pump sprays	NS		NS	0.1 mL	Should not be used for instilling liquid substances directly into the eye.
Application of test substance	Test substance is placed in the conjunctival sac.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 2004)**
Use of anesthetics prior to instillation of test substance	NS	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Local anesthetic may be used prior to instillation of test substance.	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Anesthetic may be used after 24 hours if it does not influence response of the eye to irritants.
Observation					
Observation Period	At least 48 hours. Extended if irritation persists.	At least 72 hours, except when rabbit shows severe pain or distress, or early severe/corrosive effects, upon which the rabbit is humanely killed. Otherwise, sufficient to evaluate reversibility or irreversibility within 21 days.	At least 72 hours. Extended if necessary.	At least 72 hours, but not more than 21 days. Should be sufficient enough to evaluate the reversibility or irreversibility of effects within a 21-day period.	At least 72 hours, except when rabbit shows severe pain or distress, or early severe/corrosive effects, upon which the rabbit is humanely killed. Can be extended up to 21 days if effects persist.
Examination times after treatment	1, 24, 48 hours, and 4, 7 days.	1, 24, 48, 72 hours, 7, 14, 21 days.	24, 48, 72 hours, and 7 days	1, 24, 48, and 72 hours. Extended up to 21 days to assess reversibility.	1, 24, 48, and 72 hours. Can be extended up to 21 days. Observations of mild to moderate lesions until they clear or for 21 days. Observations at 7, 14, and 21 days to determine reversibility.
Observation aids	NS	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours

Test Method Component	Reference				
	Draize et al. (1944)	OECD TG 405 (April 2002)	FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)	FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)	European Union Annex V B.5 (formerly EEC; EU 2004)
Irrigation					
Washout	NS	Generally, eyes may not be washed until after 24 hours post-treatment, except for solids, which may be removed with saline or water after 1 hour.	After 24 hours post-treatment, eyes may be washed with a sodium chloride solution.	After 24 hours post-treatment, eyes may be washed with water to show whether washing palliates or exacerbates irritation.	Generally, eyes may not be washed until after 24 hours post-treatment, except for solids, which may be removed with saline or water after 1 hour.
Additional testing to determine effects of timely irrigation	NS	Not recommended unless scientifically justified.	NS	Indicated when substances are shown to be irritating. At 30 seconds after exposure, the eyes are washed with water for 30 seconds.	Possibility of washing out in case of immediate corrosive or irritating effects. Use of satellite group to investigate influence of washing is not recommended, unless scientifically justified.

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

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

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

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

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

4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems

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

The FHSA guideline states that a test substance is considered an eye irritant if four or more of six rabbits have positive ocular scores in nonirrigated eyes within 72 hours after instillation of the test substance (CPSC 2003). A positive score is defined by corneal opacity or iritis scores of ≥ 1, or conjunctival redness or chemosis scores of ≥ 2. In addition, if only one of the six rabbits shows ocular effects within 72 hours, the test substance in considered nonirritating to the eye. If two or three rabbits have positive ocular scores, the test is repeated in a second group of six rabbits. Then, if the criteria for an ocular irritant for the second test (three or more positive rabbits) or a nonirritant (0 positive rabbits) are met, a classification is made. However, if only one or two rabbits have positive scores in the second test, the test is repeated a third and final time. If one or more rabbits have positive ocular scores in the third test, the test substance is classified as an ocular irritant. If none of the rabbits have positive ocular scores in the third test, the test substance is classified as a nonirritant (CPSC 2003).

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

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

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

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

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

In the CEC (1991) study, acute toxicity data for the substances evaluated were obtained from a literature review by Botham, Mckillop, and Purchase (Part II of CEC [1991]) of Imperial Chemical Industries (UK) and a chemical profile was produced for each substance except

dibutyltin chloride, which lacked *in vivo* ocular data[2]. In order to assign a EU ocular toxicity classification, the *in vivo* rabbit eye test method used to assess the ocular toxicity of the test substances needed to meet at least four criteria:

- 0.1 mL of the test substance was applied to the animal eye.
- A single eye per animal was used for the evaluation.
- Three to six animals were tested for each substance.
- The eye was not irrigated after application of the test substance.

In general, the eyes of tested rabbits were examined 24, 48, and 72 hours after instillation of the test substance. Scoring of injury was based on the Draize system (EEC 1984)[3]. These substances were classified by the authors according to EEC (1984) and used to assess the *in vitro* test method accuracy.

For Gettings et al. (1991, 1994, 1996), *in vivo* comparative data were obtained from a modified Draize eye test. For the test method protocol, six rabbits (three male, three female) were used for each test substance. The right eye of each rabbit was anesthetized prior to instillation of 0.1 mL of test substance into the conjunctival sac. Ocular irritation was evaluated at 1 hour, and at 1, 2, 3, 4 and 7 days. If irritation persisted, ocular responses were observed at seven day intervals up to a maximum of 21 days. MAS were determined according to Williams et al. (1982). Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances were provided by the CTFA. These substances were classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems and as described in **Section 4.3** (**Appendix C**).

Existing and concurrently run *in vivo* studies, all of which were performed according to OECD TG 405 and following GLP guidelines were used to assess *in vitro* test method accuracy in Balls et al. (1995). The data were generated after 1981 and met the following criteria:

- Normally used at least three New Zealand White rabbits tested at the same time.
- 0.1 mL or the equivalent weight of substance was instilled into the conjunctival sac.
- Anesthesia was not used.
- Observations were made at least at 1, 2, and 3 days after instillation.

Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each substance were obtained from the ECETOC Reference Chemicals data bank (ECETOC 1998). Using these data, substances were classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems and as described in **Section 4.3** (**Appendix C**).

---

[2] Some of the evaluated substances did not have corresponding *in vivo* rabbit data included in the chemical profile. Only substances where *in vivo* rabbit results were obtained were used in this evaluation.

[3] According to details in the report, the criteria used in the EEC (1984) classification system appear to be identical to the criteria used in the EU (2001) classification system.

In vivo data used by Gilleron et al. (1996) data were obtained from the studies of Gautheron et al. (1994). According to the report, the studies were performed according to the French and European directives (EEC 1984, 1991). Substances were classified by the authors according to the EU (1993) classification system and used to assess the *in vitro* test method accuracy.

For Spielmann et al. (1996), *in vivo* comparative data were obtained following the Draize eye test. The studies were performed in accordance with OECD TG 405. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances were provided by ZEBET. Using this data, substances were classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems and as described in **Section 4.3** (**Appendix C**).

For Hagino et al. (1999), *in vivo* comparative data were collected using the conventional Draize eye test protocol. Ocular irritation was evaluated at 1 hour and each day after treatment, up to a maximum of 14 days. Detailed *in vivo* data for each test substance, consisting of cornea, iris, and conjunctiva scores for each rabbit were provided by the National Institute of Health Sciences (NIHS). These substances were classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems and as described in **Section 4.3** (**Appendix C**).

For Bagley et al. (1992), Vinardell and Macián (1994), Kojima et al. (1995), and Gilleron et al. (1997) comparative *in vivo* data were obtained from ECETOC Reference Chemical data bank (ECETOC 1998) and Hagino et al. (1999). The substances each of these studies were classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems and as described in **Section 4.3** (**Appendix C**).

4.3 *In Vivo* Classification Criteria Used for BRD Analysis

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

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

- At least three rabbits were tested in the study, unless a severe effect (e.g., corrosion of the cornea) was noted in a single rabbit. In such cases, substance

classification could proceed based on the effects observed in less than three rabbits.

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

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

4.3.1 GHS Classification Rules Used for BRD Analysis

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

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

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

4.3.2 EPA Classification Rules Used for BRD Analysis

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

Table 4-3 Criteria for Classification of Rabbits According to the GHS Classification System (Modified from UN 2003)

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

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

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

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

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

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

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

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

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

4.3.3 EU Classification Rules Used for BRD Analysis

Substance classification using the EU classification system was conducted sequentially (EU 2001). While average Draize scores are used for classification, the calculation of average scores for the EU system depends on the number of rabbits tested in a study (see **Section 4.1.3** for additional details). Depending on the number of rabbits tested, the appropriate average scores were calculated, then the substance was classified based on the number of rabbits with a minimal positive average (for studies that used three rabbits) or the overall average (for studies that used more than three rabbits). The criteria used for substance classification are in **Table 4-6**.

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

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

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

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

NICEATM staff made attempts to obtain original HET-CAM and *in vivo* reference data for substances. A *Federal Register* (*FR*) notice (Vol. 69. No. 57, pp. 13589-12861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original HET-CAM (and comparative *in vivo* rabbit) data was published on March 24, 2004. A second request for original HET-CAM (and comparative *in vivo* rabbit) was published on February 28, 2005 (Vol. 69, No. 38, pp. 9661-9662). In addition, NICEATM staff contacted authors of selected published HET-CAM studies to request the original HET-CAM and *in vivo* reference data. In response to these efforts, the following *in vivo* data were obtained:

- *In vivo* data for the substances evaluated by Spielmann et al. (1996) were obtained from Drs. H. Spielmann and M. Liebsch of ZEBET. Individual animal scores for each substance tested were provided.
- *In vivo* data for the substances evaluated by Hagino et al. (1999) were obtained from Dr. Yasuo Ohno of NIHS. Individual animal scores for each substance tested , in addition to other substances, were provided.
- Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances were provided by the CTFA for Gettings et al. (1991, 1994, 1996).

4.5 *In Vivo* Data Quality

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

The extent to which the *in vivo* rabbit eye studies used to provide the comparative data in the published HET-CAM validation studies were compliant with GLP guidelines is based on the information provided in the published reports. Although an attempt was made to obtain the original study records, such records could not be obtained. Based on the available information, the reports that were identified as following GLP guidelines or used data obtained according to GLP guidelines were Gettings et al. (1991, 1994, 1996), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999).

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

Due to the possibility of irreversible eye injury that could impair vision or cause blindness, human ocular irritancy studies are not routinely conducted. The only exceptions are for products intended for actual human eye use (e.g., contact lens solutions, ophthalmic pharmaceuticals) or cosmetic/personal care products that are known not to cause more than minimal to mild responses in rabbits. Bruner et al. (1998) and Cater et al. (2004) reported on studies conducted in humans of cosmetic and surfactant-based personal care formulations.

However, all of the substances tested were classified as mild irritants or nonirritants and corresponding HET-CAM tests were not conducted. Procter & Gamble provided information from human exposures to three consumer-product formulations as a comparison to the EU ocular toxicity classifications (EU 2001), assigned based on results from the low volume eye test (LVET). However, because all three of these formulations were classified as nonirritants or mild irritants, based on results obtained in the LVET, evaluation of the accuracy of the HET-CAM test method for identifying ocular corrosives and severe irritants in humans is not possible.

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

Several U.S. Federal agencies (OSHA, CPSC, and the National Institute for Occupational Safety and Health [NIOSH]) were contacted for data resulting from accidental human exposures. Based on emergency department reports for work related eye-injuries, NIOSH estimated that approximately 39,200 chemical-related eye injuries occurred in 1998, (NIOSH 2004). Approximately 10,000 of these cases were attributed to an unidentified or unspecified chemical. Additional cases (<2500 each) were reported for injuries related to specific chemicals or chemical/product classes, which included[4]:

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

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

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

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

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

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

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

Accuracy of the *in vivo* test method would ideally be assessed by comparison of ocular effects observed in the rabbit to those effects produced in humans. A review of the literature indicates that there are few studies in which rabbit and human responses have been carefully compared under controlled conditions to assess the accuracy of the *in vivo* test method. Therefore, most studies conduct retrospective evaluations and comparisons of responses between humans and rabbits. A review indicates that a number of studies show that responses to mild to moderate irritants were generally similar between rabbits and humans (Lewin and Guillery 1913; Suker 1913; Leopold 1945; Carpenter and Smyth 1946; McLaughlin 1946; Nakano 1958; Barkman 1969; Grant 1974). A review of these studies can be found in McDonald et al. (1987). For a severe irritant, Grant (1974) and Butscher (1953) showed that accidental exposure to neat thioglycolic acid produced similar responses in humans and rabbits.

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

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

Based largely on the protocol of Draize et al. (1944), the original regulatory requirements for eye irritation testing mandated the use of at least six rabbits. In recognition of animal welfare concerns, several evaluations were conducted to assess the reliability of the test method and

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

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

[This Page Intentionally Left Blank]

5.0 HET-CAM TEST METHOD DATA AND RESULTS

5.1 Description of the HET-CAM Test Method Protocols Used To Generate Data

As noted in **Section 3.1**, 12 published reports contained sufficient data on which to conduct an analysis of HET-CAM test method accuracy. These reports are: CEC (1991), Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Gilleron et al. (1996, 1997), Spielmann et al. (1996), and Hagino et al. (1999).

The HET-CAM protocols used by these investigators are similar to each other, with a few exceptions (see **Appendix B1** for a comparative summary of test method components). Fertilized hen's eggs were incubated using conditions established by the investigator. A portion of the eggshell was removed and the CAM exposed. Then, the test substance was applied to the CAM surface. After a predetermined exposure period, the test substance was rinsed from the CAM. Irritant effects in the CAM blood vessels and albumen were subjectively assessed and either the times to the development of irritant endpoints were determined or the severity of the irritant endpoints was scored at predetermined time intervals.

Examples of some of the test method components that differed among the HET-CAM protocols used to generate data used in the accuracy analysis of **Section 6.0** include:

- Relative humidity during egg incubation ranged from 52.5 to 62.5%.
- Volume or quantity of the test substance applied to the CAM (when reported) was either 0.1 or 0.3 mL for liquids and 0.3 g for solids.
- Number of replicate eggs per test substance ranged from three to six.
- Some studies included concurrent positive control substances, while others did not.

The extent to which the differences among the various protocols impact on HET-CAM study results and the classification of a test substance as an ocular corrosive or severe irritant is unknown.

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

NICEATM staff made attempts to obtain original HET-CAM data for substances that also had been tested *in vivo* using the standard rabbit eye test. An *FR* notice (Vol. 69. No. 57, pp. 13589-12861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original HET-CAM (and comparative *in vivo* rabbit) data was published on March 24, 2004. A second request for original HET-CAM (and comparative *in vivo* rabbit) was published on February 28, 2005 (Vol. 69, No. 38, pp. 9661-9662). In addition, NICEATM staff contacted authors of selected published HET-CAM studies to request the original HET-CAM data. In response to these efforts, the following *in vitro* data were obtained:

- Summaries of HET-CAM results (e.g., Q-Scores) were obtained for the 60 substances evaluated by Balls et al. (1995) from European Centre for the

Validation of Alternative Methods (ECVAM). The summary data included the substance name and the average HET-CAM score for the substance.

- *In vitro* data for the substances evaluated in Spielmann et al. (1996) were obtained from Drs. H. Spielmann and M. Liebsch. The data provided included the overall HET-CAM scores obtained by each laboratory for each substance evaluated. *In vitro* data for two control substances also were provided.
- Drs. Philippe Vanparys and Freddy Van Goethem provided individual endpoint scores for each egg evaluated for substances described in Gilleron et al. (1996, 1997). *In vitro* data for four control substances also were provided.

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

As described in **Section 2.0**, the approach used to analyze HET-CAM study data varied and depended on the method used to collect the data. For test method protocols that evaluated the time to development of endpoints (i.e., hemorrhage, lysis, coagulation) which are correlated with ocular corrosivity or irritation (**Section 2.2.9.1**), an IS, Q-Score, or mtc value was calculated. For test method protocols that evaluated the severity of the toxic response (**Section 2.2.9.3**), an S-Score was calculated. For test method protocols that evaluated the lowest test substance concentration needed to produce a minimal response on the CAM (**Section 2.2.9.2**), the ITC was determined. The ITC was typically combined with the IS for the test substance to evaluate ocular irritation or corrosivity potential of a substance.

The focus of the accuracy analysis in this BRD is on the ability of the HET-CAM test method to identify ocular corrosives or severe irritants, as defined by the GHS, EPA, and EU classification systems (EPA 1996; EU 2001; UN 2003). However, because of variations between *in vitro* analysis methods and the historical HET-CAM classification systems developed there were some retrospective evaluations that needed to be conducted. For example, no single irritancy classification scheme for distinguishing between nonirritants and various classes of irritants has been applied to *in vitro* HET-CAM data. Depending on the type of *in vitro* data collected and the method used to analyze the data, various irritation classification schemes have been developed. Even when HET-CAM data were evaluated using a common approach (e.g., IS), investigators used different decision criteria for classifying test substances as nonirritants or irritants.

Furthermore, most of the irritancy classification schemes used by the *in vitro* studies were not developed to meet the needs of the ocular irritation classification schemes currently used by the U.S. (EPA 1996), the EU (EU 2001), or the GHS (UN 2003). Therefore, substances classified based on *in vitro* data were usually defined as "severe irritant" or "mild irritant." These substances were not typically classified, based on *in vitro* data, according to the categories of the GHS (UN 2003), EPA (1996), or EU (2001) classification systems (e.g., Category 1 for the GHS classification system, Category I for the EPA classification system, or R41 for the EU classification system). It is noted that there have been attempts by some investigators (Gettings et al. 1991, 1994, and 1996; Spielmann et al. 1996) to correlate HET-CAM scores with the ocular irritation classification scheme described by the FHSA

classification system (CPSC 1988) and by the EU classification system (EU 1992), respectively (see **Section 2.2.13**).

To evaluate the ability of HET-CAM to identify ocular corrosives and severe irritants, as defined by the EPA (1996), GHS (UN 2003), and EU (2001) classification systems, HET-CAM results obtained using each of the four different analysis methods were assigned an ocular irritancy classification based on the *in vitro* classification system most commonly used for that particular data analysis method. Thus, substances were classified in categories, based on the *in vitro* score, ranging from nonirritant to severe irritant. EU classifications were assigned, based on the *in vitro* results, for the substances tested in Spielmann et al. (1996). These investigator assigned classifications then were used in evaluating the ability of HET-CAM to identify ocular corrosives and severe irritants as defined by the EU classification system (EU 1992).

For some of the studies evaluated, the HET-CAM results for different testing laboratories were available (Balls et al. 1995; Spielmann et al. 1996; Hagino et al. 1999). In these cases, an overall "consensus classification call" was made for each multiply tested substance. The result of each testing laboratory (e.g., IS value) was converted to the corresponding irritation classification. The classification obtained by a majority of the testing laboratories was used to develop a "consensus classification call." In those cases where the same number of testing laboratories had different results, the more severe result was used for the overall classification call (e.g., if two testing laboratories classified a substance as a moderate irritant and two testing laboratories classified the same substance as a severe irritant; the overall classification call was severe irritant).

Some investigators (e.g., Gettings et al. 1996) classified the ocular irritancy potential of test substances using two or more different analysis methods. In such cases, these data were reclassified according to the approach used most commonly for each *in vitro* classification scheme and an accuracy assessment was conducted for each analysis method.

5.3.1 IS

5.3.1.1 *IS Analysis Method*

For those test method protocols that assigned a score to each of the endpoints evaluated at preset time intervals, the values assigned to each endpoint were totaled to give an IS value for the test substance (i.e., IS[A] analysis method). The possible IS values range from 0 (for test substances that do not induce development of any of the toxic endpoints of interest over the range of time intervals) to 21 (for test substances that induced development of all three toxic endpoints within 30 seconds of application of the test substance) (Luepke 1985).

For those test method protocols that noted the time that a specific endpoint was first observed, the IS value was calculated (i.e., IS[B] analysis method) using the formula (Kalweit et al. 1987, 1990):

$$\left(\left(\frac{(301 - \textit{Hemorrhage time})}{300}\right) \times 5\right) + \left(\left(\frac{(301 - \textit{Lysis time})}{300}\right) \times 7\right) + \left(\left(\frac{(301 - \textit{Coagulation time})}{300}\right) \times 9\right)$$

where:

Hemorrhage time = time (in seconds) of the first appearance of blood hemorrhages
Lysis time = time (in seconds) of the first appearance of vessel lysis
Coagulation time = time (in seconds) of the first appearance of protein coagulation

The IS value, when calculated using this formula, has a maximal value of 21.

When the development of hyperemia, injection, or another toxic endpoint was evaluated instead of vessel lysis, the time to first appearance for the alternative endpoint replaced the lysis time point.

5.3.1.2 *IS Classification Scheme*

For studies that used the analysis methods developed by Luepke (1985) or Kalweit et al. (1987, 1990), the ocular irritancy classification scheme described in **Table 5-1** was used for the accuracy analysis presented in this BRD (see **Section 6.0**). Therefore, substances with an IS(A) or IS(B) value of nine or greater were classified as severe irritants for the purposes of this analysis. The rationale for the decision criteria used in this classification scheme were not provided and the correlation of these categories to irritancy categories described by the EPA (1996), GHS (UN 2003), and EU (2001) classification systems is unknown.

Table 5-1 IS Classification Scheme Used to Classify Substances For Accuracy Analysis[1]

HET-CAM Score Range	Irritation Category
0 to 0.9	Nonirritant
1 to 4.9	Slight Irritation
5 to 8.9	Moderate Irritation
9 to 21	Severe Irritation

[1]According to Luepke (1985) and Kalweit et al. (1987, 1990).

5.3.2 Q-Score

5.3.2.1 *Q-Score Analysis Method*

To determine the Q-Score, the irritation potential of both the test substance and a reference substance are evaluated. The irritation potential could be determined using any approach, but typically was expressed as an IS value. The IS value of the test substance was then compared to the IS value of the reference standard to calculate a ratio, which was then used to assess the irritation potential of the test substance.

5.3.2.2 *Q-Score Classification Scheme*

The study that used Q-Scores to classify the ocular irritation potential of test substances used the classification scheme of Balls et al. (1995) (see **Table 5-2**). This classification scheme was used in the BRD; substances with a Q-Score of at least 2 were classified as a severe irritant. The rationale for the decision criteria used in this classification scheme were not provided and the correlation of these categories to irritancy categories described by the EPA (1996), GHS (UN 2003), and EU (2001) classification systems is unknown.

Table 5-2 Q-Score Classification Scheme Used to Classify Substances For Accuracy Analysis[1]

Q-Score	Irritation Category
< 1.5	Nonirritant
1.5 ≤ Q < 2	Moderate
≥ 2	Severe

[1]Classification scheme according to Balls et al. (1995)

5.3.3 mtc10

5.3.3.1 *mtc10 Analysis Method*

To determine the mtc10, the mean detection time for the appearance of the coagulation endpoint when using a 10% solution was evaluated. The mean is calculated over the total number of replicate eggs used for each experiment.

5.3.3.2 *mtc10 Classification Scheme*

Two different cut-off values were used to classify a substance as a severe irritant. Linear discriminant analysis was performed, assuming equal *a priori* probabilities. With this linear model, a value of 174 seconds (for 142 chemicals) and 139 seconds (for 189 chemicals) was calculated to separate the severe substances (i.e., R41 chemicals) from the nonsevere substances (**Table 5-3**). For the accuracy analyses described in **Section 6.0** of the BRD, two evaluations are provided. Substances with an mtc value less than 174 seconds and 139 seconds were classified as severe irritants (R41). The classification scheme was developed for the EU classification system (EU 1992).

Table 5-3 mtc10 Classification Scheme Used to Classify Substances For Accuracy Analysis[1]

mtc10 (Range 1)	mtc10 (Range 2)	Irritation Category
< 174 seconds	< 139 seconds	R41
≥ 174 seconds	≥ 139 seconds	Remainder

[1]From Spielmann et al. (1996).

5.3.4 IS and ITC

5.3.4.1 *IS and ITC Analysis Method*

This analysis method combines two different parameters to determine the irritancy potential of a test substance. The IS value is determined for each test substance at a 10% concentration and the ITC is defined as the lowest concentration producing a slight or weak response on the CAM after application of the test substance.

5.3.4.2 *IS and ITC Classification Scheme*

For the accuracy analysis, substances with (a) an ITC value less than 1%, or (b) an ITC value between 1% and 2.5% and an IS value of at least 16 were classified as severe irritants (R41) (**Table 5-4**).

Table 5-4 IS and ITC Classification Scheme Used to Classify Substances for Accuracy Analysis[1]

ITC (% Concentration)	IS Value (10% Concentration)	EU Irritation Category[2]
> 10%	< 16	None/slight (Nonirritant)
> 10%	> 16	Moderate (Nonirritant)
< 10%	< 16	Moderate (Nonirritant)
< 10%	> 16	Irritant (R36)
< 2.5%	< 16	Irritant (R36)
1% < ITC ≤ 2.5%	≥ 16	Severe (R41)
≤ 1%		Severe (R41)

[1]According to Spielmann et al. (1996)
[2]EU (1992)

5.3.5 S-Score

5.3.5.1 *S-Score Analysis Method*

This score represents the highest total score for any endpoint evaluated for a test substance. The severity scores assigned for each endpoint (which range from 0 to 3 and are assigned at a single user-defined time point after treatment) are totaled across the replicate eggs evaluated per test substance to produce a total score for each irritation endpoint (i.e., three total scores). The toxic endpoint that yields the highest score is the S-Score for the test substance. Many of the test method protocols that evaluated the irritation potential of test substances using this method of analysis advocated the use of six eggs per test substance. In such situations, the maximal S-Score is 18.

5.3.5.2 *S-Score Classification Scheme*

Substances with an S-Score of at least 15 were classified as a severe irritant for the analysis described in the BRD (see **Table 5-5**). The rationale for the decision criteria used in this classification scheme were not provided and the correlation of these categories to irritancy categories described by the EPA (1996), GHS (UN 2003), and EU (2001) classification systems is unknown.

Table 5-5 S-Score Classification Scheme Used to Classify Substances For Accuracy Analysis[1]

S-Score	Irritation Category
< 6	Nonirritant
6 ≤ S < 15	Moderate
≥ 15	Severe

[1]Classification scheme according to Balls et al. (1995); based on six replicate eggs per test substance.

5.4 Summary of Results

A total of 260 test substances were evaluated in 383 HET-CAM studies. A summary of results used to evaluate test method accuracy is shown in **Appendix C.** This table, sorted by reference, provides the CASRN, the concentration tested, the calculated *in vitro* score, the *in vitro* irritation classification of the test substance (based on the irritation classification schemes in **Section 5.3**), and the literature source. Other supporting information, such as purity of the test substance, was included in the table to the extent that this information was available.

5.4.1 CEC (1991)

In vitro data for 15 substances evaluated in 26 studies were extracted. The substances were evaluated in up to seven laboratories. IS(B) values, calculated using the mathematical model developed by Kalweit et al. (1987), were presented in the report. Each tested substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. EU irritancy classifications, based on *in vivo* studies and results, were available for these substances. Therefore, accuracy of the *in vitro* results could only be compared to the EU classification system.

5.4.2 Gettings et al. (1991)

In the CTFA Evaluation of Alternatives Program – Phase I, ten hydroalcoholic formulations were evaluated in one laboratory. Mean IS(B) values, calculated using the mathematical model developed by Kalweit et al. (1987), were presented in the report for nine of the formulations. Each formulation was classified based on the *in vitro* classification system described in **Section 5.3.1**. Comparative *in vivo* data for these formulations were obtained from the FDA and the CTFA, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

The report also described an *in vitro* analysis approach developed by Bartnik et al. (1987). The Bartnik et al. approach was not used in the accuracy analysis conducted in this BRD since the quantitative aspects of this model were not available. However, the study authors' conclusions regarding the accuracy of the HET-CAM test method using the Bartnik et al. (1987) analysis method are addressed in **Section 9.0**.

5.4.3 Gettings et al. (1994)

In the CTFA Evaluation of Alternatives Program – Phase II, 18 oil/water formulations were evaluated in one laboratory. Mean IS(A) and IS(B) values, calculated using the mathematical models developed by Leupke (1985) and Kalweit et al. (1987), respectively, were presented. Each formulation was classified based on the *in vitro* classification system described in **Section 5.3.1**. Comparative *in vivo* data for these formulations were obtained from the CTFA, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

5.4.4 Gettings et al. (1996)

In the CTFA Evaluation of Alternatives Program – Phase III, 25 surfactant-based personal care cleansing formulations were evaluated in one laboratory. Mean IS(A) and IS(B) values,

calculated using the mathematical models developed by Leupke (1985) and Kalweit et al. (1987), respectively, were presented. Each formulation was classified based on the *in vitro* classification system described in **Section 5.3.1**. Comparative *in vivo* data for these formulations were obtained from the CTFA, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

The report also described an *in vitro* analysis approach where the IS value was divided by the ITC to yield a ratio that was used to describe the irritation potential of the test substance. This approach was not used in the accuracy analysis conducted in this BRD since the quantitative aspects of this *in vitro* model were not available. However, the study authors' conclusions regarding the accuracy of the HET-CAM test method using this analysis approach are addressed in **Section 9.0**.

5.4.5 Bagley et al. (1992)

In vitro data for two substances were extracted from this report. The mean IS(A) values, calculated using the mathematical model described by Luepke (1985), were provided in the study report. Each substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. Comparative *in vivo* data for this evaluation were obtained from published literature, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

Although HET-CAM data also were reported for seven additional chemicals and 20 consumer product formulations in this publication, detailed *in vivo* reference data were not available for these substances. Therefore, the HET-CAM data for these substances are not included in this analysis. The study authors' conclusions regarding the accuracy of the HET-CAM test method for these substances are addressed in **Section 9.0**.

5.4.6 Vinardell and Macián (1994)

In vitro data for two test chemicals were extracted from this report. The mean IS(A) values, calculated using the mathematical model described by Luepke (1985), were provided in the study report. Each substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. Comparative *in vivo* data for this evaluation were obtained from published literature, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

Although HET-CAM data also were reported in this publication for two additional chemicals and six consumer product formulations, detailed *in vivo* reference data were not available for these substances. Therefore, the HET-CAM data for these substances are not included in this analysis. The study authors' conclusions regarding the accuracy of the HET-CAM test method for these substances are addressed in **Section 9.0**.

5.4.7 Balls et al. (1995)

In this evaluation of the HET-CAM test method, 52 test substances were evaluated in two to four laboratories. Four of these substances were tested at two different concentrations and two were tested at three concentrations, for a total of 60 different tests. The Q-Score and the S-Score were obtained for each substance in each laboratory. Tested substances were

classified based on the *in vitro* classification system described in **Section 5.3.2** and **Section 5.3.3**. Detailed *in vivo* data for the 60 studies were obtained from ECETOC (1998), allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

5.4.8 Kojima et al. (1995)

In vitro data were extracted for five test substances. The mean IS(A) values, calculated using the mathematical model described by Luepke (1985), were provided in the study report. Each substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. Comparative *in vivo* data for this evaluation were obtained from published literature, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

Although HET-CAM data also were reported for 18 other substances, detailed *in vivo* reference data were not available for these substances, precluding their use in an analysis of accuracy. However, the study authors' conclusions regarding the accuracy of the HET-CAM test method for these substances are addressed in **Section 9.0**.

5.4.9 Gilleron et al. (1996)

In this evaluation of the HET-CAM test method, 46 substances were evaluated in a single laboratory. Average HET-CAM IS(B) values, calculated using the mathematical model described by Kalweit et al. (1987), were provided in the report and individual endpoint scores for each egg evaluated for a substance were obtained. Each substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. EU irritancy classifications, based on *in vivo* studies and results, were available for these substances. Therefore, accuracy of the *in vitro* results could only be compared to the EU classification system.

5.4.10 Spielmann et al. (1996)

In the publication, two different analysis methods were presented. As mentioned in **Section 3.3.1**, the IS and ITC and the mtc10 evaluation were evaluated. Each substance was classified based on the appropriate *in vitro* classification system provided in Spielmann et al. (1996). EU irritancy classification, based on *in vivo* studies and results, were provided in the paper. The results of the accuracy analyses for these analysis methods are presented in **Section 6.0**.

In addition to the above two analysis methods, an additional analysis was conducted using the data available from this report and obtained from the study authors. For this additional analysis, IS(B) values for 112 substances that were evaluated in two to three laboratories were classified based on the *in vitro* classification system described in **Section 5.3.1**. The IS(B) values were calculated using the mathematical model described by Kalweit et al. (1987). Detailed *in vivo* data for the test substances were provided, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

5.4.11 Gilleron et al. (1997)

In this evaluation of the HET-CAM test method, 52 substances were evaluated and compared to 60 different *in vivo* studies. Average HET-CAM IS(B) values, calculated using the mathematical model described by Kalweit et al. (1987), were provided in the report and individual endpoint scores for each egg evaluated for a substance were obtained. Each substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. Detailed *in vivo* data for the substances were obtained from ECETOC (1998), allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

5.4.12 Hagino et al. (1999)

In this evaluation of the HET-CAM test method, 14 substances were evaluated in five laboratories. Three of these substances were tested at two different concentrations, for a total of 17 different tests. Average HET-CAM IS(A) values, calculated using the mathematical model described in Luepke (1985), from each testing laboratory and the overall average HET-CAM IS(A) from all the testing laboratories were provided in the report. Each substance was classified based on the *in vitro* classification system described in **Section 5.3.1**. Detailed *in vivo* data for the test substances (including the different concentrations tested) were provided, allowing for an accuracy assessment compared to the EPA (1996), GHS (UN 2003), and EU (2001) classification systems.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, coded chemicals should be used in all validation studies and all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). Data quality was evaluated by a review of the methods section in literature references and the submitted reports. Thus, data quality can be evaluated only to the extent this information was provided in the published reports. Based on the available information, the reports that were identified as following GLP guidelines or used data obtained according to GLP guidelines were Gettings et al. (1991, 1994, 1996), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999). Furthermore, based on the available information, the reports that identified using coded chemicals were Gettings et al. (1991, 1994, 1996), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999). Detailed information on coding procedures used in different studies is provided in **Section 3.4**.

5.6 Lot-to-lot Consistency of Test Substances

Ideally, a single lot of each substance would be used during the validation of a test method. In situations where multiple lots of the same chemical must be used, then lot-to-lot consistency of test substances needs to be evaluated to ensure that the same substance is being evaluated over the duration of the study. The procedures used in evaluating lot-to-lot consistency were evaluated by what was described in the published reports. No attempt was made to review original records to assess the procedures used to evaluate different batches of tested substances.

Gettings et al. (1991, 1994, 1996) noted that all substances were dispensed from a single source to ensure test substance consistency. The substances were placed in a secondary container, labeled with appropriate chemical code information, and then provided to the participating testing laboratories. No information was provided in the report about the time frame in which the studies were conducted, or whether more than one lot of a substance was tested.

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

No information was provided in any of the remaining reports about maintaining lot-to-lot consistency.

5.7 Availability of Data for External Audit

Availability of original study data, for the reports considered in the accuracy and reliability analysis, for an external audit has not been determined.

[This Page Intentionally Left Blank]

6.0 HET-CAM TEST METHOD ACCURACY

6.1 Accuracy of the HET-CAM Test Method

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

- accuracy (concordance): the proportion of correct outcomes (positive and negative) of a test method
- sensitivity: the proportion of all positive substances that are classified as positive
- specificity: the proportion of all negative substances that are classified as negative
- positive predictivity: the proportion of correct positive responses among substances testing positive
- negative predictivity: the proportion of correct negative responses among substances testing negative
- false positive rate: the proportion of all negative substances that are falsely identified as positive
- false negative rate: the proportion of all positive substances that are falsely identified as negative

The ability of the HET-CAM test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems[1], was evaluated using two approaches. In the first approach, the performance of HET-CAM was assessed separately for each *in vitro-in vivo* comparative study (i.e., "per study" approach) reviewed in **Sections 4.0** and **5.0**. Within the "per study" analysis approach, there were two different analyses used. In the second approach, the performance of HET-CAM was assessed after pooling data across comparative studies that used the same data analysis method (i.e., IS, IS and ITC, Q-Score, or S-Score).

As mentioned above, for the "per study" accuracy analysis approach, two different types of analyses were used. In the first analysis, the HET-CAM ocular irritancy potential of each substance in each report was determined (**Appendix C**). When the same substance was evaluated in multiple laboratories within the same study (see Balls et al. 1995, Spielmann et al. 1996, and Hagino et al. 1999 in **Appendix C**), the HET-CAM ocular irritancy potential for each independent test result was determined. An overall HET-CAM ocular irritancy classification was the assigned for each substance in the study based on the majority of ocular irritancy classification calls (e.g., if two laboratories classified a substance as a nonirritant and three laboratories classified a substance as a severe irritant; the overall *in vitro* irritancy classification for the substance used in this analysis would be severe irritant). When there was an even number of different irritancy classifications for substances (e.g., two laboratories classified a substance as

---

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

a nonirritant and two laboratories classified a substance as a severe irritant), the more severe irritancy classification was used for the overall classification for the substance (severe irritant, in this case). Once the ocular irritancy potential classification was determined for each substance in each of the studies, the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems, was determined for each study. The overall *in vitro* and *in vivo* classifications assigned to each substance are provided in **Appendix D**.

In the second analysis used in the "per study" evaluation, each classification obtained when the same substance was evaluated in multiple laboratories was used separately to assess test method accuracy (i.e., results were not combined across multiple laboratories to develop an overall HET-CAM ocular irritancy classification). The ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the three different classification systems, was then determined for reports where multiple results were available for tested substances. This analysis was applied to the CEC (1991), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999) studies.

In the second approach in evaluating the accuracy of HET-CAM, results from the different studies using the same HET-CAM analysis approach were combined. As discussed in **Section 2.0**, there are several different data analysis methods that have been used (i.e., IS, IS and ITC, Q-Score, S-Score). Therefore, an accuracy assessment was conducted for each analysis method described. When the same substance was evaluated in multiple laboratories, the overall HET-CAM ocular irritancy classification was based on the majority of calls among all of the laboratories in the studies (see **Appendix C**). Once the ocular irritancy classification was determined for each substance, the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems, was determined for each analysis method (**Appendix D**). Since the test methods protocols used in different studies to generate HET-CAM test results are not identical (see **Appendix A** for comparisons of key components of test method protocols), care should be used when interpreting the results of these analyses.

The three ocular hazard classification systems (GHS [UN 2003], EPA [1996], and EU [2001]) considered during each approach use different classification systems and decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results (see **Sections 1.0** and **4.0**). All three classification systems are based on individual animal response data in terms of the magnitude of the response and on the extent to which induced ocular lesions fail to reverse by day 21. Thus, to evaluate the accuracy of the HET-CAM test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at the different observation times are needed for each substance. However, these data were not consistently available in the reports considered, which limited the number of test results that could be used to assess test method accuracy. Furthermore, most of the *in vivo* classifications used for the analyses presented in this section are based on the results of a single study. Unless otherwise indicated, variability in the *in vivo* classification is unknown.

6.1.1 GHS Classification System: HET-CAM Test Method Accuracy

6.1.1.1 *Overall Test Method Accuracy*

Accuracy analyses for ocular corrosives and severe irritants, as defined by the GHS classification system (UN 2003)[2], were evaluated for the following reports: Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Spielmann et al. (1996), Gilleron et al. (1997), and Hagino et al. (1999). Of these reports, Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999) provided HET-CAM test data for substances tested in multiple laboratories.

In these studies, HET-CAM test data was provided for a total of 376 substances, 260 of which had sufficient comparative *in vivo* data that could be used to assign an ocular irritancy classification according to the GHS classification system (UN 2003). Of these 260 substances, 92 substances were classified as GHS severe irritants based on results from the *in vivo* rabbit eye test. *In vivo* and *in vitro* irritancy classifications of test substances are provided in **Appendix C** and **Appendix D**.

For one set of data (Spielmann et al. 1996), a large number of substances were available to compare the accuracy of the test method when substances were evaluated at a 10% and 100% concentration *in vitro* and 100% *in vivo*. Therefore, a comparison of the accuracy statistics of these two *in vitro* concentrations was possible. To include the additional HET-CAM test data, which were tested at 10% and 100% concentrations, appropriate data were combined with each of the Spielmann et al. (1996) data sets. These combined data sets were used to evaluate the accuracy of the IS(B) test method, when using a 10% (IS[B]-10) or 100% (IS[B]-100) concentration *in vitro*, to predict the effects produced *in vivo* at 100%. As a corollary to this evaluation, the accuracy of the IS(A) method, when substances were tested at 10% or 100% concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration also was evaluated.

Based on the data provided in the reports and when results across multiply tested substances were combined to generate a single consensus call per test substance, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the GHS classification system (UN 2003) of 41% to 83%, a sensitivity of 25% to 100%, a specificity of 9% to 100%, a false positive rate of 0% to 91%, and a false negative rate of 0% to 75%[3,4]. The performance characteristics for each report are provided in **Table 6–1**.

The performance statistic ranges for Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999), when results from different testing laboratories are considered separately rather than

---

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

[3] The ranges provided do not include the results obtained for Bagley et al. (1992) and Vinardell and Macián (1994); the number of chemicals evaluated (two each) was deemed to few to consider.

[4] For substances where there were two *in vivo* studies with discordant results (e.g., one study classified the substance as a Category 1 and a second study classified the substance as a Category 2A), the more severe irritancy classification was used for the accuracy analysis.

combined, are: 47% to 80% for accuracy, 27% to 87% for sensitivity, 46% to 82% for specificity, 18% to 54% for the false positive rate, 15% to 73% for the false negative rate. These performance characteristics also are provided in **Table 6-1**.

The overall performance statistics, arranged by HET-CAM data analysis method, are provided in **Table 6-2**. Based on the combined test result approach, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the GHS classification system (UN 2003), of 44% to 85%, a sensitivity of 25% to 100%, a specificity of 40% to 100%, a false positive rate of 0% to 60%, and a false negative rate of 0% to 75%

The IS(A)-100 analysis method (substances were tested *in vitro* at a concentration of 100% and compared to substances tested *in vivo* at 100%) had the highest accuracy for predicting ocular corrosives and severe irritants (85% [17/20]). It is noted that for the IS(A)-100 analysis method evaluation represents 20 substances that are mostly formulations. Comparatively, the IS(B) approach (which has a larger database and contains many individual chemicals) had the highest accuracy when 10% concentration tested *in vitro* was compared to 100% concentration tested *in vivo*. The false positive and false negative rates for this analysis method were 33% (20/61) and 30% (12/40), respectively.

6.1.1.2 *Discordant Results According to the GHS Classification System*

To evaluate discordant responses of the HET-CAM test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed for each analysis method evaluated. These included specific classes of chemicals with sufficiently robust numbers of substances (n ≥ 5), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants and surfactant-based formulations, pH, physical form).

IS(A)-10 and IS(A)-100 Analysis Methods

The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration, and (b) substances tested at a 100% concentration. As is shown in **Table 6-3**, the false negative rate of the IS(A) analysis method is higher when test substances are tested at a 10% concentration (75% [12/16]) when compared to 100% (0% [0/2]). However, the false positive rate of the IS(A) analysis method is lower for the 10% concentration (0% [0/8]) compared to the 100% concentration (17% [3/18]).

As indicated in **Table 6-3**, formulations were the only chemical class with a sufficient number of substances that allowed for an evaluation. Twelve out of sixteen (75%) of formulations (all of which were surfactant-based formulations and all of which were tested as liquids *in vivo*) were underpredicted by the IS(A)-10 analysis method. Comparatively, 18% (3/17) formulations (oil water formulations) evaluated by the IS(A)-100 analysis method were overpredicted. With regard to physical form for the IS(A)-100 analysis method, the false positive and false negative rates were 17% (3/18) and 0% (0/2), respectively for liquids.

Substances were more likely to be underpredicted if (a) the *in vivo* effect was based on a persistent lesion and (b) if the concentration of the test substance *in vitro* was 100% (**Table 6-3**).

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

Data Source	Anal.[1]	n[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gettings et al. (1991)	IS(B)	9/10	78	7/9	100	3/3	67	4/6	60	3/5	100	4/4	33	2/6	0	0/3
Gettings et al. (1994)	IS(A)	18/18	83	15/18	100	1/1	82	14/17	25	1/4	100	14/14	18	3/17	0	0/1
Gettings et al. (1994)	IS(B)	18/18	78	14/18	100	1/1	76	13/17	20	1/5	100	13/13	24	4/17	0	0/1
Gettings et al. (1996)	IS(A)	24/25	50	12/24	25	4/16	100	8/8	100	4/4	40	8/20	0	0/8	75	12/16
Gettings et al. (1996)	IS(B)	24/25	71	17/24	56	9/16	100	8/8	100	9/9	53	8/15	0	0/8	44	7/16
Bagley et al. (1992)	IS(A)	2/32	0	0/2	-	-	0	0/2	0	0/2	-	-	100	2/2	-	-
Vinardell and Macián (1994)	IS(B)	2/13	50	1/2	-	-	50	1/2	0	0/1	100	1/1	50	1/2	-	-
Balls et al. (1995)	Q	43/59	63	27/43	100	15/15	43	12/28	48	15/31	100	12/12	57	16/28	0	0/15
	Q*	162/177	62	101/162	87	45/52	51	56/110	45	45/99	88	56/63	49	54/110	13	7/52
Balls et al. (1995)	S	16/59	44	7/16	36	4/11	60	3/5	67	4/6	30	3/10	40	2/5	64	7/11
	S*	47/54	47	22/47	27	8/30	82	14/17	73	8/11	38	14/36	18	3/17	73	22/30
Kojima et al. (1995)	IS(A)	5/24	60	3/5	50	2/4	100	1/1	100	2/2	33	1/3	0	0/1	50	2/4
Spielmann et al. (1996)	IS(B)-10	77/120	68	52/77	79	19/24	62	33/53	49	19/39	87	33/38	38	20/53	21	5/24
	IS(B)-10*	157/236	75	118/157	74	37/50	75	81/107	59	37/63	86	81/94	24	26/107	26	13/50
Spielmann et al. (1996)	IS(B)-100	75/120	55	41/75	88	21/24	39	20/51	40	21/52	87	20/23	61	31/51	13	3/24
	IS(B)-100*	150/236	58	87/150	85	40/47	46	47/103	42	40/96	87	47/54	54	56/103	15	7/47
Gilleron et al. (1997)	IS(B)	54/60	41	22/54	86	19/22	9	3/32	40	19/48	50	3/6	91	29/32	14	3/22
Hagino et al. (1999)	IS(A)	15/17	80	12/15	100	8/8	57	4/7	73	8/11	100	4/4	43	3/7	0	0/8
	IS(A)*	75/85	67	50/75	90	36/40	40	14/35	63	36/57	78	14/18	60	21/35	10	4/40

Abbreviation: GHS = United Nations Globally Harmonized System (UN 2003).

[1]Anal. = Data collection/analysis method used to transform the sample data into HET-CAM scores. IS(A) = Method described in Luepke (1985); IS(B), IS(B)-10, and IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995). For those analysis methods marked with an "*", *in vitro* results across multiple testing laboratories were not pooled to develop an overall HET-CAM classification for the test substance. In these analyses, the accuracy evaluation was based on individual study results for substances evaluated in multiple laboratories. Additional information on this approach is provided in **Section 6.1**
[2]n = Number of substances included in this analysis/the total number of substances in the study.
[3]The data on which the percentage calculation is based.

Table 6-2 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by HET-CAM Analysis Method

Analysis Method[1]	n[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
IS(A)-10[4]	24	50	12/24	25	4/16	100	8/8	100	4/4	40	8/20	0	0/8	75	12/16
IS(A)-100[4]	20	85	17/20	100	2/2	83	15/18	40	2/5	100	15/15	17	3/18	0	0/2
IS(A)	60	63	38/60	48	13/27	76	25/33	62	13/21	64	25/39	24	8/33	52	14/27
IS(B)-10[4]	101	68	69/101	70	28/40	67	41/61	58	28/48	77	41/53	33	20/61	30	12/40
IS(B)-100[4]	138	54	75/138	87	34/39	41	41/99	37	34/92	89	41/46	59	58/99	13	5/39
IS(B)[5]	106	58	61/106	79	33/42	44	28/64	48	33/69	76	28/37	56	36/64	21	9/42
Q-Score	43	63	27/43	100	15/15	43	12/28	48	15/31	100	12/12	57	16/28	0	0/15
S-Score	16	44	7/16	36	4/11	60	3/5	67	4/6	30	3/10	40	2/5	64	7/11

Abbreviation: GHS = Globally Harmonized System (UN 2003)
[1]IS(A), IS(A)-10, IS(A)-100 = Method described in Luepke (1985); IS(B), IS(B)-10, IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995).
[2]n = Number of substances evaluated in each study.
[3]The data on which the percentage calculation is based.
[4]The analysis compares the ability of the specified concentration tested *in vitro* (IS(A)-10 represents the 10% concentration tested *in vitro*) to predict the effect produced by the undiluted test substance tested *in vivo*.
[5]This analysis excludes substances evaluated in Spielmann et al. (1996).

Table 6-3 False Positive and False Negative Rates of the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall IS(A)-10	24	0	0/8	75	12/16
Overall IS(A)-100	20	17	3/18	0	0/2
Chemical Class[3]-IS(A)-10					
Formulation	24	0	0/8	75	12/16
Chemical Class[3]-IS(A)-100					
Formulation	18	18	3/17	0	0/1
Properties of Interest					
Physical Form: IS(A)-10					
Liquids	24	0	0/8	75	12/16
Physical Form: IS(A)-100					
Liquids	20	17	3/18	0	0/2
Solids	0	-	-	-	-
Surfactant-Based Formulations – IS(A)-10	24	0	0/8	75	12/16
Category 1 Subgroup-IS(A)-10[4]					
- Total	2	-	-	0	0/2
- 4 (CO=4 at any time)	1	-	-	0	0/1
- 3 (severity/persistence)	0	-	-	-	-
- 2 (severity)	0	-	-	-	-
- 2-4 combined[5]	1	-	-	0	0/1
- 1 (persistence)	1	-	-	0	0/1
Category 1 Subgroup-IS(A)-100[4]					
- Total	16	-	-	75	12/16
- 4 (CO=4 at any time)	0	-	-	-	-
- 3 (severity/persistence)	0	-	-	-	-
- 2 (severity)	0	-	-	-	-
- 2-4 combined[5]	0	-	-	-	-
- 1 (persistence)	16	-	-	75	12/16

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).

[1]n = number of substances

[2]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*; False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

[3] Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

[4] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent

[5] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

IS(B)-10 and IS(B)-100 Analysis Methods
Due to the various concentrations of test substances evaluated in this test method, different permutations of these sub-analyses are provided for comparative purposes. The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration in the entire database and (b) substances tested at a 100% concentration in the entire database. As is shown in **Table 6-4**, the false negative rate of the IS(B) analysis method is higher when test substances are tested at a 10% concentration (30%, 12/40) when compared to 100% (13%, 5/39). However, the false positive rate of the IS(B) analysis method is lower for the 10% concentration (33%, 20/61) compared to the 100% concentration (59%, 58/99).

As indicated in **Table 6-4**, there were some trends in the performance of the HET-CAM test method among subgroups of the tested substances. The chemical class of substances that was consistently overpredicted (i.e., were false positives) by both analysis methods is alcohols. Eight out of a total of 16 (89% [8/9]) and 14 out of a total of 24 alcohols (88% [14/16]) were overpredicted by the IS(B)-10 and IS(B)-100 analysis methods, respectively. Additional chemical classes that were overpredicted by both analysis methods were ethers, amines, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% [0/8] and 26% [6/23]). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines and ethers. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% (3/16) and 37% (7/19), respectively for liquids and 58% (11/19) and 13% (1/8) for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 65% (33/51) and 0% (0/9), respectively for liquids and 67% (16/24) and 24% (4/17) for solids.

Information regarding the pH of test substances was only available for a subset of the substances tested at a concentration of 10% or 100% using the IS(B) analysis method. Among the substances that were tested at a 10% concentration, two out of a total of 35 test substances were underpredicted (false negative rate: 13%; 2/16). Among these two, both were acidic (pH < 7.0). For substances tested at a 100% concentration, two out of 35 test substances were underpredicted. Of these substances, one was acidic (pH < 7.0) and one was basic (pH > 7.0). For substances that were overpredicted, basic substances were more overpredicted than acidic substances when tested at a 10% concentration *in vitro* (false positive rate of basic substances = 80% [4/5] vs. false positive rate of acidic substances: 50% [7/14]) (see **Table 6-4**).

Finally, substances were more likely to be underpredicted if (a) the *in vivo* effect was based on a persistent lesion, and (b) if the concentration of the test substance *in vitro* was 10% (**Table 6-4**).

Table 6-4 False Positive and False Negative Rates of the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall IS(B)-10	101	33	20/61	30	12/40
Overall IS(B)-100	138	59	58/99	13	5/39
Chemical Class[3]-IS(B)-10					
Alcohols	16	89	8/9	25	2/7
Aldehyde	5	0	0/4	100	1/1
Amine	7	60	3/5	50	1/2
Ether	14	50	5/10	50	2/4
Formulation	24	0	0/8	44	7/16
Heterocyclic compound	7	86	6/7	-	0/0
Organic salt	7	57	4/7	-	0/0
Chemical Class[3]-IS(B)-100					
Alcohols	24	88	14/16	13	1/8
Aldehyde	6	80	4/5	0	0/1
Amine	9	83	5/6	33	1/3
Carboxylic acid/Carboxylic acid salt	11	60	3/5	17	1/6
Ester	12	90	9/10	0	0/2
Ether	16	50	6/12	25	1/4
Formulation	27	26	6/23	0	0/4
Heterocyclic compound	12	78	7/9	33	1/3
Inorganic salt	5	100	2/2	0	0/3
Ketone	6	67	4/6	-	0/0
Organic salt	9	86	6/7	0	0/2
Properties of Interest					
Physical Form: IS(B)-10					
Liquids/Solutions	35	19	3/16	37	7/19
Solids	27	58	11/19	13	1/8
Unknown	39	23	6/26	31	4/13
Physical Form: IS(B)-100					
Liquids	60	65	33/51	0	0/9
Solids	41	67	16/24	24	4/17
Unknown	37	38	9/24	8	1/13
Surfactant – Total IS(B)-100	2	50	1/2	-	0/0
-nonionic	2	50	1/2	-	0/0
-anionic	0	-	-	-	-
-cationic	0	-	-	-	-
Surfactant-Based Formulations – IS(B)-10	24	0	0/8	44	7/16
pH – IS(B)-10	35[4]	58	11/19	13	2/16
- acidic (pH < 7.0)	24	50	7/14	20	2/10
- basic (pH > 7.0)	11	80	4/5	0	0/6

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
pH – IS(B)-100	35[4]	68	13/19	13	2/16
- acidic (pH < 7.0)	23	69	9/13	10	1/10
- basic (pH > 7.0)	12	67	4/6	17	1/6
Category 1 Subgroup-IS(B)-10[5]					
- Total	40	-	-	30	12/40
- 4 (CO=4 at any time)	13	-	-	15	2/13
- 3 (severity/persistence)	0	-	-	-	-
- 2 (severity)	0	-	-	-	-
- 2-4 combined[6]	13	-	-	15	2/13
- 1 (persistence)	27	-	-	37	10/27
Category 1 Subgroup-IS(B)-100[5]					
- Total	38[7]	-	-	11	4/38
- 4 (CO=4 at any time)	19	-	-	11	2/19
- 3 (severity/persistence)	1	-	-	100	1/1
- 2 (severity)	2	-	-	0	0/2
- 2-4 combined[6]	22	-	-	14	3/22
- 1 (persistence)	16	-	-	6	1/16

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as negative *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3] Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] Total number of GHS Category 1 substances for which pH information was obtained.
[5] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent
[6] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.
[7]The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

Q-Score Analysis Method

As is shown in **Table 6-5**, the false positive and negative rates of the Q-Score analysis method are 57% (16/28) and 0% (0/15), respectively.

As indicated in **Table 6-5**, the chemical classes that were overpredicted by the Q-Score analysis method were alcohols and esters. The false negative rate was 0% for all chemical classes shown in the table.

With regard to physical form of the substances, the false positive and false negative rates were 59% (16/27) and 0% (0/11), respectively for liquids and 0% (0/1) and 0% (0/4) for solids.

Table 6-5 False Positive and False Negative Rates of the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall Q-Score	43	57	16/28	0	0/15
Chemical Class[3]					
Alcohols	10	50	4/8	0	0/2
Carboxylic Acid/Carboxylic Acid Salt	5	100	2/2	0	0/3
Ester	7	43	3/7	-	0/0
Heterocyclic compound	7	50	1/2	0	0/5
Onium	7	0	0/2	0	0/5
Properties of Interest					
Physical Form:					
Liquids	38	59	16/27	0	0/11
Solids	5	0	0/1	0	0/4
Category 1 Subgroup[4]					
- Total	14[6]	-	-	0	0/14
- 4 (CO=4 at any time)	7	-	-	0	0/7
- 3 (severity/persistence)	0	-	-	-	0/0
- 2 (severity)	3	-	-	0	0/3
- 2-4 combined[5]	10	-	-	0	0/10
- 1 (persistence)	4	-	-	0	0/4

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent
[5] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.
[6] The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

Since there was an overall 0% (0/15) false negative rate, there was a 0% false negative rate when the accuracy of the analysis method was evaluated when compared to different *in vivo* lesion types.

S-Score Analysis Method

As is shown in **Table 6-6**, the false positive and false negative rates of the S-Score analysis method are 40% (2/5) and 64% (7/11).

Table 6-6 False Positive and False Negative Rates of the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall S-Score	16	40	2/5	64	7/11
Chemical Class[3]					
Carboxylic Acid/Carboxylic Acid Salt	5	0	0/1	75	3/4
Properties of Interest					
Physical Form:					
Liquids	0	-	-	-	-
Solids	16	40	2/5	64	7/11
Category 1 Subgroup[4]					
- Total	10[6]	-	-	60	6/10
- 4 (CO=4 at any time)	5	-	-	80	4/5
- 3 (severity/persistence)	1	-	-	100	1/1
- 2 (severity)	1	-	-	100	1/1
- 2-4 combined[5]	7	-	-	86	6/7
- 1 (persistence)	3	-	-	0	0/3

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).
[1]n = number of substances
[2]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*; False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[3] Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent
[5] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.
[6] The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

The only chemical class with sufficient substances to conduct an analysis was carboxylic acids/carboxylic acid salts. In this chemical class, the false negative rate was 75% (3/4) while the false positive rate was 0% (0/1). With regard to physical form of the substances, all substances tested using this analysis method were solids; thus, the false negative rate was 64% (7/11). Finally, substances were more likely to be underpredicted if (a) the *in vivo* effect was based on a corneal opacity of 4 at any time.

6.1.2 EPA Classification System: HET-CAM Test Method Accuracy

6.1.2.1 *Overall Test Method Accuracy*

Accuracy analyses for ocular corrosives and severe irritants, as defined by the EPA classification system (EPA 1996), were evaluated for the following reports: Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Spielmann et al. (1996), Gilleron et al. (1997), and Hagino et al. (1999). Of these reports, Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999) provided HET-CAM test data for substances tested in multiple laboratories.

In these studies, HET-CAM test data was provided for a total of 376 substances, 256 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the EPA classification system (EPA 1996)[5]. Based on results from the *in vivo* rabbit eye test, 76 of these 256 substances were classified as severe irritants (i.e., Category I). *In vivo* and *in vitro* irritancy classifications of test substances are provided in **Appendix C** and **Appendix D**.

As described in the previous section (see **Section 6.1.1**), a large number of substances were available to compare the accuracy of the test method when substances were evaluated at a 10% and 100% concentration *in vitro* and 100% *in vivo*. As conducted previously, to include the additional HET-CAM test data, which were tested at 10% and 100% concentrations, appropriate data were combined with each of the Spielmann et al. data sets. These combined data sets were used to evaluate the accuracy of the IS(B) test method, when using a 10% (IS(B)-10) or 100% (IS(B)-100) concentration *in vitro*, to predict the effects produced *in vivo* at 100%. As a corollary to this evaluation, the accuracy of the IS(A) method, when substances were tested at 10% or 100% concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration was evaluated.

Based on the data provided in the ten reports and when results across multiply tested substances were combined to generate a single consensus call per test substance, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EPA classification system (EPA 1996), of 38% to 83%, a sensitivity of 24% to 100%, a specificity of 12% to 100%, a false positive rate of 0% to 88%, and a false negative rate of 0% to 76%[6]. The performance characteristics for each report are provided in **Table 6-7**.

The performance statistic ranges for Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999), when results from different testing laboratories are considered separately rather than combined, are: of 53% to 72% for accuracy, 32% to 94% for sensitivity, 35% to 83% for specificity, 17% to 65% for the false positive rate, and 6% to 68% for the false negative rate. These performance characteristics are provided in **Table 6-7**.

The overall performance statistics, arranged by HET-CAM data analysis method, are provided in **Table 6-8**. Based on the combined test result approach, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EPA classification system (EPA 1996), of 48% to 85%, a sensitivity of 24% to 100%, a specificity of 41% to 100%, a false positive rate of 0% to 59%, and a false negative rate of 0% to 76%.

---

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

[6] For substances where there were two *in vivo* studies with discordant results (e.g., one study classified the substance as a Category I and a second study classified the substance as a Category II), the more severe irritancy classification was used for the accuracy analysis.

Table 6-7 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by Study

Data Source	Anal.[1]	n[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gettings et al. (1991)	IS(B)	9/10	78	7/9	100	3/3	67	4/6	60	3/5	100	4/4	33	2/6	0	0/3
Gettings et al. (1994)	IS(A)	18/18	83	15/18	100	1/1	82	14/17	25	1/4	100	14/14	0	3/17	0	0/1
Gettings et al. (1994)	IS(B)	18/18	78	14/18	100	1/1	76	13/17	20	1/5	100	13/13	24	4/17	0	0/1
Gettings et al. (1996)	IS(A)	25/25	48	12/25	24	4/17	100	8/8	100	4/4	38	8/21	0	0/8	76	13/17
Gettings et al. (1996)	IS(B)	25/25	72	18/25	59	10/17	100	8/8	100	10/10	53	8/15	0	0/8	41	7/17
Bagley et al. (1992)	IS(A)	2/32	0	0/2	-	-	0	0/2	0	0/2	-	-	100	2/2	-	-
Vinardell and Macián (1994)	IS(B)	2/13	50	1/2	-	-	50	1/2	0	0/1	100	1/1	50	1/2	-	-
Balls et al. (1995)	Q	44/59	61	27/44	100	14/14	43	13/30	45	14/31	100	13/13	57	17/30	0	0/14
	Q*	163/177	63	103/163	92	44/48	51	59/115	44	44/100	94	59/63	49	56/115	8	4/48
Balls et al. (1995)	S	15/20	53	8/15	44	4/9	67	4/6	67	4/6	44	4/9	33	2/6	56	5/9
	S*	43/54	53	23/43	32	8/25	83	15/18	73	8/11	47	15/32	17	3/18	68	17/25
Kojima et al. (1995)	IS(A)	5/24	80	4/5	67	2/3	100	2/2	100	2/2	67	2/3	0	0/2	33	1/3
Spielmann et al. (1996)	IS(B)-10	74/120	64	47/74	80	12/15	59	35/59	33	12/36	92	35/38	41	24/59	20	3/15
	IS(B)-10*	148/236	72	107/148	72	21/29	72	86/119	39	21/54	91	86/94	28	33/119	28	8/29
Spielmann et al. (1996)	IS(B)-100	71/120	51	36/71	93	14/15	39	22/56	29	14/48	96	22/23	61	34/56	7	1/15
	IS(B)-100*	141/236	55	77/141	89	25/28	46	52/113	29	25/86	95	52/55	54	61/113	11	3/28
Gilleron et al. (1997)	IS(B)	53/60	38	20/53	84	16/19	12	4/34	35	16/46	57	4/7	88	30/34	16	3/19
Hagino et al. (1999)	IS(A)	15/17	73	11/15	100	7/7	50	4/8	64	7/11	100	4/4	50	4/8	0	0/7
	IS(A)*	75/135	63	47/75	94	33/35	35	14/40	56	33/59	88	14/16	65	26/40	6	2/35

Abbreviation: EPA = U.S. Environmental Protection Agency.

[1]Anal. = Data collection/analysis method used to transform the sample data into HET-CAM scores. IS(A) = Method described in Luepke (1985); IS(B), IS(B)-10, and IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995). For those analysis methods marked with an "*", *in vitro* results across multiple testing laboratories were not pooled to develop an overall HET-CAM classification for test substances. In these analyses, the accuracy evaluation was based on individual study results for substances evaluated in multiple laboratories. Additional information on this approach is provided in **Section 6.1**.
[2]n = Number of substances included in this analysis/the total number of substances in the study.
[3]The data on which the percentage calculation is based.

Table 6-8 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by HET-CAM Analysis Method

Analysis Method[1]	n[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
IS(A)-10[4]	25	48	12/25	24	4/17	100	8/8	100	4/4	38	8/21	0	0/8	76	13/17
IS(A)-100[4]	20	85	17/20	100	2/2	83	15/18	40	2/5	100	15/15	17	3/18	0	0/2
IS(A)	61	66	40/61	46	12/26	74	26/35	57	12/21	65	26/40	26	9/35	54	14/26
IS(B)-10[4]	98	65	64/98	68	21/31	64	43/67	47	21/45	81	43/53	36	24/67	32	10/31
IS(B)-100[4]	133	52	69/133	89	25/28	42	44/105	29	25/86	94	44/47	58	61/105	11	3/28
IS(B)[5]	106	57	60/106	78	31/40	49	29/66	46	31/68	76	29/38	56	37/66	22	9/40
Q-Score	44	61	27/44	100	14/14	43	13/30	45	14/31	100	13/13	57	17/30	0	0/14
S-Score	15	53	8/15	44	4/9	67	4/6	67	4/6	44	4/9	33	2/6	56	5/9

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]IS(A), IS(A)-10, IS(A)-100 = Method described in Luepke (1985); IS(B), IS(B)-10, IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995).
[2]n = Number of substances evaluated in each study.
[3]The data on which the percentage calculation is based.
[4]The analysis compares the ability of the specified concentration tested *in vitro* (IS(A)-10 represents the 10% concentration tested *in vitro*) to predict the effect produced by the undiluted test substance tested *in vivo*.
[5]This analysis excluded substances evaluated in Spielmann et al. (1996).

The IS(A)-100 analysis approach, when substances were tested *in vitro* at a concentration of 100% and compared to substances tested *in vivo* at 100%, had the highest accuracy for predicting ocular corrosives and severe irritants (85% [17/20]), as classified by the EPA (EPA 1996). It is noted that the database used for this evaluation represents 20 substances that are mostly formulations. Comparatively, the IS(B) approach (which has a larger database and contains many individual chemicals) had the highest accuracy when 10% concentration tested *in vitro* was compared to 100% concentration tested *in vivo*. The false positive and false negative rates for this analysis method were 36% (24/67) and 32% (10/31), respectively.

6.1.2.2 *Discordant Results According to the EPA Classification System*

To evaluate discordant responses of the HET-CAM test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed for each analysis method evaluated. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants and surfactant-based formulations, pH, physical form).

IS(A)-10 and IS(A)-100 Analysis Methods

The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration, and (b) substances tested at a 100% concentration. As is shown in **Table 6-9**, the false negative rate of the IS(A) analysis method is higher when test substances are tested at a 10% concentration (76% [13/17]) when compared to 100% (0% [0/2]). However, the false positive rate of the IS(A) analysis method is lower for the 10% concentration (0% [0/8]) compared to the 100% concentration (17% [3/18]).

As indicated in **Table 6-9**, formulations were the only chemical class with a sufficient number of substances that allowed for an evaluation. Thirteen out of seventeen (76%) of formulations (all of which were surfactant-based formulations and all of which were tested as liquids *in vivo*) were underpredicted by the IS(A)-10 analysis method. Comparatively, 18% (3/17) formulations (oil water formulations) evaluated by the IS(A)-100 analysis method were overpredicted. With regard to physical form for the IS(A)-100 analysis method, the false positive and false negative rates were 17% (3/18) and 0% (0/1), respectively for liquids.

IS(B)-10 and IS(B)-100 Analysis Methods

Due to the various concentrations of test substances evaluated in this test method, different permutations of these sub-analyses are provided for comparative purposes. The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration in the entire database and (b) substances tested at a 100% concentration in the entire database. As is shown in **Table 6-10**, the false negative rate of the IS(B) analysis method is higher when test substances are tested at a 10% concentration (32%, 10/31) when compared to 100% (11%, 3/28). However, the false positive rate of the IS(B) analysis method is lower for the 10% concentration (36%, 26/67) compared to the 100% concentration (58%, 61/105).

Table 6-9 False Positive and False Negative Rates of the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EPA Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall IS(A)-10	25	0	0/8	76	13/17
Overall IS(A)-100	20	17	3/18	0	0/2
Chemical Class[3]-IS(A)-10					
Formulation	25	0	0/8	76	13/17
Chemical Class[3]-IS(A)-100					
Formulation	18	18	3/17	0	0/1
Properties of Interest					
Physical Form: IS(A)-10 **Liquids**	25	0	0/8	76	13/17
Physical Form: IS(A)-100 **Liquids**	20	17	3/18	0	0/2
Solids	0	-	-	-	-
Surfactant-Based Formulations – IS(A)-10	25	0	0/8	76	13/17

Abbreviations: EPA = U.S. Environmental Protection Agency (EPA 1996).
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

As indicated in **Table 6-10**, there were some trends in the performance of the HET-CAM test method among subgroups of the tested substances. The chemical class of substances that was consistently overpredicted according the EPA classification system (i.e., were false positives) by both analysis methods is alcohols. Eight out of a total of 15 (89% [8/9]) and 14 out of a total of 23 alcohols (88% [14/16]) were overpredicted by the IS(B)-10 and IS(B)-100 analysis methods, respectively. Additional chemical classes that were overpredicted by both analysis methods were ethers, amines, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% [0/8]) and 26% [6/23]). Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% (3/16) and 37% (7/19), respectively for liquids and 65% (15/23) and 0% (0/1) for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 65% (33/51) and 0% (0/9), respectively for liquids and 66% (19/29) and 25% (2/8) for solids.

Table 6-10 False Positive and False Negative Rates of the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EPA Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall IS(B)-10	98	36	24/67	32	10/31
Overall IS(B)-100	133	58	61/105	11	3/28
Chemical Class[3]-IS(B)-10					
Alcohols	15	89	8/9	33	2/6
Aldehyde	6	80	4/5	0	0/1
Amine	5	60	3/5	-	0/0
Ether	11	50	5/10	0	0/1
Formulation	25	0	0/8	41	7/17
Heterocyclic compound	8	88	7/8	-	0/0
Organic salt	7	57	4/7	-	0/0
Chemical Class[3]-IS(B)-100					
Alcohols	23	88	14/16	14	1/7
Aldehyde	6	80	4/5	0	0/1
Amine	8	71	5/7	100	1/1
Carboxylic acid/Carboxylic acid salt	9	60	3/5	25	1/4
Ester	11	90	9/10	0	0/1
Ether	13	50	6/12	0	0/1
Formulation	27	26	6/23	0	0/4
Heterocyclic compound	12	78	7/9	33	1/3
Inorganic salt	5	100	4/4	0	0/1
Ketone	6	67	4/6	-	0/0
Organic salt	10	75	6/8	0	0/2
Properties of Interest					
Physical Form: IS(B)-10					
Liquids/Solutions	35	19	3/16	37	7/19
Solids	24	65	15/23	0	0/1
Unknown	39	21	6/28	27	3/11
Physical Form: IS(B)-100					
Liquids	60	65	33/51	0	0/9
Solids	37	66	19/29	25	2/8
Unknown	36	36	9/25	9	1/11
Surfactant – Total IS(B)-100	2	50	1/2	-	0/0
-nonionic	2	50	1/2	-	0/0
-anionic	0	-	-	-	-
-cationic	0	-	-	-	-
Surfactant-Based Formulations – IS(B)-10	25	0	0/8	41	7/17
pH – IS(B)-10	32[4]	58	14/24	0	0/8
- acidic (pH < 7.0)	19	53	8/15	0	0/4
- basic (pH > 7.0)	13	78	7/9	0	0/4

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
pH – IS(B)-100	30[4]	73	16/22	0	0/8
- acidic (pH < 7.0)	18	64	9/14	0	0/4
- basic (pH > 7.0)	12	88	7/8	0	0/4

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] Total number of EPA Category I substances for which pH information was obtained.

Information regarding the pH of test substances was available for a subset of the substances tested. Among the substances that were tested at a 10% concentration, zero out of a total of 32 test substances were underpredicted (false negative rate: 0% [0/8]). The false positive rate for substances (a) tested at 10% concentration and (b) pH information was available was 58% (14/24). When these substances were separated based on pH, the false positive rate for acidic substances was 53% (8/15) and for basic substances was 78% (7/9). For substances tested at a 100% concentration, the false negative rate for substances for which pH information was available was 0% (0/8). Basic test substances had a higher false positive rate than acidic substances when tested at a 100% concentration *in vitro* (false positive rate of basic substances = 88% [7/8] vs. false positive rate of acidic substances: 64% [9/14]) (see **Table 6-10**).

Q-Score Analysis Method
As is shown in **Table 6-11**, the false positive and negative rates of the Q-Score analysis method are 57% (17/30) and 0% (0/14), respectively.

As indicated in **Table 6-11**, there were some trends in the performance of the Q-Score analysis method among subgroups of the tested substances. The chemical classes that were overpredicted according the EPA classification system (i.e., were false positives) were alcohols, carboxylic acids/carboxylic acid salts, esters, and heterocyclic compounds. The false negative rate was 0% for all chemical classes shown in the table.

With regard to physical form of the substances overpredicted by the Q-Score analysis method, the false positive and false negative rates were 61% (17/28) and 0% (0/10), respectively for liquids, and 0% (0/1) and 0% (0/4) for solids.

S-Score Analysis Method
As is shown in **Table 6-12**, the false positive and false negative rates of the S-Score analysis method are 33% (2/6) and 56% (5/9).

There were insufficient substances in any single chemical class evaluated (n ≥ 5) to assess the ability of the S-Score analysis method to predict specific classes. With regard to physical form of the substances, most of the substances evaluated with this method were solids. The false positive rate and false negative rate of solids was 33% (2/6) and 56% (5/9), respectively (**Table 6-12**).

Table 6-11 False Positive and False Negative Rates of the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the EPA Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall Q-Score	44	57	17/30	0	0/14
Chemical Class[3]					
Alcohols	10	50	4/8	0	0/2
Carboxylic Acid/Carboxylic Acid Salt	6	100	2/2	0	0/4
Ester	7	43	3/7	-	0/0
Heterocyclic compound	6	50	1/2	0	0/4
Onium	6	0	0/2	0	0/4
Property of Interest					
Physical Form:					
Liquids/Solutions	38	61	17/28	0	0/10
Solids	6	0	0/2	0	0/4

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

Table 6-12 False Positive and False Negative Rates of the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the EPA Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall S-Score	15	33	2/6	56	5/9
Property of Interest					
Physical Form					
Liquids	0	-	-	-	-
Solids	15	33	2/6	56	5/9

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

6.1.3 EU Classification System: HET-CAM Test Method Accuracy

6.1.3.1 *Overall Test Method Accuracy*

Accuracy analyses for ocular corrosives and severe irritants, as defined by the EU classification system (EU 2001), were evaluated for the following reports: CEC (1991), Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Spielmann et al. (1996), Gilleron et al. (1996, 1997), and Hagino et al. (1999). Of these reports, CEC (1991), Balls et al. (1996), Spielmann et al. (1996), and Hagino et al. (1999) provided HET-CAM data for substances tested in multiple laboratories.

In these studies, HET-CAM test data was provided for a total of 381 substances, 312 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the EU

classification system (EU 2001)[7]. Based on results from the *in vivo* rabbit eye test, 85 of these 312 substances were classified as severe irritants (i.e., R41). *In vitro* and *in vivo* classifications of these substances are provided in **Appendix C** and **Appendix D**.

As described in **Section 6.1.1.1**, a large number of substances were available to compare the accuracy of the test method when substances were evaluated at a 10% and 100% concentration *in vitro* and 100% *in vivo*. As conducted previously, to include the additional HET-CAM test data, which were tested at 10% and 100% concentrations, appropriate data were combined with each of the Spielmann et al. data sets. These combined data sets were used to evaluate the overall accuracy of the IS(B) test method, when using a 10% (IS(B)-10) or 100% (IS(B)-100) concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration. As a corollary to this evaluation, the accuracy of the IS(A) method, when substances were tested at 10% or 100% concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration was evaluated.

In addition to the analysis methods described previously, two additional analysis methods were evaluated for performance when compared to the EU classification system. These methods, the IS and ITC and the mtc, were evaluated and the results presented in Spielmann et al. (1996). The results of the analysis discussed in the report were included in this section for comparison.

Based on the data provided in these reports and when results across multiply tested substances were combined to generate a single consensus call per test substance, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EU classification system (EU 2001), of 40% to 88%, a sensitivity of 25% to 100%, a specificity of 10% to 100%, a false positive rate of 0% to 90%, and a false negative rate of 0% to 75%[8]. The performance characteristics for each report are provided in **Table 6-13**.

The performance statistic ranges for CEC (1991), Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999), when results from different testing laboratories are considered separately rather than combined, are: of 55% to 76% for accuracy, 35% to 91% for sensitivity, 38% to 82% for specificity, 18% to 62% for the false positive rate, and 9% to 65% for the false negative rate. These performance characteristics are provided in **Table 6-13**.

The overall performance statistics, arranged by HET-CAM data analysis method, are provided in **Table 6-14**. Based on the combined test result approach, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EU classification system (EU 2001), of 50% to 85%, a sensitivity of 25% to 100%, a specificity of 46% to 100%, a false positive rate of 0% to 54%, and a false negative rate of 0% to 75%.

---

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

[8] For substances where there were two *in vivo* studies with discordant results (e.g., one study classified the substance as a Category I and a second study classified the substance as a Category II), the more severe irritancy classification was used for the accuracy analysis.

Table 6-13 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System, by Study

Data Source	Anal.[1]	n[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
CEC (1991)	IS(B)	26/32	62	16/26	86	6/7	53	10/19	40	6/15	91	10/11	47	9/19	14	1/7
	IS(B)*	130/130	68	88/130	67	22/33	68	66/97	42	22/53	86	66/77	32	31/97	33	11/33
Gettings et al. (1991)	IS(B)	8/10	88	7/8	100	3/3	80	4/5	75	3/4	100	4/4	20	1/5	0	0/3
Gettings et al. (1994)	IS(A)	18/18	83	15/18	100	1/1	82	14/17	25	1/4	100	14/14	18	13/17	0	0/1
Gettings et al. (1994)	IS(B)	18/18	78	14/18	100	1/1	76	13/17	20	1/5	100	13/13	24	4/17	0	0/1
Gettings et al. (1996)	IS(A)	24/25	50	12/24	25	4/16	100	8/8	100	4/4	40	8/20	0	0/8	75	12/16
Gettings et al. (1996)	IS(B)	24/25	71	17/24	56	9/16	100	8/8	100	9/9	53	8/15	0	0/8	44	7/16
Bagley et al. (1992)	IS(A)	2/32	0	0/2	-	-	0	0/2	0	0/2	-	-	100	2/2	-	-
Vinardell and Macián (1994)	IS(B)	2/13	50	1/2	-	-	50	1/2	0	0/1	100	1/1	50	1/2	-	-
Balls et al. (1995)	Q	39/49	64	25/39	100	13/13	46	12/26	48	13/27	100	12/12	54	14/26	0	0/13
	Q*	146/177	64	94/146	91	40/44	53	54/102	45	40/88	93	54/58	47	48/102	9	4/44
Balls et al. (1995)	S	14/59	50	7/14	44	4/9	60	3/5	67	4/6	38	3/8	40	2/5	56	5/9
	S*	40/54	55	22/40	35	8/23	82	14/17	73	8/11	48	14/29	18	3/17	65	15/23
Kojima et al. (1995)	IS(A)	4/24	75	3/4	67	2/3	100	1/1	100	2/2	50	1/2	0	0/1	33	1/3
Spielmann et al. (1996)[4]	IS and ITC	118/118	71	84/118	42	19/45	89	65/73	70	19/27	71	65/91	11	8/73	58	26/45
Spielmann et al. (1996)[4]	mtc10	142	76	108/142	52	25/48	88	83/94	70	25/36	78	83/106	12	11/94	48	23/48
Spielmann et al. (1996)[4]	mtc10	189	77	145/189	53	30/57	87	115/132	64	30/47	81	115/142	13	17/132	47	27/57
Spielmann et al. (1996)	IS(B)-10	71/120	66	47/71	82	14/17	61	33/54	40	14/35	92	33/36	39	21/54	18	3/17
	IS(B)-10*	144/236	76	109/144	77	27/35	75	82/109	50	27/50	91	82/90	25	27/109	23	8/35

Data Source	Anal.[1]	n[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Spielmann et al. (1996)	IS(B)-100	69/120	52	36/69	94	16/17	38	20/52	33	16/48	95	20/21	62	32/52	6	1/17
	IS(B)-100*	138/236	70	97/138	91	30/33	45	47/105	34	30/88	94	47/50	55	58/105	9	3/33
Gilleron et al. (1996)	IS(B)	46/46	57	26/46	67	2/3	56	24/43	10	2/21	96	24/25	44	19/43	33	1/3
Gilleron et al. (1997)	IS(B)	48/60	40	19/48	89	16/18	10	3/30	37	16/43	60	3/5	90	27/30	11	2/18
Hagino et al. (1999)	IS(A)	15/17	73	11/15	100	7/7	50	4/8	64	7/11	100	4/4	50	4/8	0	0/7
	IS(A)*	75/85	63	47/75	91	32/35	38	15/40	56	32/57	83	15/18	62	25/40	9	3/35

Abbreviation: EU = European Union.

[1]Anal. = Data collection/analysis method used to transform the sample data into HET-CAM scores. IS(A) = Method described in Luepke (1985); IS(B), IS(B)-10, and IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995). For those analysis methods marked with an "*", *in vitro* results across multiple testing laboratories were not pooled to develop an overall HET-CAM classification for the test substances. In these analyses, the accuracy evaluation was based on individual study results for substances evaluated in multiple laboratories. Additional information on this approach is provided in **Section 6.1**.

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

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

[4]Results were calculated based on the results presented in Spielmann et al. (1996). Classification of *in vivo* results is described in Spielmann et al. (1996).

Table 6-14 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System, by HET-CAM Analysis Method

Analysis Method[1]	N[2]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
IS(A)-10[4]	24	50	12/24	25	4/16	100	8/8	100	4/4	40	8/16	0	0/8	75	12/16
IS(A)-100[4]	20	85	17/20	100	2/2	83	15/18	40	2/5	100	15/15	17	3/18	0	0/2
IS(A)	59	66	39/59	48	12/25	74	25/34	57	12/21	66	25/38	26	9/34	52	13/25
IS(B)-10[4]	95	67	64/95	70	23/31	66	41/62	52	23/44	80	41/51	34	21/62	30	10/33
IS(B)-100[4]	164	57	94/164	93	31/33	48	63/131	31	31/99	97	63/65	52	68/131	6	2/33
IS(B)[5]	161	60	97/161	80	37/46	52	60/115	40	37/92	87	60/69	48	55/115	20	9/46
Q-Score	39	64	25/39	100	13/13	46	12/26	48	13/27	100	12/12	54	14/26	0	0/13
S-Score	14	50	7/14	44	4/9	60	3/5	67	4/6	38	3/8	40	2/5	56	5/9
mtc10[6]	142	76	108/142	52	25/48	88	83/94	70	25/36	78	83/106	12	11/94	48	23/48
mtc10[6]	189	77	145/189	53	30/57	87	115/132	64	30/47	81	115/142	13	17/132	47	27/57
IS and ITC[6]	118	71	84/118	42	19/45	89	65/73	70	19/27	71	65/91	11	8/73	58	26/45

Abbreviation: EU=European Union (EU [2001]).

[1]IS(A), IS(A)-10, IS(A)-100 = method described in Luepke (1985); IS(B), IS(B)-10, IS(B)-100 = method described in Kalweit et al. (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).

[2]N = number of substances evaluated in each study.

[3]Data used to calculate the percentage.

[4]The analysis compares the ability of the specified concentration tested *in vitro* (IS(A)-10 represents the 10% concentration tested *in vitro*) to predict the effect produced by the undiluted test substance tested *in vivo*.

[5]This analysis excluded substances evaluated in Spielmann et al. (1996).

[6]Results were calculated based on the results presented in Spielmann et al. (1996). Classification of *in vivo* results is described in Spielmann et al. (1996).

The IS(A)-100 analysis approach, when substances were tested *in vitro* at a concentration of 100% and compared to substances tested *in vivo* at 100%, had the highest accuracy for predicting ocular corrosives and severe irritants (85% [17/20]), as classified by the EU (EU 2001). It is noted that these results that the database used for this evaluation represents 20 substances that are mostly formulations. Comparatively, the IS(B) approach (which has a larger database and contains many individual chemicals) had the highest accuracy when 10% concentration tested *in vitro* was compared to 100% concentration tested *in vivo*. The false positive and false negative rates for this analysis method were 34% (21/62) and 30% (10/33), respectively.

6.1.3.2 *Discordant Results According to the EU Classification System*

To evaluate discordant responses of the HET-CAM test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed for each analysis method evaluated. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants and surfactant-based formulations, pH, physical form).

IS(A)-10 and IS(A)-100 Analysis Method

The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration, and (b) substances tested at a 100% concentration. As is shown in **Table 6-15**, the false negative rate of the IS(A) analysis method is higher when test substances are tested at a 10% concentration (75% [12/16]) when compared to 100% (0% [0/2]). However, the false positive rate of the IS(A) analysis method is lower for the 10% concentration (0% [0/8]) compared to the 100% concentration (17% [3/18]).

As indicated in **Table 6-15**, formulations were the only chemical class with a sufficient number of substances that allowed for an evaluation. Twelve out of sixteen (75%) of formulations (all of which were surfactant-based formulations and all of which were tested as liquids *in vivo*) were underpredicted by the IS(A)-10 analysis method. Comparatively, 18% (3/17) formulations (oil-water formulations) evaluated by the IS(A)-100 analysis method were overpredicted. With regard to physical form for the IS(A)-100 analysis method, the false positive and false negative rates were 17% (3/18) and 0% (0/2), respectively for liquids.

IS(B)-10 and IS(B)-100

Due to the various concentrations of test substances evaluated in this test method, different permutations of these sub-analyses are provided for comparative purposes. The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration in the entire database, and (b) substances tested at a 100% concentration in the entire database. As is shown in **Table 6-16**, the false negative rate of the IS(B) analysis method is higher when test substances are tested at a 10% concentration (30%, 10/33) when compared to 100% (6%, 2/33). However, the false positive rate of the IS(B) analysis method is lower for the 10% concentration (34%, 21/62) compared to the 100% concentration (52%, 68/131).

Table 6-15 False Positive and False Negative Rates of the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EU Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall IS(A)-10	24	0	0/8	75	12/16
Overall IS(A)-100	20	17	3/18	0	0/2
Chemical Class[3]-IS(A)-10					
Formulation	24	0	0/8	75	12/16
Chemical Class[3]-IS(A)-100					
Formulation	18	18	3/17	0	0/1
Properties of Interest					
Physical Form: IS(A)-10 **Liquids**	24	0	0/8	75	12/16
Physical Form: IS(A)-100 **Liquids** **Solids**	20 0	17 -	3/18 -	0 -	0/2 -
Surfactant-Based Formulations – IS(A)-10	24	0	0/8	75	12/16

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro;* False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

As indicated in **Table 6-16**, there were some trends in the performance of the HET-CAM test method among subgroups of the tested substances. The chemical class of substances that was consistently overpredicted according the GHS classification system (i.e., were false positives) by both analysis methods is alcohols. Nine out of a total of 15 (90% [9/10]) and 19 out of at total of 31 alcohols (79% [19/24]) were overpredicted by the IS(B)-10 and IS(B)-100 analysis methods, respectively. Additional chemical classes that were overpredicted by both analysis methods were ethers, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% [0/8] and 23% [5/22]). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% (3/16) and 38% (7/18), respectively for liquids and 60% (12/20) and 0% (0/3) for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 61% (40/66) and 0% (0/8), respectively for liquids and 48% (19/40) and 8% (1/13) for solids.

Table 6-16 False Positive and False Negative Rates of the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EU Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall IS(B)-10	95	34	21/62	30	10/33
Overall IS(B)-100	164	52	68/131	6	2/33
Chemical Class[3]-IS(B)-10					
Alcohols	15	90	9/10	40	2/5
Aldehyde	5	0	0/4	100	1/1
Amine	14	30	3/10	50	2/4
Ether	12	50	5/10	0	0/2
Formulation	24	0	0/8	44	7/16
Heterocyclic compound	7	86	6/7	-	0/0
Organic salt	7	57	4/7	-	0/0
Chemical Class[3]-IS(B)-100					
Alcohols	31	79	19/24	14	1/7
Aldehyde	6	80	4/5	0	0/1
Amine	15	64	7/11	25	1/4
Carboxylic acid/Carboxylic acid salt	15	50	5/10	0	0/5
Ester	12	90	9/10	0	0/2
Ether	17	47	7/15	0	0/2
Formulation	27	23	5/22	0	0/4
Heterocyclic compound	16	58	7/12	25	1/4
Ketone	10	70	7/10	-	0/0
Organic salt	12	80	8/10	0	0/2
Organic sulfur containing compound	7	50	2/4	0	0/3
Properties of Interest					
Physical Form: IS(B)-10					
Liquids/Solutions	34	19	3/16	38	7/18
Solids	23	60	12/20	0	0/3
Unknown	38	23	6/26	25	3/12
Physical Form: IS(B)-100					
Liquids	74	61	40/66	0	0/8
Solids	53	48	19/40	8	1/13
Unknown	37	36	9/25	83	1/12
Surfactant – Total IS(B)-100	10	44	4/9	0	0/1
-nonionic	6	33	2/6	-	0/0
-anionic	1	100	1/1	-	0/0
-cationic	2	100	1/1	0	0/1
-zwitterionic	1	0	0/1	-	0/0
Surfactant-Based Formulations – IS(B)-10	24	0	0/8	44	7/16
pH – IS(B)-10	30[4]	58	11/19	0	0/11
- acidic (pH < 7.0)	21	50	7/14	0	0/7

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
- basic (pH > 7.0)	9	80	4/5	0	0/4
pH – IS(B)-100	29[4]	72	13/18	0	0/11
- acidic (pH < 7.0)	20	69	9/13	0	0/7
- basic (pH > 7.0)	9	80	4/5	0	0/4

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro;* False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4]Total number of EU R41 substances for which pH information was obtained.

Information regarding the pH of test substances was available for a subset of the substances tested. Among the substances that were tested at a 10% concentration, 11 out of 30 test substances were overpredicted (false positive rate: 58% [11/19]). Among these, seven were acidic (pH < 7.0), and four were basic. For substances tested at a 100% concentration, 13 out of 29 test substances were overpredicted. Of these substances, nine were acidic (pH < 7.0), and four were basic (pH > 7.0). For substances that were underpredicted, there was a 0% false negative rate for both analysis methods (see **Table 6-16**).

Q-Score Analysis Method

As is shown in **Table 6-17**, the false positive and negative rates of the Q-Score analysis method are 56% (14/26) and 0% (0/13), respectively.

Table 6-17 False Positive and False Negative Rates of the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the EU Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall Q-Score	39	56	14/26	0	0/13
Chemical Class[3]					
Alcohols	9	43	3/7	0	0/2
Ester	7	43	3/7	-	0/0
Heterocyclic compound	7	50	1/2	0	0/5
Onium	7	0	0/2	0	0/5
Properties of Interest					
Physical Form:					
Liquids	34	56	14/25	0	0/9
Solids	5	0	0/1	0	0/4

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

There were some trends in the performance of the Q-Score analysis method among subgroups of the tested substances. The chemical class of substances that were consistently overpredicted was alcohols, ester, and heterocyclic compounds. The false negative rate was 0% for all chemical classes shown in the table.

With regard to physical form of the substances overpredicted by the Q-Score analysis method, the false positive and false negative rates were 56% (14/25) and 0% (0/9) for liquids and 0% (0/1) and 0% (0/4) for solids, respectively.

S-Score Analysis Method

As is shown in **Table 6-18**, the false positive and false negative rates of the S-Score analysis method are 40% (2/5) and 56% (5/9).

Table 6-18 False Positive and False Negative Rates of the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the EU Classification System

Category	n[1]	False Positive Rate[2]		False Negative Rate[2]	
		%	No.	%	No.
Overall S-Score	14	40	2/5	56	5/9
Properties of Interest					
Physical Form:					
Liquids	0	-	-	-	-
Solids	14	40	2/5	56	5/9

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro;* False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

There were insufficient substances in any single chemical class evaluated (n ≥ 5) to assess the ability of the S-Score analysis method to predict specific classes. With regard to physical form of the substances, all substances tested using this analysis method were solids. As shown in **Table 6-18**, the false positive and false negative rates for solids ranged from 40% (2/5) to 56% (5/9).

IS and ITC Analysis Method

Based on the information provided in Spielmann et al. (1996), there were eight substances that were identified as false positives. These substances were:

- (-)-phenylephrine
- theophylline sodium acetate
- (+)-phenylephrine
- sodium cyanate
- sodium lauryl ether sulfate
- hyton
- *p*-anisidine
- rubinrot Y

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

While there were some differences in results among the three hazard classification systems evaluated (i.e., EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that HET-CAM test method performance was comparable among the three systems (see **Table 6-19**).

Table 6-19 Ranges of Performance Statistics for Evaluated Analysis Methods for GHS, EPA, and EU Classification Systems

Analysis Methods	Accuracy	Sensitivity	Specificity	False Positive Rates	False Negative Rates
IS(A)-10	48-50%	24-25%	100%	0%	75-76%
IS(A)-100	85%	100%	83%	17%	0%
IS(B)-10	65-68%	68-70%	64-67%	33-36%	30-32%
IS(B)-100	51-57%	87-93%	40-47%	52-59%	6-13%
Q-Score	61-64%	100%	43-46%	54-57%	0%
S-Score	44-50%	36-44%	60-67%	33-40%	56-64%

Abbreviations: EPA = U.S. Environmental Protection Agency, EU = European Union, GHS = Globally Harmonized System.
Based on data presented in **Tables 6-2, 6-8,** and **6-14**. A single value indicates the same percentage results for all three hazard classification systems.

Given the relatively homogeneous performance of the HET-CAM test method among the three classification systems, the discussion for the IS(A)-10 and IS(A)-100, IS(B)-10 and IS(B)-100, Q-Score, and S-Score analysis methods encompasses all three hazard classification systems, unless otherwise indicated. Additional information on the mtc and/or the IS and ITC analysis method can be obtained from Spielmann et al. (1996) (**Section 9.0** provides a summary of the report).

6.2.1 Discordance Among Chemical Classes

6.2.1.1 *IS(A)-10 and IS(A)-100 Analysis Methods*

Most of the substances evaluated by these analysis methods were formulations. For the IS(A)-10 analysis method, which evaluated mostly surfactant-based formulations, the false negative rates ranged from 75% to 76%, while the false positive rate was 0% for all classification systems. Comparatively, the IS(A)-100 analysis method, which evaluated primarily oil-water formulations, had a higher false positive rate than false negative rate.

6.2.1.2 *IS(B)-10 and IS(B)-100 Analysis Methods*

The chemical class of substances that was consistently overpredicted according the GHS classification system (i.e., were false positives) by both analysis methods is alcohols (89% to 90% for the IS(B)-10 analysis method and 79% to 88% for the IS(B)-100 analysis method).

Additional chemical classes that were overpredicted by both analysis methods were ethers (50% for IS(B)-10 and 47 to 50% for IS(B)-100), organic salts (57% for IS(B)-10 and 75% to 86% for IS(B)-100), and heterocyclic compounds (86% to 88% for IS(B)-10 and 58 to 78% for IS(B)-100). Formulations appeared to have the lowest false positive rates for both analysis methods (0% for IS(B)-10 and 23% to 26% for IS(B)-100). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

The broad range in the accuracy results from some of the chemical classes evaluated (e.g., heterocyclic compounds evaluated with the IS(B)-100 analysis method) appears to be due to the greater number of substances within this chemical class that were evaluated by the EU classification system and not the GHS or EPA classification systems. As mentioned earlier in this section (see **Section 6.1**), insufficient *in vivo* data was available for some of the substances evaluated, which did not allow for classification according to all three classification systems.

6.2.1.3 *Q-Score Analysis Method*

The accuracy analysis indicated that alcohols and esters are often overpredicted (43 to 50% and 43%) false positive rate, depending on the classification system used) in the Q-score analysis method. The numbers of substances among the remaining chemical classes were too few to resolve any definitive trends in overprediction by the Q-Score analysis method. The false negative rate for all chemical classes with a sufficient number of substances ($n \geq 5$) was 0%.

6.2.1.4 *S-Score Analysis Method*

Due to the limited database for this analysis method, a chemical class evaluation could only be conducted for carboxylic acids/carboxylic acid salts for the GHS classification system. For this chemical class and classification system, the false negative rate was 75% (3/4) and the false positive rate was 0% (0/1).

6.2.2 Discordance Among Physical or Chemical Properties of Interest

6.2.2.1 *IS(A)-10 and IS(A)-100 Analysis Methods*

With regard to physical form of the substances tested by these analysis methods, a majority of the substances were tested as liquids/solutions *in vitro* and *in vivo*. Therefore, the false negative and false positive rates for these analysis methods were similar or the same as to the overall false positive and false negative rates. That is the false negative and false positive rates for liquids were 75% to 76% and 0% for the IS(A)-10 analysis method and 0% and 18% for the IS(A)-100 analysis method. No solids were evaluated using the IS(A)-10 analysis method, while the false negative and false positive rates were 0% for the IS(A)-100 analysis method.

For the GHS classification scheme, the evaluation indicated that substances were more likely to be underpredicted if (a) the *in vivo* lesion was based on persistence of effect and (b) if the *in vitro* test concentration was 100%.

6.2.2.2 *IS(B)-10 and IS(B)-100 Analysis Methods*

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% and 37% to 38%, respectively for liquids and 58% to 65% and 0% to 13% for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 61% to 65% and 0%, respectively for liquids and 48% to 67% and 8% to 24% for solids. The physical form of many of the tested substances was unknown based on the available information. Therefore, there were numerous tested compounds (36 to 39 substances) for each hazard classification system that were not included in this evaluation.

The broad range in the accuracy results from some of the physical properties (e.g. IS(B)-100 solids) evaluated appears to be due to the greater number of substances within this class that were evaluated by the EU classification system and not the GHS or EPA classification systems. As mentioned earlier in this section (see **Section 6.1**), insufficient *in vivo* data was available for some of the substances evaluated, which did not allow for classification according to all three classification systems.

Information regarding the pH of test substances was available for a subset of the substances tested (29 to 35 substances). Overall, substances were observed to have a higher false positive rate when (a) tested at a 100% concentration and (b) had a pH greater than 7.0.

For the GHS classification scheme, the evaluation indicated that substances were more likely to be underpredicted if (a) the *in vivo* lesion was based on persistence of effect and (b) if the *in vitro* test concentration was 10%.

6.2.2.3 *Q-Score Analysis Method*

With regard to physical form of the substances overpredicted by the Q-Score analysis method, 14 to 17 were liquids and none were solids. The ranges of false positive and false negative rates for liquids were 56% to 61% and 0%, respectively. The false positive and false negative rates for solids were 0% for both parameters. There was insufficient information for the other evaluated categories (e.g., surfactant-based formulations) to conduct an analysis.

6.2.2.4 *S-Score Analysis Method*

With regard to physical form of the substances overpredicted by the S-Score analysis method, 14 to 16 were solids. There were no liquids evaluated with analysis method. The false negative rates for solids ranged from 56% to 64% (5/9 to 7/11) and the false positive rates ranged from 33% to 40% (2/6 to 2/5). There was insufficient information for the other evaluated categories (e.g., surfactant-based formulations) to conduct an analysis.

7.0 HET-CAM TEST METHOD RELIABILITY

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

Quantitative HET-CAM test method data were available for replicate eggs within individual experiments or for replicate experiments within an individual laboratory for two studies (Gilleron et al. 1996, 1997). Therefore, an evaluation of the repeatability and/or intralaboratory reproducibility of the HET-CAM test method could be conducted. Additionally, comparable HET-CAM data were available for multiple laboratories within each of three to four comparative validation studies (CEC 1991, Balls et al. 1995, Spielmann et al. 1996, and Hagino et al. 1999), which allowed for an evaluation of the interlaboratory reproducibility of the HET-CAM test method.

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the HET-CAM Test Method

There was limited information on the rationale for substance selection used in various multilaboratory studies to evaluate the reliability of the HET-CAM test method. Most reports indicated that substances were selected for inclusion based on available *in vivo* rabbit eye data for comparison, to cover the range of ocular irritation potential, and to include substances with different physicochemical properties (e.g., solids, liquids).

The selection of substances used in the CEC (1991) evaluation was based on the following criteria:

- The substances should be representative of currently used industrial chemicals and should represent a range of chemical structures.
- The substances should cover the range of eye effects from nonirritant to severe irritant.

- The *in vivo* rabbit eye studies should have been conducted in accordance with EEC criteria, and the animal data should be sufficient to allow an irritancy classification to be definitively assigned to the test substance.
- Whenever possible, the substances should have been used in previous validation studies.

As noted previously, the EC/HO validation study reported on by Balls et al. (1995) evaluated the performance and reproducibility of the HET-CAM test method using 60 substances (i.e., there were 52 different substances with four substances tested at two different concentrations and two substances tested at three concentrations). A description of the requirements for inclusion into the study was provided in **Section 3.0**.

Gilleron et al. (1996, 1997) selected substances that represented a broad spectrum of ocular irritancies, chemical classes, and chemical structures. Substances also were selected on the basis of availability of historical *in vivo* data, to avoid conducting additional tests for the validation study. Additionally, substances evaluated in the Gilleron et al. (1997) study were the same as those previously evaluated by Balls et al. (1995).

Spielmann et al. (1996) selected substances that represented a broad spectrum of ocular irritancies, chemical classes, and chemical structures. Substances also were selected on the basis of availability of historical *in vivo* data.

Hagino et al. (1999) evaluated substances that were major ingredients in cosmetic formulations and preparations. These substances included surfactants and solvents.

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Quantitative Assessment of Intralaboratory Repeatability

An analysis of interlaboratory repeatability has included such approaches as:

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

Two of the reports discussed included intralaboratory repeatability data (Gilleron et al. 1996, 1997)[1]. Using these data, the consistency of HET-CAM IS(B) results obtained among identically-treated eggs within an experiment was evaluated using a CV analysis. Considering the number of replicate eggs tested in each experiment, no attempt was made to use ANOVA to determine if any individual egg score differed from any other egg scores.

7.2.1.1 Gilleron et al. (1996)

Individual egg results for 46 substances analyzed by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1996) were received in response to a request from NICEATM. In the data provided to NICEATM, the original test results for nine of the 46 substances included in the 1996 publication (laurylsulfobetaine, deoxycholic acid,

[1] Transformed data for these studies are available in **Appendices E1** and **E2**.

ethylacetoacetate, methyl isobutyl ketone, methanol, N-laurylsarcosine, promethazine hydrochloride, 2-methoxyethanol, benzethonium chloride, and imidazole) were no longer available. Since alternative HET-CAM test data generated were available for these substances, these data were provided to NICEATM. The overall replicate egg mean and median %CV values were evaluated with and without the inclusion of the data for these nine substances.

For each test substance, three different eggs were used in each of at least three replicate experiments. For this evaluation, the %CV values were determined for each endpoint (hemorrhage, lysis, coagulation) and for the overall *in vitro* IS(B) score. For each of the endpoints, there were experiments where test substances did not produce any effects (i.e., the average score of the three replicate eggs and standard deviation [SD] of the scores were both 0) (see **Appendix E3**). For the three endpoints evaluated, 69 of 146 experiments (47%) resulted in an average score and SD of zero for the hemorrhage and lysis endpoints. Additionally, 47 of 146 experiments (32%) resulted in a total average score and SD of zero for the coagulation endpoint. For the overall *in vitro* IS(B) score, 21 of 146 experiments (14%) resulted in an average score and SD of zero (**Appendix E3**). For three test substances (anthracene, ethylenediaminetetraacetic acid [EDTA] dipotassium, and iminodibenzyl), the overall IS(B) analysis method score and SD were zero for each of the eggs tested. The replicate egg repeatability %CV values for individual experiments, excluding studies where such values could not be calculated, ranged from 0.12 to 173.21 for hemorrhage, from 0.25 to 173.21 for lysis, from 0.00 to 173.21 for coagulation, and from 0.25 to 173.21 for the overall *in vitro* IS(B) score (see **Table 7-1** and **Appendix E3**).

The mean and median replicate egg repeatability %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E3**), excluding studies where the overall IS(B) score and SD were zero, were 32.52 and 11.49, respectively (**Table 7-1**). When the data for the nine substances noted were removed, the mean and median replicate egg repeatability %CV values for the overall IS(B) scores were 41.48 and 17.54, respectively (**Table 7-1**).

7.2.1.2 *Gilleron et al. (1997)*

Individual egg results for 60 substances evaluated by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1997) were provided to NICEATM. Among the data, the original test results for four of the 60 substances included in the 1997 publication (Maneb, 1-napthalene acetic acid, Tween 20, and 1-napthalene acetic acid, sodium salt) were no longer available. Since alternative HET-CAM test data were available for these substances, these data were provided to NICEATM. The overall replicate egg mean and median %CV values were evaluated with and without the inclusion of these data.

Table 7-1 Intralaboratory Repeatability Results for HET-CAM Studies of Gilleron et al. (1996)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.64 (1.93)	2.68 (2.88)	3.59 (3.44)	7.92 (5.84)
Range of Values for All Substances	0.12-173.21	0.25-173.21	0.00-173.21	0.25-173.21
%CV for Substances[2]	117.56	107.52	95.69	73.74
Number of Experiments	146	146	146	146
Mean Value (SD) Excluding Nine Substances Where Original Data Was Not Available[1]	1.63 (1.90)	1.87 (2.57)	2.83 (3.25)	6.33 (5.43)
Range of Values Excluding Nine Substances Where Original Data Was Not Available	0.12-173.21	0.25-173.21	0.00-173.21	0.35-173.21
%CV Excluding Nine Substances Where Original Data Was Not Available[2]	116.13	137.49	115.07	85.84
Number of Experiments	111	111	111	111
Mean Overall *In Vitro* Score %CV for All Substances	32.52			
Median Overall *In Vitro* Score %CV for all Substances	11.49			
Mean Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	41.48			
Median Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	17.54			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E3**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

For each test substance, three different eggs were used in each of at least three replicate experiments. For this evaluation, the %CV values were determined for each endpoint (hemorrhage, lysis, coagulation) and for the overall *in vitro* IS(B) score. For each of the endpoints, there were experiments where test substances did not produce any effects (i.e., the average score of the three replicate eggs and standard deviation [SD] of the scores were both 0) (see **Appendix E4**). For the hemorrhage endpoint, 91 of 184 experiments (49%) resulted in an average score and SD of zero for the three replicate eggs; for the lysis endpoint, 22 of 184 experiments (12%) resulted in an average score and SD of zero; while, for the coagulation endpoint, 16 of 184 experiments (9%) resulted in an average score and SD of zero. For the overall *in vitro* IS(B) score, 6 of 184 experiments (3%) resulted in an average score and SD of zero for the three replicate eggs (**Appendix E4**). For one test substance (Maneb), the overall IS(B) analysis method score and SD were zero for each of the eggs tested. The replicate egg repeatability %CV values for individual experiments, excluding studies where such values could not be calculated, ranged from 0.23 to 173.21 for hemorrhage, from 0.00 to 173.21 for lysis, from 0.37 to 173.21 for coagulation, and from 0.13 to 173.21 for the overall *in vitro* IS(B) score (**Table 7-2** and **Appendix E4**).

The mean and median replicate egg repeatability %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E4**), excluding studies where such values could not be calculated, were 7.61 and 2.24, respectively (**Table 7-2**). When the data for the four substances noted were removed the mean and median replicate egg repeatability %CV values for the overall IS(B) scores were 6.99 and 2.04, respectively (**Table 7-2**).

7.2.2 Quantitative Assessment of Intralaboratory Reproducibility

Interlaboratory variability can be evaluated by assessing the CV or by using ANOVA methods. Two studies discussed in **Section 6.0** included intralaboratory reproducibility data (Gilleron et al. 1996, 1997). For both sets of studies, quantitative HET-CAM test method data were available for studies repeated three to five times in a single laboratory.

7.2.2.1 *Gilleron et al. (1996)*

Individual experimental results for 46 substances evaluated by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1996) were received in response to a request from NICEATM. In the data provided to NICEATM, the test results for nine of the 46 substances included in the 1996 publication (laurylsulfobetaine, deoxycholic acid, ethylacetoacetate, methyl isobutyl ketone, methanol, N-laurylsarcosine, promethazine hydrochloride, 2-methoxyethanol, benzethonium chloride, and imidazole) were no longer available. Since alternative HET-CAM test data generated were available for these substances, these data were provided to NICEATM. The overall mean and median %CV values for replicate experiments were evaluated with and without the inclusion of these data.

In these studies, three different eggs were used for each experiment. Three experiments were conducted for each test substance, except for the nine substances where nonoriginal data was provided. For these substances, data for three to five experiments were provided.

Table 7-2 Intralaboratory Repeatability Results for HET-CAM Studies of Gilleron et al. (1997)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.94 (2.12)	5.60 (2.31)	6.42 (2.68)	13.96 (4.89)
Range of Values for All Substances	0.23-173.21	0.00-073.21	0.37-173.21	0.13-173.21
%CV for Substances[2]	109.10	41.24	41.78	34.99
Number of Experiments	184	184	184	184
Mean Value (SD) Excluding Four Substances Where Original Data Was Not Available[1]	2.07 (2.16)	5.75 (2.19)	6.60 (2.49)	14.42 (4.48)
Range of Values Excluding Four Substances Where Original Data Was Not Available	0.23-173.21	0.00-073.21	0.37-173.21	0.13-173.21
%CV Excluding Four Substances Where Original Data Was Not Available[2]	104.43	38.04	37.78	31.05
Number of Experiments	168	168	168	168
Mean Overall *In Vitro* Score %CV for All Substances	7.61			
Median Overall *In Vitro* Score %CV for all Substances	2.24			
Mean Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	6.99			
Median Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	2.04			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E4**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

For each of the endpoints, there were a number of experiments where the test substance did not induce any effects (i.e., the average score of the repeated experiments and SD of the scores were both 0) (see **Appendix E5**). For the overall *in vitro* IS(B) score, three of 46 experiments (7%) resulted in an average score and SD of zero for the repeated experiments (**Appendix E5**). For EDTA, the overall IS(B) analysis method score and SD were zero for all replicate experiments. The reproducibility %CV values for individual substances, excluding studies where such values could not be calculated, ranged from 2.59 to 173.21 for hemorrhage, from 1.55 to 173.21 for lysis, from 1.52 to 173.21 for coagulation, and from 6.66 to 173.21 for the overall *in vitro* IS(B) score (**Appendix E5** and **Table 7-3**).

The mean and median reproducibility %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E5**), excluding studies where such values could not be calculated, were 52.73 and 33.70, respectively (**Table 7-3**). When the data for the nine substances noted were removed, the mean and median reproducibility %CV values for the overall IS(B) scores were 60.66 and 39.15, respectively (**Table 7-3**).

7.2.2.2 Gilleron et al. (1997)

Individual experimental results for 60 substances evaluated by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1997) were provided by the authors to NICEATM. Among the data, the original test results for four of the 60 substances included in the 1997 publication (Maneb, 1-napthalene acetic acid, Tween 20, and 1-napthalene acetic acid, sodium salt) were no longer available. Since alternative HET-CAM test data were available for these substances, these data were provided to NICEATM. The overall mean and median %CV values for replicate experiments were evaluated with and without the inclusion of these data.

In these studies, three different eggs were used for each experiment. Three experiments were conducted for each test substance, except for the four substances where nonoriginal data was provided. For these substances, data for three to five experiments were provided.

For each of the endpoints, there were a number of experiments where the test substance did not induce any effects (i.e., the average score of the three replicate eggs and thus the SD of the scores were both zero) (see **Appendix E6**). For the overall *in vitro* IS(B) score, none of substances resulted in an average score and SD of zero for the three replicate experiments (**Appendix E6**). The reproducibility %CV values for individual substances, excluding studies where such values could not be calculated, ranged from 0.20 to 173.21 for hemorrhage, from 0.12 to 200.00 for lysis, from 0.00 to 173.21 for coagulation, and from 0.34 to 200.00 for the overall *in vitro* IS(B) score (**Appendix E6** and **Table 7-4**).

The mean and median reproducibility %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E6**), excluding studies where such values could not be calculated, were 17.48 and 6.34, respectively (**Table 7-4**). When the data for the nine substances noted were removed, the mean and median reproducibility %CV values for the overall IS(B) scores were 13.49 and 5.25, respectively (**Table 7-4**).

Table 7-3 Intralaboratory Reproducibility Results for HET-CAM Studies of Gilleron et al. (1996)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.64 (2.04)	2.68 (2.96)	3.59 (3.52)	7.51 (5.28)
Range of Values for All Substances	2.59-173.21	1.55-173.21	1.52-173.21	6.66-173.21
%CV for Substances[2]	124.12	110.41	97.92	70.35
Mean Value (SD) Excluding Nine Substances Where Original Data Was Not Available[1]	1.63 (2.01)	1.87 (2.66)	2.83 (3.34)	6.33 (5.06)
Range of Values Excluding Nine Substances Where Original Data Was Not Available	2.59-173.21	1.55-173.21	4.84-173.21	14.33-173.21
%CV Excluding Nine Substances Where Original Data Was Not Available[2]	123.08	142.31	118.37	79.92
Mean Overall *In Vitro* Score %CV for All Substances	52.73			
Median Overall *In Vitro* Score %CV for all Substances	33.70			
Mean Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	60.66			
Median Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	39.15			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E5**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

Table 7-4 Intralaboratory Reproducibility Results for HET-CAM Studies of Gilleron et al. (1997)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.94 (2.12)	5.60 (2.31)	6.42 (2.68)	13.96 (4.89)
Range of Values for All Substances	0.20-173.20	0.12-200.00	0.00-173.21	0.34-200.00
%CV for Substances[2]	109.10	41.24	41.78	35.00
Mean Value (SD) Excluding Four Substances Where Original Data Was Not Available[1]	2.07 (2.16)	5.75 (2.18)	6.60 (2.50)	14.42 (4.48)
Range of Values Excluding Four Substances Where Original Data Was Not Available	0.20-173.21	0.12-173.21	0.00-173.21	0.34-118.75
%CV Excluding Four Substances Where Original Data Was Not Available[2]	104.43	38.04	37.78	31.05
Mean Overall *In Vitro* Score %CV for All Substances	17.48			
Median Overall *In Vitro* Score %CV for all Substances	6.34			
Mean Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	13.49			
Median Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	5.25			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E6**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

7.2.3 Quantitative and Qualitative Assessments of Interlaboratory Reproducibility

Generally, an analysis of interlaboratory variability has included such approaches as:

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

Several of the studies discussed in **Section 6.0** included interlaboratory data for at least a subset of the substances evaluated. Using this data, the ability of the HET-CAM test method to reproducibly identify ocular corrosives and severe irritants versus nonsevere irritants (i.e., moderate and slight irritant) and nonirritants were evaluated using two approaches.

In the first approach, a qualitative assessment of reproducibility was conducted. In this evaluation, the individual laboratory *in vitro* ocular irritation classification for each substance was used to evaluate the extent of agreement among the participating laboratories in their ability to identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants. The reliability of HET-CAM was assessed separately for each study (i.e., publication) with multiple laboratory data (see CEC 1991, Balls et al. 1995, Spielmann et al. 1996, Hagino et al. 1999). In an alternative approach, the reliability of HET-CAM was assessed after pooling data across comparative studies that used the same data analysis method. The analysis methods where there was interlaboratory data were IS(A), IS(B)-10, IS(B)-100, Q-Score, and S-Score for the GHS and EPA classification systems. For the EU classification system, all the same HET-CAM analysis methods could be evaluated, as well as the IS(B) analysis method.

Substances classified, based on HET-CAM test data, as corrosive/severe irritants or nonsevere irritants/nonirritants were further classified by their *in vivo* rabbit eye test results, as determined within the GHS (UN 2003), EPA (1996), and EU (2001) classification systems.

Because the focus of this reliability assessment is on the interlaboratory reproducibility of HET-CAM in identifying corrosives/severe irritants versus nonsevere irritants/nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants. For example, three laboratories could classify a chemical as a nonirritant and one laboratory could classify the same chemical as a moderate irritant. Within this analysis, where a nonirritant and moderate irritant classification would be placed together, this distribution of classification calls would be considered as 100% agreement between laboratories.

In the second approach, a quantitative assessment of reproducibility was determined. CVs for test substances, where laboratory scores were available, for substances tested were reported or determined. The reproducibility of HET-CAM was assessed for studies (i.e.,

publication) reviewed in **Sections 4.0** and **5.0** where individual testing laboratory data was available (see CEC 1991, Balls et al. 1995, Spielmann et al. 1996, Hagino et al. 1999).

As discussed in **Section 2.0**, there is no standardized data collection method for HET-CAM studies and several different analysis methods have been developed (i.e., IS, Q-Score, S-Score). Therefore, the reliability assessments conducted in this section were evaluated according to each of the analysis methods described.

7.2.3.1 *Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System*

Interlaboratory reproducibility for the HET-CAM test method was evaluated for the following reports: Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999). The agreement of classification calls among participating laboratories and its relationship to the GHS *in vivo* classification (UN 2003) for the substances tested in each report is provided in **Table 7-5**.

The participating laboratories were in 100% agreement in regard to the GHS ocular irritancy classification for 21 (45%) of the 47 substances tested when using the Q-Score (Balls et al. 1995). The extent of agreement between testing laboratories was greatest for substances correctly identified as GHS corrosives or severe irritants (60% [9/15] accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). Comparatively, greater disparity between laboratory substance classifications was observed for false positives (i.e., positive *in vitro* but negative *in vivo*) and those substances accurately classified as nonsevere irritants. For instance, 75% (12/16) of the false positives and 58% (7/12) of the correctly identified nonsevere irritants exhibited less than 100% agreement in the GHS irritancy classifications among laboratories.

In addition to the Q-Score, Balls et al. (1995) evaluated irritancy potential for some substances using an S-Score. The participating laboratories were in 100% agreement in regard to the GHS ocular irritancy classification for 13 (68%) of the 19 tested substances. Substances that were classified as false negatives and false positives exhibited the most discordant results, with 29% (2/7) of the false negatives and 100% (2/2) of the false positives exhibiting less than 100% classification agreement between testing laboratories. There was complete agreement among testing laboratories for substances correctly classified as severe irritants or nonsevere/nonirritants, based on the GHS classification system (UN 2003).

The participating laboratories were in 100% agreement for 85 (79%) of 107 substances evaluated with the IS(B)-10 analysis method (Spielmann et al. 1996). The extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM (94% [31/33]). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (56% [10/18] false positives had less than 100% concordance between testing laboratories).

Table 7-5 Evaluation of the Reliability of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the GHS Classification System, by Study

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with ≤50% Agreement among Labs
Balls et al. (1995)	Q	+/+	2 4	4 11	3 (75%)[4] 6 (55%)	-	- 4 (36%)	-	-	1 (25%) 1 (9%)
		+/-	-	-	-	-	-	-	-	-
		-/+	4	16	4 (25%)	-	9 (56%)	-	-	3 (19%)
		-/-	2 4	1 11	1 (100%) 4 (36%)	-	- 7 (64%)	-	-	-
		?/-	2	1	1 (100%)	-	-	-	-	-
		?/+	3 4	1 2	1 (100%) 1 (50%)	-	- 1 (50%)	-	-	-
		Total	2-4	47	21 (45%)	-	21 (45%)	-	-	5 (10%)
Balls et al. (1995)	S	+/+	2	4	4 (100%)	-	-	-	-	-
		+/-	2 3 4	1 4 2	1 (100%) 2 (50%) 2 (100%)	-	-	- 2 (50%) -	-	-
		-/+	2 4	1 1	-	-	-	-	-	1 (100%) 1 (100%)
		-/-	3 4	1 2	1 (100%) 2 (100%)	-	-	-	-	-
		?/-	3	1	-	-	-	1 (100%)	-	-
		?/+	2	2	1 (50%)	-	-	-	-	1 (50%)
		Total	2-4	19	13 (68%)	-	-	3 (16%)	-	3 (16%)

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with ≤50% Agreement among Labs
Spielmann et al. (1996)	IS(B) -10	+/+	2	18	16 (89%)	-	-	-	-	2 (11%)
			3	1	-	-	-	1 (100%)	-	-
		+/-	2	4	4 (100%)	-	-	-	-	-
			3	1	-	-	-	1 (100%)	-	-
		-/+	2	16	7 (44%)	-	-	-	-	9 (56%)
			3	2	1 (50%)	-	-	-	-	1 (50%)
		-/-	2	31	30 (97%)	-	-	-	-	1 (3%)
			3	2	1 (50%)	-	-	1 (50%)	-	-
		?/-	2	10	10 (100%)	-	-	-	-	-
			3	2	1 (50%)	-	-	1 (50%)	-	-
		?/+	2	16	14 (88%)	-	-	-	-	2 (11%)
			3	4	1 (25%)	-	-	2 (50%)	-	1 (25%)
		Total		107	85 (79%)			6 (6%)		16 (15%)
Spielmann et al. (1996)	IS(B) -100	+/+	2	17	16 (94%)	-	-	-	-	1 (6%)
			3	2	1 (50%)	-	-	1 (50%)	-	-
		+/-	2	2	2 (100%)	-	-	-	-	-
		-/+	2	27	20 (74%)	-	-	-	-	7 (26%)
			3	4	1 (25%)	-	-	3 (75%)	-	-
		-/-	2	17	16 (94%)	-	-	-	-	1 (6%)
		?/-	2	6	6 (100%)	-	-	-	-	-
			3	2	2 (100%)	-	-	-	-	-
		?/+	2	18	15 (83%)	-	-	-	-	3 (17%)
			3	4	2 (50%)	-	-	2 (50%)	-	-
		Total		99	81 (82%)			6 (6%)		12 (12%)
Hagino et al. (1999)	IS(A)	+/+	5	8	5 (63%)	2 (25%)	-	-	1 (12%)	-
		+/-	-	-			-	-	-	-
		-/+	5	3	3 (100%)		-	-	-	-
		-/-	5	4	1 (25%)	1 (25%)	-	-	2 (50%)	-
		?/-	-	-			-	-	-	-
		?/+	5	2	2 (100%)		-	-	-	-
		Total	2-4	17	11 (64%)	3 (18%)	-	-	3 (18%)	-

Abbreviation: GHS = Globally Harmonized System (UN 2003).
[1]Anal = analysis method used to transform the sample data into HET-CAM scores. IS(A) = method described in Luepke (1985); IS(B)-10 and IS(B)-100 = method described in Kalweit et al. (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).
[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category 1); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category 2A or 2B) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), a GHS classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.
[3]N indicates number of substances.
[4]Number in parentheses indicates percentage of tested chemicals.

For the IS(B)-100 analysis method (Spielmann et al. 1996), the participating laboratories were in 100% agreement for 81 (82%) of 99 substances evaluated. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM (94% [16/17]). Greater disparity between laboratory substance classifications was observed for substances that were identified as false positives (32% [10/31] false positives had less than 100% concordance between testing laboratories).

For the report by Hagino et al. (1999), there was 100% agreement in regard to the GHS ocular irritancy classification for 11 (64%) of the 17 substances. Discordance in the classification results was present for substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants. Discordance in the results obtained by different laboratories ranged from 37% (3/8) to 75% (3/4) of the substances within these two groups. Substances classified as false positives had the greatest extent of agreement among laboratories.

The overall reliability statistics, evaluated by HET-CAM data analysis method, for the IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-5**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For the IS(A) analysis method, the addition of two additional test substances evaluated by Kojima et al. (1995) yielded an overall concordance pattern consistent with what was observed for the Hagino et al. (1999) data alone. For the IS(B)-100 analysis method, the additional data from different testing laboratories were obtained from Gilleron et al. (1996, 1997) and Vinardell and Macián (1994). As with the IS(A) analysis method, the addition of IS(B)-100 results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Spielmann et al. (1996).

7.2.3.2 *Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System*

Reliability analyses for the HET-CAM test method were evaluated for the following two reports: Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999). The agreement of classification calls among participating laboratories and its relationship to the EPA (1996) *in vivo* classification for the substances tested in each report is provided in **Table 7-6**.

The participating laboratories were in 100% agreement in regard to the EPA ocular irritancy classification for 21 (45%) of the 47 substances tested when using the Q-Score (Balls et al. 1995). The extent of agreement between testing laboratories was greatest for substances correctly identified as EPA corrosives or severe irritants (71% [10/14] of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between laboratory substance classifications was observed for false positives (i.e., positive *in vitro* but negative *in vivo*) and those substances accurately classified as nonsevere irritants. For instance, 76% (13/17) of the false positives and 58% (7/12) of the correctly identified nonsevere irritants exhibited less than 100% agreement in the EPA irritancy classifications among laboratories.

Table 7-6 Evaluation of the Reliability of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the EPA Classification System, by Study

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with 50% or Less Agreement among Labs
Balls et al. (1995)	Q	+/+	2 4	4 10	3 (75%)[4] 7 (70%)	-	- 3 (30%)	-	-	1 (25%)
		+/-	-	-	-	-	-	-	-	-
		-/+	4	17	4 (24%)	-	9 (52%)	-	-	4 (24%)
		-/-	2 4	1 11	1 (100%) 4 (36%)	-	- 7 (64%)	-	-	-
		?/-	2	1	1 (100%)	-		-	-	-
		?/+	3 4	1 2	1 (100%)	-	- 2 (50%)	-	-	-
		Total	2-4	47	21 (45%)	-	21 (45%)	-	-	5 (10%)
Balls et al. (1995)	S	+/+	2	3	3 (100%)	-	-	-	-	-
		+/-	3 4	3 2	2 (66%) 2 (100%)	-	-	1 (33%) -	-	-
		-/+	2 4	1 1	-	-	-	-	-	1 (100%) 1 (100%)
		-/-	3 4	1 2	1 (100%) 2 (100%)	-	-	-	-	-
		?/-	2 3	1 2	1 (100%)	-	-	- 2 (100%)	-	-
		?/+	2	2	1 (50%)	-	-	-	-	1 (50%)
		Total	2-4	18	12 (66%)	-	-	3 (17%)	-	3 (17%)

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with 50% or Less Agreement among Labs
Spielmann et al. (1996)	IS(B)-10	+/+	2	9	8 (89%)	-	-	-	-	1 (11%)
			3	1	-	-	-	1 (100%)	-	-
		+/-	2	3	3 (100%)	-	-	-	-	-
		-/+	2	18	9 (50%)	-	-	-	-	9 (50%)
			3	3	1 (33%)	-	-	1 (33%)	-	1 (33%)
		-/-	2	31	31 (100%)	-	-	-	-	-
			3	2	1 (50%)	-	-	1 (50%)	-	-
		?/-	2	10	10 (100%)	-	-	-	-	-
			3	3	1 (33%)	-	-	2 (66%)	-	-
		?/+	2	21	19 (90%)	-	-	-	-	2 (10%)
			3	3	1 (33%)	-	-	1 (33%)	-	1 (33%)
		Total	2-3	104	84 (81%)			6 (6%)		14 (13%)
Spielmann et al. (1996)	IS(B)-100	+/+	2	10	9 (90%)	-	-	-	-	1 (10%)
			3	1	1 (100%)	-	-	-	-	-
		+/-	2	1	1 (100%)	-	-	-	-	-
		-/+	2	29	22 (76%)	-	-	-	-	7 (24%)
			3	4	1 (25%)	-	-	3 (75%)	-	-
		-/-	2	17	16 (94%)	-	-	-	-	1 (6%)
			3	1	1 (100%)					-
		?/-	2	7	7 (100%)	-	-	-	-	-
			3	1	1 (100%)	-	-	-	-	-
		?/+	2	21	19 (90%)	-	-	-	-	2 (10%)
			3	5	2 (40%)	-	-	3 (60%)	-	-
		Total	2-3	97	80 (82%)			6 (6%)		11 (11%)
Hagino et al. (1999)	IS(A)	+/+	5	7	5 (71%)	2 (29%)	-	-	-	-
		+/-	-	-	-	-	-	-	-	-
		-/+	5	4	4 (100%)	-	-	-	-	-
		-/-	5	3	1 (33%)	-	-	-	2 (66%)	-
		?/-	-	-	-	-	-	-	-	-
		?/+	5	2	1 (50%)	-	-	-	1 (50%)	-
		Total	-	16	11 (69%)	3 (27%)	-	-	3 (27%)	-

Abbreviation: EPA = U.S. Environmental Protection Agency (EPA 1996).
[1]Anal = analysis method used to transform the sample data into HET-CAM scores. IS(A) = method described in Luepke (1985); IS(B) = method described in Kalweit et al (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).
[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III, or IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 6.1** of the HET-CAM BRD for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro.*
[3]N indicates number of substances.
[4]Number in parentheses indicates percentage of tested chemicals.

In addition to the Q-Score, Balls et al. (1995) evaluated irritancy potential for some substances by using an S-Score. The participating laboratories were in 100% agreement in regard to the EPA ocular irritancy classification for 12 (66%) of the 18 tested substances. Substances that were classified as false negatives and false positives exhibited the most discordant results, with 20% (1/5) of false negatives and 100% (2/2) of false positives exhibiting less than 100% classification agreement among testing laboratories. There was complete agreement among testing laboratories for substances correctly classified as severe irritants or nonsevere/nonirritants, based on the EPA classification system.

The participating laboratories were in 100% agreement for 84 of the 104 (81%) substances evaluated using the IS(B)-10 analysis method (Spielmann et al. 1996). The extent of agreement between testing laboratories was greatest for substances correctly identified as EPA nonsevere irritants or nonirritants by HET-CAM (97% [32/33]). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (52% [11/21] false positive had less than 100% concordance between testing laboratories).

For the IS(B)-100 analysis method (Spielmann et al. 1996), the participating laboratories were in 100% agreement 80 (82%) of the 97 substances tested. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as EPA nonsevere irritants or nonirritants by HET-CAM (94% [17/18]). Greater disparity between laboratory substance classifications was observed for substances that were identified as false positives (33% [10/33] false positive had less than 100% concordance between testing laboratories).

For the report by Hagino et al. (1999), there was 100% agreement in regard to the EPA ocular irritancy classification for 11 (69%) of the 16 substances. Discordance in the classification results was observed for substances that were correctly identified as nonsevere irritants/nonirritants. Of the three correctly identified nonsevere irritants/nonirritants, two substances had less than 100% classification agreement among the laboratories. For EPA severe irritants, there was 100% laboratory agreement for 71% (5/7) of the tested substances.

The overall reliability statistics, evaluated by HET-CAM data analysis method, for the IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-6**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For the IS(A) analysis method, the addition of two additional test substances evaluated by Kojima et al. (1995) yielded an overall concordance pattern consistent with what was observed for the Hagino et al. (1999) data alone. For the IS(B)-100 analysis method, the additional data from different testing laboratories were obtained from Gilleron et al. (1996, 1997) and Vinardell and Macián (1994). As with the IS(A) analysis method, the addition of IS(B)-100 results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Spielmann et al. (1996).

7.2.3.3 *Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System*

Reliability analyses for the HET-CAM test method were evaluated for the following four reports: CEC (1991), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999). The agreement of classification calls among participating laboratories and its relationship to the EU (2001) *in vivo* classification for the substances tested in each report is provided in **Table 7-7**.

For the CEC evaluation, the participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 6 (23%) of the 26 substances tested when using the IS(B) analysis method. The extent of agreement among laboratories was greatest for accurately identified EU corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (50% [3/6] of the identified EU corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives and those substances accurately classified as EU nonsevere irritants/nonirritants. For instance, 100% (9/9) of the false positives and 70% (7/10) of the correctly identified EU nonsevere irritants/nonirritants exhibited less than 100% agreement among laboratories in irritancy classifications.

The participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 21 (45%) of the 47 substances tested when using the Q-Score (Balls et al. 1995). The extent of agreement between testing laboratories was greatest for substances correctly identified as EU corrosives or severe irritants (69% [9/13] of accurately identified EU corrosives/severe irritants exhibited 100% classification agreement among testing laboratories). Comparatively, greater disparity between laboratory substance classifications was observed for false positives and accurately identified EU nonsevere irritants/nonirritants. For instance, 71% (10/14) of the false positives and 58% (7/12) of the correctly identified EU nonsevere irritants/nonirritants exhibited less than 100% agreement among laboratories in irritancy classifications.

In addition to the Q-Score, Balls et al. (1995) evaluated irritancy potential for some substances using an S-Score. The participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 12 (66%) of the 18 tested substances. Substances classified as false positives exhibited the most discordant results, with 100% (2/2) exhibiting less than 100% agreement in classification among laboratories.

The participating laboratories were in 100% agreement for 84 of the 106 (79%) substances evaluated with the IS(B)-10 analysis method (Spielmann et al. 1996). The extent of agreement between testing laboratories was greatest for substances correctly identified as EU nonsevere irritants or nonirritants by HET-CAM (93% [31/33]). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (58% [11/19] false positive had less than 100% concordance between testing laboratories).

Table 7-7 Evaluation of the Reliability of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the EU Classification System, by Study

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 75-99% Agreement among Labs	Substances with 50-74% Agreement among Labs	Substances with 25-49% Agreement among Labs
CEC (1991)	IS(B)	+/+	3	3	3 (100%)	-	-	-
			5	1	-	-	1 (100%)	-
			6	2	-	1 (50%)	1 (50%)	-
		+/-	7	1	-	1 (100%)	-	-
		-/+	3	3	-	-	1 (33%)	2 (66%)
			7	6	-	1 (17%)	2 (34%)	3 (51%)
		-/-	3	6	3 (50%)	-	2 (33%)	1 (17%)
			7	4	-	2 (50%)	2 (50%)	-
		?/-	-	-	-	-	-	-
		?/+	-	-	-	-	-	-
		Total	3-7	26	6 (23%)	5 (19%)	9 (35%)	6 (23%)
Balls et al. (1995)	Q	+/+	2	4	3 (75%)[4]	-	1 (25%)	-
			4	9	6 (67%)	3 (37%)		
		+/-	-	-	-	-	-	-
		-/+	4	14	4 (28%)	7 (50%)	3 (21%)	-
		-/-	2	1	1 (100%)	-	-	-
			4	11	4 (36%)	7 (63%)		
		?/-	2	1	1 (100%)	-	-	-
		?/+	3	1	1 (100%)	-	-	-
			4	6	1 (17%)	4 (67%)	1 (17%)	
		Total	2-4	47	21 (45%)	21 (45%)	5 (10%)	-

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 75-99% Agreement among Labs	Substances with 50-74% Agreement among Labs	Substances with 25-49% Agreement among Labs
Balls et al. (1995)	S	+/+	2	3	3 (100%)	-	-	
		+/-	2 3 4	1 3 2	1 (100%) 2 (66%) 2 (100%)	-	- 1 (33%)	
		-/+	2 4	1 1	-	-	1 (100%) 1 (100%)	
		-/-	3 4	1 2	1 (100%) 2 (100%)	-	-	
		?/-	3	2	-	-	2 (100%)	
		?/+	2	2	1 (50%)	-	1 (50%)	
		Total	2-4	18	12 (66%)	-	6 (34%)	
Spielmann et al. (1996)	IS(B)-10	+/+	2 3	12 1	11 (92%) -	- -	1 (8%) 1 (100%)	- -
		+/-	2	3	3 (100%)	-	-	-
		-/+	2 3	17 2	7 (41%) 1 (50%)	- -	- 1 (50%)	10 (59%) -
		-/-	2 3	31 2	30 (97%) 1 (50%)	- -	1 (3%) 1 (50%)	- -
		?/-	2 3	11 3	11 (100%) 1 (33%)	- -	- 2 (66%)	- -
		?/+	2 3	20 4	18 (90%) 1 (25%)	-	2 (10%) 2 (50%)	- 1 (25%)
		Total	2-3	106	84 (79%)		11 (10%)	11 (10%)

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 75-99% Agreement among Labs	Substances with 50-74% Agreement among Labs	Substances with 25-49% Agreement among Labs
Spielmann et al. (1996)	IS(B)-100	+/+	2	12	11 (92%)	-	1 (8%)	-
			3	1	1 (100%)	-	-	-
		+/-	2	1	1 (100%)	-	-	-
		-/+	2	28	21 (75%)	-	-	7 (25%)
			3	4	1 (25%)	-	3 (75%)	-
		-/-	2	17	16 (94%)	-	-	1 (6%)
		?/-	2	7	7 (100%)	-	-	-
			3	2	2 (100%)	-	-	-
		?/+	2	21	18 (86%)	-	-	3 (24%)
			3	2	2 (100%)	-	-	-
		Total	2-3	95	80 (84%)		4 (4%)	11 (11%)
Hagino et al. (1999)	IS(A)	+/+	5	7	5 (71%)	1 (14%)	1 (14%)	-
		+/-	-	-	-	-	-	-
		-/+	5	4	3 (75%)	1 (25%)	-	-
		-/-	5	3	1 (33%)	-	2 (66%)	-
		?/-	-	-	-	-	-	-
		?/+	5	2	2 (100%)	-	-	-
		Total	2-4	16	11 (69%)	2 (12%)	3 (19%)	-

Abbreviation: EU = European Union (EU 2001).
[1]Anal = analysis method used to transform the sample data into HET-CAM scores. IS(A) = method described in Luepke (1985); IS(B) = method described in Kalweit et al. (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).
[2]A "+" indicates that the substance was assigned an overall classification of corrosive or severe irritant (Category R41); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category R36) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (i.e., insufficient dose volume), an EU classification could not be made. See Section 6.1 of the Draft HET-CAM BRD for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.
[3]N indicates number of substances.
[4]Number in parentheses indicates percentage of tested chemicals.

For the IS(B)-100 analysis method (Spielmann et al. 1996), the participating laboratories were in 100% agreement for 80 (84%) of the 95 substances tested. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as EU nonsevere irritants or nonirritants by HET-CAM (94% [16/17]). Greater disparity between individual substance classifications was observed for substances that were identified as false positives (31% [10/32] false positive had less than 100% concordance between testing laboratories).

For the report by Hagino et al. (1999), there was 100% agreement in regard to the EU ocular irritancy classification for 11 (69%) of the 16 substances. Discordance in the classification results was observed for substances that were correctly identified as EU nonsevere irritants/nonirritants. Of three correctly identified EU nonsevere irritants/nonirritants, two substances exhibited less than 100% classification agreement among laboratories. Of the seven correctly identified EU corrosives/severe irritants, five substances (71%) produced the same classification in all five laboratories.

The overall reliability statistics, evaluated by HET-CAM data analysis method, for the IS(B), IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-7**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For the IS(A) analysis method, the addition of two additional test substances evaluated by Kojima et al. (1995) yielded an overall concordance pattern consistent with what was observed for the Hagino et al. (1999) data alone. For the IS(B)-100 analysis method, the additional data from different testing laboratories were obtained from Gilleron et al. (1996, 1997) and Vinardell and Macián (1994). As with the IS(A) analysis method, the addition of the results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Spielmann et al. (1996).

7.2.3.4 *Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results Based On Qualitative Analyses*

For each of the *in vivo* classifications systems, there were few substances that were evaluated in all reports discussed. Therefore, a direct comparison of the reliability of the analysis methods used by each report is limited. In general the ability of the HET-CAM test method to identify corrosives and severe irritants (for substances where there is repeated data to assess reproducibility and reliability) was similar between hazard classification systems evaluated. Therefore, conclusions about the HET-CAM reproducibility for one *in vivo* classification system generally apply to all classification systems, unless otherwise noted.

For the IS(A) analysis method, there were four false positive substances. The chemical classes represented by these substances included amidine, ether, carboxylic acid, amine, and alcohol. There were no chemicals or substances tested multiple times in different laboratories that were classified as false negatives to allow for an evaluation of common chemical product classes.

For the IS(B)-10 analysis method, the most common chemical classes shown to overpredicted, and where there were discordant results between testing laboratories was

alcohols. Other chemical classes, where discordant results were observed, included amines and phenols. Substances that were underpredicted tended to be underpredicted by all the testing laboratories that evaluated the substance.

For the IS(B)-100 analysis method, most of the substances evaluated produced the same response in all testing laboratories. Of the substances where there were discordant results, the chemical classes included, esters, aldehydes, and amines.

For alcohols that were evaluated using the Q-Score analysis method and were defined as false positives by the HET-CAM test method, the extent of agreement among laboratories was 75% (i.e., three of four laboratories classified the alcohol as a severe irritant). The extent of agreement among laboratories for the classification of esters (e.g., methyl acetate, ethyl-2-methylacetoacetate), which were false positives, ranged between 50% and 75%. Compared to the Q-Score, there were not enough tested substances within each *in vivo/in vitro* combination for S-Score or the IS(A) analysis methods to draw similar conclusions for the *in vivo* classification system.

7.2.3.5 *Quantitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category*

CEC (1991): Between three and five laboratories evaluated each substance tested at 100% concentration. A subset of substances was evaluated at a concentration of 10% by three testing laboratories. Based on the two different data sets, two different evaluations were conducted[2]. For the substances tested at a 100% concentration *in vitro*, substances tested by five laboratories (excluding Laboratories #5 and #6) were assessed[3]. For the substances tested at a 10% concentration *in vitro*, substances tested by three laboratories were assessed.

Using these criteria, %CV values for 14 substances evaluated at 100% concentration and 12 substances evaluated at 10% concentration were determined. The mean and median %CV values for substances evaluated at 100% concentration *in vitro* were 31.86 and 33.04, respectively (**Table 7-8**). The mean and median %CV values for substances evaluated at 10% concentration *in vitro* were 66.29 and 60.75, respectively (**Table 7-9**).

Balls et al. (1995): This evaluation used two different analysis methods, the S-Score and Q-Score. A description of each of these analyses methods is provided in **Section 5.0**. The use of these different analysis methods was dependent upon the transparency of the test materials. For substances where the reactions on the CAM could be observed the Q-Score was calculated. Comparatively, for substances where the reactions on the CAM could not be observed the S-Score was calculated (**Appendix A** provides a description of the differences in the test method protocols used for each analysis method).

---

[2] Data for these analyses are available in **Appendix F1.**

[3] Data from these testing laboratories were excluded from this analysis because the study report indicated that both laboratories had difficulty in identifying lysis and thrombosis/coagulation.

Table 7-8 %CV Values for Substances Evaluated at 100% Concentration *In Vitro* Using the IS(B) Analysis Method (from CEC 1991)

Substance[1]	Conc.	Mean IS(B) Score	SD	%CV Values
2-Butoxyethyl acetate	100%	4.76	0.31	6.58
Butanol	100%	11.44	1.0	8.71
Chloroform	100%	12.8	2.43	18.98
Triacetin	100%	4.18	0.91	21.76
Glycerol	100%	9.32	2.62	28.14
Tributyltin chloride	100%	8.94	2.88	32.21
Dimethyl sulfoxide	100%	9.88	3.24	32.83
Sodium dodecyl sulfate	100%	10.02	3.33	33.25
Triethanolamine	100%	8.52	2.94	34.55
Toluene	100%	11.04	4.31	39.06
2-Methoxyethanol	100%	9.14	3.72	40.65
Mercuric chloride	100%	10.52	4.57	43.44
n-Hexane	100%	5.04	3.16	62.78
Brij 35	100%	5.58	4.18	74.90
Mean %CV of substances tested at 100%				**31.86**
Median %CV of substances tested at 100%				**33.04**
Range %CV of substances tested at 100%				**6.58-74.90**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; SD = standard deviation.
[1]Substances organized by increasing %CV values.

Between two and four laboratories evaluated each substance tested in this report. For this evaluation, only substances tested by all four laboratories were assessed[4]. Using this criteria, %CV values for 40 substances evaluated using the Q-Score and five substances evaluated using the S-Score were determined. The average and median %CV values for substances evaluated with the Q-Score were 49.83 and 42.50 (range of %CVs: 15.09 to 157.25), respectively (**Table 7-10**). The average and median %CV values for substances evaluated with the S-Score were 84.42 and 71.90 (range of %CVs: 68.47 to 116.4), respectively (**Table 7-11**).

The average and median %CV values for GHS Category 1 substances (UN 2003), based on *in vivo* results, were 36.26 and 38.93 for the Q-Score and 81.53 and 81.53 for the S-Score. The average and median %CV value for EPA Category I substances (EPA 1996), based on *in vivo* results, were, 33.54 and 34.81 for the Q-Score and 81.53 and 81.53 for the S-Score.

[4] Individual laboratory data is available in **Appendix C**.

Table 7-9 %CV Values for Substances Evaluated at 10% Concentration *In Vitro* Using the IS(B) Analysis Method (from CEC 1991)

Substance[1]	Conc.	Mean IS(B) Score	SD	%CV Values
Dimethyl sulfoxide	10%	4.20	0.17	4.12
Tributyltin chloride	10%	12.13	3.11	25.61
Acetic acid	10%	14.67	5.08	34.67
Butanol	10%	10.50	5.01	47.70
Glycerol	10%	5.57	2.74	49.27
Sodium dodecyl sulfate	10%	12.53	6.79	54.15
Chloroform	10%	7.20	4.85	67.36
2-Butoxyethyl acetate	10%	2.43	2.15	88.56
Triacetin	10%	6.30	6.36	100.88
2-Methoxyethanol	10%	3.37	3.51	104.19
Triethanolamine	10%	5.07	5.46	107.86
n-Hexane	10%	4.60	5.11	111.08
Mean %CV of substances tested at 10%				**66.29**
Median %CV of substances tested at 10%				**60.75**
Range %CV of substances tested at 10%				**4.12-111.08**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; SD = standard deviation.
[1]Substances organized by increasing %CV values.

Spielmann et al. (1996): Individual laboratory results on tested substances were provided in response to a request by NICEATM[5]. In the evaluation, substances were evaluated at a 10% and 100% concentration in at least two different testing laboratories. Therefore, evaluation of the reliability of the test method was conducted for each concentration tested. Additionally, in order to resolve discrepancies in results between testing laboratories, some substances were tested in one additional testing laboratory (substances are italicized in **Table 7-12**). In order to determine if the substance tested in three laboratories affected the overall %CV values, an evaluation of the overall %CV values was conducted with these substances removed.

The average and median %CV values for substances tested at 10% concentration were 60.17 and 42.65, respectively. For substances tested at 100% concentration, the average and median %CV values were lower: 35.21 and 26.22, respectively. When substances that were tested in three different testing laboratories were removed from the assessment, little change was seen in the mean and median %CV values for both concentrations tested (**Table 7-12**).

[5] Individual laboratory data is available in **Appendix C**.

Table 7-10 %CV Values for Substances Evaluated Using the Q-Score Analysis Method (from Balls et al. 1995)

Substance[1]	Conc.	GHS Category 1	EPA Category I	Mean Q-Score	SD	%CV Values
2,2-Dimethylbutanoic acid	-	-	X	12.78	1.93	15.09
Trichloroacetic acid	30%	X	X	12.32	1.89	15.35
Benzalkonium chloride	1%	X	X	4.18	0.68	16.29
Sodium hydroxide	1%	-	-	5.42	0.99	18.20
Butyl acetate	-	-	-	1.63	0.31	18.95
Methyl cyanoacetate	-	-	-	1.38	0.34	24.84
Sodium lauryl sulfate	-	-	-	2.12	0.53	25.25
Triton X-100	5%	-	-	2.25	0.61	27.14
Octanol	-	-	-	1.67	0.47	28.15
Cyclohexanol	-	X	X	4.91	1.42	29.01
Benzalkonium chloride	10%	X	X	5.59	1.72	30.68
Ethyl-2-methylacetoacetate	-	-	-	2.09	0.66	31.74
Methyl isobutyl ketone	-	-	-	1.67	0.53	31.76
Cetylpyridinium bromide	6%	X	-	2.29	0.75	32.56
Triton X-100	10%	-	-	2.32	0.82	35.62
Hexanol	-	-	-	3.88	1.45	37.40
Methyl ethyl ketone	-	-	-	4.60	1.72	37.45
Toluene	-	-	-	3.73	1.41	37.98
Sodium lauryl sulfate	15%	X	X	2.84	1.11	38.93
Cetylpyridinium bromide	10%	X	X	2.98	1.21	40.60
Parafluoraniline	-	-	-	3.55	1.57	44.31
Polyethylene glycol 400	-	-	-	1.03	0.46	44.41
Pyridine	-	X	X	8.74	3.88	44.42
Tween 20	-	X	-	0.58	0.27	45.98
Sodium hydroxide	10%	X	X	13.44	6.74	50.12
Isobutanol	-	-	-	3.82	1.98	51.99
Trichloroacetic acid	3%	-	-	10.79	5.68	52.67
Benzalkonium chloride	5%	X	X	4.76	2.61	54.87
Ethyl acetate	-	-	-	2.52	1.39	55.11
Methyl acetate	-	-	-	3.03	1.70	56.12
Ethanol	-	-	-	6.13	3.75	61.16
Acetone	-	-	-	10.75	7.41	68.95
Glycerol	-	-	-	0.79	0.56	70.83
Isopropanol	-	-	-	5.96	4.23	71.93
2,6-Dichlorobenzoyl chloride	-	-	-	5.85	4.23	72.44
2-Ethyl-1-hexanol	-	-	-	1.49	1.12	74.75
Ethyl trimethyl acetate	-	-	-	0.40	0.41	103.70
gamma-Butyrolactone	-	-	-	8.67	9.12	105.19
Cetylpyridinium bromide	0.1%	-	-	0.86	1.15	134.05
Methylcyclopentane	-	-	-	2.42	3.81	157.25

Substance[1]	Conc.	GHS Category 1	EPA Category I	Mean Q-Score	SD	%CV Values
Mean for All Substances (n=40)	-	-	-	-	-	**49.83**
Median for All Substances	-	-	-	-	-	**42.50**
Range for All Substances	-	-	-	-	-	**15.09-157.25**
Mean for Severe Irritants (GHS) (n=11)	-	-	-	-	-	**36.26**
Median for Severe Irritants	-	-	-	-	-	**38.93**
Range for Severe Irritants	-	-	-	-	-	**15.35-54.87**
Mean for Severe Irritants (EPA) (n=8)	-	-	-	-	-	**33.54**
Median for Severe Irritants	-	-	-	-	-	**34.81**
Range for Severe Irritants	-	-	-	-	-	**15.35-54.87**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; EPA = U.S. Environmental Protection Agency (EPA 1996); GHS = Globally Harmonized System (UN 2003); SD = standard deviation.
[1]Substances organized by increasing %CV values.

Table 7-11 %CV Values for Substances Evaluated Using the S-Score Analysis Method (from Balls et al. 1995)

Substance[2]	GHS Category 1	EPA Category I	Mean S-Score	Standard Deviation	%CV
4-Carboxybenzaldehyde	-	-	4	2.83	70.71
Fomasafen	-	-	5.25	3.77	71.90
1-Napthalene acetic acid	X	X	5.75	5.44	94.59
Sodium oxalate	X	X	8	5.48	68.47
Dibenzyl phosphate	-	-	8.25	9.60	116.42
Mean for All Substances (n=5)	-	-	-	-	**84.42**
Median for All Substances	-	-	-	-	**71.90**
Range for All Substances	-	-	-	-	**68.47-116.4**
Mean for Severe Irritants (GHS) (n=2)	-	-	-	-	**81.53**
Median for Severe Irritants	-	-	-	-	**81.53**
Range for Severe Irritants	-	-	-	-	**68.47-94.59**
Mean for Severe Irritants (EPA) (n=2)	-	-	-	-	**81.53**
Median for Severe Irritants	-	-	-	-	**81.53**
Range for Severe Irritants	-	-	-	-	**68.47-94.59**

Abbreviations: %CV = percent coefficient of variation; EPA = U.S. Environmental Protection Agency (EPA 1996); GHS = Globally Harmonized System (UN 2003).
[1]Substances organized by increasing %CV values.

Table 7-12 %CV Values for Substances Evaluated Using IS(B) Analysis Method (from Spielmann et al. 1996)

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
7-Acetoxyheptanal		1.55	2.19	141.42	10.95	8.56	78.14
n-Acetyl-methionine	1115-47-5	9.85	5.30	53.84	-	-	-
Ambuphylline	5634-34-4	13.25	3.61	27.22	14.85	2.90	19.52
4-Amino-5-methoxy-2-methylbenzenesulfonic acid	6471-78-9	9.80	4.34	44.29	12.17	3.20	26.31
Anisole	100-66-3	3.65	5.16	141.42	18.80	0.42	2.26
B 25		0.00	0.00	-	0.00	0.00	-
n-Butanal	123-72-8	3.95	3.89	98.46	19.20	1.56	8.10
n-Butanol	71-36-3	13.95	6.15	44.10	16.60	5.09	30.67
Butyl carbamate	592-35-8	6.80	5.93	87.21	12.67	1.93	15.27
Caffeine sodium benzoate	8000-95-1	6.37	1.66	26.11	13.10	5.31	40.52
Caffeine sodium salicylate	8002-85-5	8.60	1.70	19.73	17.40	1.98	11.38
Camphen	79-92-5	6.00	5.66	94.28	-	-	-
Cerium-2-ethylhexanoate	24593-34-8	7.40	0.71	9.56	17.18	2.93	17.09
1-Chloroctane-8-ol		5.55	1.77	31.85	16.50	3.11	18.86
3-Cyclohexene-1-methanol	1679-51-2	10.95	1.20	10.98	18.95	0.07	0.37
DC 8		0.00	0.00	-	2.50	3.54	141.42
1,4-Dibutoxybenzene	104-36-9	2.10	2.97	141.42	-	-	-
Diepoxid 126	2386-87-0	5.50	3.38	61.42	10.53	4.82	45.78
2,5-Dimethylhexanediol	110-03-2	6.65	3.61	54.23	13.85	3.89	28.08
3,6-Dimethyloctanol		0.15	0.21	141.42	4.30	0.00	0.00
4,4-Dimethyl-3-oxo-pentanenitrile	59997-51-2	4.95	0.92	18.57	6.20	0.71	11.40
1-(2,6-Dimethylphenoxy)-2-propanone	53012-41-2	7.42	9.99	134.67	11.80	7.60	64.42
Diphocars		14.70	5.09	34.63	15.10	3.96	26.22
1,2-Dodecanediol	1119-87-5	5.48	5.75	104.84	3.20	1.27	39.77
DTPA Pentasodium salt	140-01-2	15.58	0.11	0.73	19.65	0.35	1.80
Ede 140		1.70	2.40	141.42	2.30	3.25	141.42
1,2-Epoxydodecane	2855-19-8	2.05	2.90	141.42	4.95	5.02	101.42
Ethiosan		1.90	2.69	141.42	-	-	-
Ethyl butanal	97-96-1	1.80	2.55	141.42	18.05	0.92	5.09

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
Gadopentetic acid dimeglumine salt	86050-77-3	4.70	2.40	51.15	5.70	3.54	62.03
Genomoll	115-96-8	9.30	0.14	1.52	10.75	1.20	11.18
C12/C14-Glucoside		9.57	1.01	10.57	16.50	0.20	1.21
L-Glutamic acid hydrochloride	138-15-8	12.95	1.77	13.65	13.45	2.47	18.40
Glycediol		0.90	1.27	141.42	2.04	2.06	101.21
Granuform	30525-89-4	1.45	2.05	141.42	0.00	0.00	#DIV/0!
Hexahydrofarnesyl-acetone	502-69-2	1.75	0.78	44.45	6.10	2.69	44.05
Hexamethylenetetramine	100-97-0	5.05	1.06	21.00	11.15	0.07	0.63
1,2,6-Hexanetriol	106-69-4	7.90	5.09	64.45	17.05	2.47	14.52
Hnol		0.40	0.57	141.42	4.05	2.76	68.09
Hoe MBF		0.00	0.00	-	0.18	0.25	141.42
Hydo 98		11.65	1.77	15.17	-	-	-
2-Hydroxyethyl imino disodium acetate	135-37-5	11.15	3.18	28.54	13.25	3.18	24.01
2-Hydroxyisobutyric acid	594-61-6	12.85	2.90	22.56	13.45	3.04	22.61
Hypo 20		3.60	5.09	141.42	6.51	3.38	51.92
Hypo 36		4.10	0.14	3.45	12.95	4.17	32.22
Hypo 45		5.17	5.15	99.62	8.33	3.76	45.16
Hypo 54		4.15	0.21	5.11	4.15	0.07	1.70
Hyton		15.25	2.47	16.23	18.40	0.28	1.54
Iminodiacetic acid	142-73-4	8.25	7.43	90.01	6.85	5.98	87.23
Isobornyl acetate	125-12-2	2.90	1.70	58.52	6.35	2.47	38.97
Isobutanal	78-84-2	1.05	1.48	141.42	19.70	0.42	2.15
Isodecylglucoside		13.55	5.16	38.10	14.35	5.16	35.97
Isononylaldehyde	35127-50-5	0.00	0.00	-	7.25	3.89	53.64
alpha-Ketoglutaric acid	328-50-7	18.95	0.21	1.12	19.75	0.07	0.36
alpha-Lactid	4511-42-6	8.60	6.08	70.66	3.90	2.75	70.55
L-Lysine Monohydrate	39665-12-8	9.13	1.24	13.56	13.65	4.60	33.67
3-Mercapto-1,2,4-triazole	3179-31-5	11.30	9.90	87.61	-	-	-
m-Methoxybenzaldehyde	591-31-1	3.15	1.34	42.65	12.65	1.48	11.74
Methyl acetate	79-20-9	4.35	0.07	1.63	17.95	2.62	14.58
Methylpentynol	77-75-8	13.85	2.19	15.83	16.50	5.09	30.86

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
N-(2-Methylphenyl)-imidodi-carbonimidic diamide	93-69-6	17.40	0.42	2.44	-	-	-
2-Methyl-1-propanol	78-83-1	17.80	0.14	0.79	19.80	0.85	4.29
Methyltriglycol	112-35-6	4.50	0.57	12.57	14.75	3.18	21.57
Methyltriglycol	112-35-6	7.00	5.66	80.81	16.60	5.37	32.37
Napt		3.10	1.70	54.74	8.00	3.25	40.66
Nitro-bis-octylamide		0.85	1.20	141.42	4.05	3.46	85.55
Olak		17.50	1.98	11.31	18.25	1.77	9.69
Ölesulf		16.85	0.07	0.42	19.25	0.49	2.57
Phenylephrine hydrochloride	61-76-7	9.85	1.77	17.95	19.10	1.13	5.92
Phenylthiourea	103-85-5	2.00	2.83	141.42	1.55	2.19	141.42
Phosphonat A		6.70	0.14	2.11	6.80	4.67	68.63
Acefyllin piperazinate	18833-13-1	7.13	9.95	139.49	12.97	3.45	26.63
PO 2		2.15	3.04	141.42	0.15	0.21	141.42
Polyethylene glycol butyl ether	9004-77-7	13.30	3.39	25.52	19.25	0.07	0.37
Polyethylene glycol dimethyl ether	24991-55-7	2.05	2.90	141.42	13.70	8.63	62.97
Polyethylene glycol	25322-68-3	0.50	0.71	141.42	7.15	0.78	10.88
Polyhexamethylene guanidine		10.10	1.27	12.60	15.05	0.64	4.23
Polysolvan	7397-62-8	16.15	0.49	3.06	17.65	2.47	14.02
Potassium cyanate	590-28-3	17.30	2.12	12.26	17.65	2.47	14.02
Potassium hexacyanoferrate II	14459-95-1	16.50	1.84	11.14	11.75	7.71	65.60
Potassium hexacyanoferrate III	13756-66-2	5.23	1.45	27.74	6.08	0.53	8.73
2-Pseudojonon		5.75	4.17	72.56	5.70	2.26	39.70
RK Blau		2.00	2.83	141.42	-	-	-
Sacyclo		1.70	2.40	141.42	3.85	0.78	20.20
Sept		7.00	4.24	60.61	17.85	2.76	15.45
Trimethoxypropylsilane	1067-25-0	3.80	0.14	3.72	9.10	6.51	71.49
Trimethoxyoctylsilane	3069-40-7	5.00	4.10	82.02	9.20	1.13	12.30
Silan 165	29055-11-6	0.35	0.49	141.42	5.65	2.19	38.80
Silan 167	41453-78-5	1.40	1.84	131.32	3.50	1.70	48.49
Silan 253	18784-74-2	3.00	0.00	0.00	12.30	3.39	27.59
Sodium bisulfite	7631-90-5	13.30	0.85	6.38	18.40	2.26	12.30

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
Sodium sulfite	7757-83-7	12.25	1.34	10.97	14.20	2.69	18.92
Sodium cyanate	917-61-3	12.65	3.04	24.04	9.45	1.77	18.71
Sodium disilicate	13870-28-5	20.20	0.71	3.50	17.40	1.13	6.50
Sodium hydrogen sulfate	7681-38-1	17.75	1.48	8.37	18.65	0.78	4.17
Sodium lauryl ether sulfate	3088-31-1	14.10	5.09	36.11	18.45	0.78	4.22
Sodium monochloroacetate	3926-62-3	3.75	5.30	141.42	13.45	3.75	27.86
Sodiumpyrosulfite	7681-57-4	14.87	2.41	16.22	14.60	3.05	20.90
4-((2-Sulfatoethyl)sulfonyl)-aniline	2494-89-5	19.05	1.48	7.79	-	-	-
TA 01946 Alkylsilan		8.80	1.70	19.28	13.10	4.38	33.47
Theophylline sodium acetate	8002-89-9	9.40	5.66	60.18	-	-	-
Tocla		16.30	4.81	29.50	16.95	4.88	28.78
Triisooctylamine	25549-16-0	0.40	0.57	141.42	9.05	7.14	78.91
2,2,3-Trimethyl-3-cyclo-pentene-1-acetaldehyde	4501-58-0	2.60	0.42	16.32	12.20	3.54	28.98
Trioxane	110-88-3	11.33	2.93	25.91	17.90	0.14	0.79
Wessalith Slurry		6.57	4.86	74.00	9.90	8.20	82.85
Xanthinol nicotinate	437-74-1	7.65	5.16	67.48	13.20	5.94	45.00
Mean %CV Value				**60.17**			**35.21**
Median %CV Value				**42.65**			**26.22**
Range %CVs				**0-141.42**			**0-141.42**
Mean %CV Value (Minus Substances Tested in 3 Laboratories)				**58.07**			**34.62**
Median %CV Value (Minus Substances Tested in 3 Laboratories)				**31.85**			**21.57**
Range %CVs (Minus Substances Tested in 3 Laboratories)				**0-141.42**			**0-141.42**

Abbreviations: CV = coefficient of variation; CASRN = Chemical Abstract Service Registry Number.
[1]Italicized substances represent chemicals that were tested in three testing laboratories. Data for these substances were removed to determine their impact on the calculated %CV values for this data set.

Hagino et al. (1999) and Ohno et al. (1999): The Japanese Ministry of Health and Welfare evaluated the HET-CAM test method in five different laboratories as part of a validation effort to assess alternative ocular irritation test method. Nine, 15, and 14 cosmetic ingredients were evaluated in the first, second, and third steps of the validation study, respectively. These studies used the IS(A) analysis method to assess potential irritancy classifications. Average individual laboratory results and standard deviations for tested substances were reported in Hagino et al. (1999). **Appendix F2** provides the average IS(A) values for each testing laboratory for each substance evaluated in this validation effort.

The interlaboratory reproducibility was evaluated by comparing the mean %CV values. The evaluation showed that for the chemicals evaluated, the mean %CV values were: 50.2 for the ten substances evaluated in the first phase of the validation study, 114.0 for the 44 substances evaluated in the second phase of the validation study, and 39.2 for the 42 substances evaluated in the third phase of the validation study. The mean %CV value for all 96 substances (when the three phases were pooled) was 74.6. The investigators proposed that the relatively high %CV was caused by variations of the results of nonirritants, which had low *in vitro* scores. When nonirritants were removed from the analysis, the mean %CV value was 45.8 (n=68).

The average and median %CV for substances classified as GHS Category 1 (UN 2003) for the substances described in Hagino et al. (1999)[6], which described the third validation phase, were 24.4 and 27.0, respectively (see **Table 7-13**)[7]. The average and median %CV for substances classified as EPA Category I (EPA 1996) were 25.86 and 26.43, respectively (see **Table 7-13**).

Table 7-13 %CV Values for Substances Evaluated Using the IS(A) Analysis Method (from Hagino et al. 1999)

Substance[1]	Conc.	GHS Category 1	EPA Category I	Mean IS(A) Score	SD	%CV
Acetic acid	10%	X	X	17.35	1.34	7.73
Stearyltrimethylammonium chloride	10%	X	X	13.60	3.00	22.08
Potassium laurate	10%	X	X	15.32	4.00	26.18
Domiphen bromide	10%	X	X	14.05	3.71	26.43
Butanol	10%	X		9.70	2.69	27.72
di-(2-Ethylhexyl) sodium sulfosuccinate	10%	X	X	9.45	2.62	27.78
Cetyltrimethylammonium bromide	10%	X	X	14.15	4.46	31.55
Lactic acid	100%	X	X	14	5.50	39.26
Mean for Severe Irritants (GHS) (n=8)						**26.09**
Median for Severe Irritants						**27.08**
Range for Severe Irritants						**7.73-39.26**
Mean for Severe Irritants (EPA) (n=7)						**25.86**
Median for Severe Irritants						**26.43**
Range for Severe Irritants						**7.73-39.26**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; EPA = U.S. Environmental Protection Agency (EPA 1996); GHS = Globally Harmonized System (UN 2003).
[1]Substances organized by increasing %CV values.

---

[6] Percent CV values were not determined for the other phases, because average data were not provided in literature references.
[7] Individual laboratory data is available in **Appendix C**.

7.2.4 Additional Analyses of Interlaboratory Reproducibility

7.2.4.1 *Balls et al. (1995)*

The investigators of this study presented interlaboratory correlation coefficients between each pair wise combination of laboratories that were involved in the testing phase of the validation study. For example, interlaboratory correlation coefficients of the *in vitro* data for all the tested substances were developed for Laboratory A when compared to Laboratory B, C, D, and E. Summary of the interlaboratory correlation coefficients calculated in this analysis are provided in **Table 7-14** (see **Appendix G** for all correlation coefficients derived from comparing each laboratory with every other laboratory).

Table 7-14 Interlaboratory Correlation Coefficients in Balls et al. (1995)

Index Score	Interlaboratory Pearson's Correlation (r) of the *In Vitro* Data
Full set of test substances (11-49 depending on endpoint)	
HET-CAM Q-Score	0.473-0.790
HET-CAM S-Score	-0.171-0.808
HET-CAM Q-Score, with cut-off at 2	0.449-0.814
HET-CAM S-Score, with cut-off at 2	-0.316- -0.043
Chemicals soluble in water (5-25 depending on endpoint)	
HET-CAM Q-Score	0.355-0.711
HET-CAM S-Score	0.420-0.949
HET-CAM Q-Score, with cut-off at 2	0.470-0.927
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Chemicals insoluble in water (4-12 depending on endpoint)	
HET-CAM Q-Score	0.580-0.944
HET-CAM S-Score	-0.910-0.852
HET-CAM Q-Score, with cut-off at 2	0.562-0.816
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Surfactants (12)	
HET-CAM Q-Score	0.438-0.876
HET-CAM S-Score	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.420-0.966
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Solids (7-17 depending on endpoint)	
HET-CAM Q-Score	0.500
HET-CAM S-Score	-0.171-0.808
HET-CAM Q-Score, with cut-off at 2	0.985
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Solutions (14 depending on endpoint)	
HET-CAM Q-Score	0.712-0.880
HET-CAM S-Score	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.590-0.974
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Liquids (26)	
HET-CAM Q-Score	0.221-0.755
HET-CAM S-Score	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.591-0.771
HET-CAM S-Score, with cut-off at 2	Not Evaluated

For some of the endpoints, the range in correlation coefficients was rather large (e.g., correlation coefficients for chemicals insoluble in water ranged for the HET-CAM S-score endpoint from -0.910 to 0.852). There also were a large number of negative correlation coefficients noted. Review of the results did not indicate that there was one specific laboratory that yielded consistently high or low correlation coefficients.

7.2.4.2 *Blein et al. (1991)*

The investigators assessed the intralaboratory reproducibility with four substances (propylene glycol, Tween 20, SDS, and benzalkonium chloride). There was no rationale provided for the selection of these substances. The report indicated that the reproducibility of results for each substance was good within each laboratory (data not provided). Interlaboratory reproducibility evaluations were conducted with the same four substances and results with diluted and undiluted substances were examined. This analysis indicated that there were no significant differences (p = 0.055) in HET-CAM scores between the laboratories when diluted products were evaluated. However, there was a significant difference (p = 0.01) in HET-CAM scores when undiluted products were evaluated.[8]

7.2.4.3 *Doucet et al. (1999)*

Comparative screening of 40 cosmetic formulations was conducted to assess the usefulness of *in vitro* ocular toxicity test methods, including the HET-CAM test method, when compared to the currently accepted *in vivo* rabbit eye test method. The formulations were classified as skin care products (10), sunscreen products (10), surfactant based products (10), and alcoholic products (10). In this study, the *in vitro* scores (calculated as IS) were calculated as described in Luepke (1985). A substance with an IS value greater than five was defined as an irritant. Comparative *in vivo* results (calculated as MMAS) were calculated from concurrently conducted studies run according to the method described by Draize et al. (1944). A substance with an MMAS value greater than 15 was defined as an irritant.

An intralaboratory evaluation was conducted with one of the surfactant-based products (not identified) classified as an irritant. Twenty HET-CAM experiments with this substance were conducted; a %CV of 10 was obtained.

7.2.4.4 *Spielmann et al. (1991)*

Starting in 1988, a national validation study on two alternative ocular toxicity test methods was initiated by ZEBET. In this report, 27 coded substances that represented a variety of chemical and toxicological properties were evaluated in 12 laboratories to assess intralaboratory and interlaboratory reproducibility of the test method. All but four substances were evaluated at 10% concentrations; these four remaining substances were evaluated at concentrations ranging from 0.5% to 100%. The lowest concentration required to produce a slight reaction on the CAM also was determined. In this report, the *in vitro* scores (calculated as IS) were calculated as described in Kalweit et al. (1990). The irritation classification scheme used in the evaluation was performed according to Luepke (1985). The *in vivo* results (classified into irritation categories per an investigator defined classification

---

[8] In the report, the authors refer to a table (Table 2) that contains the results of the interlaboratory analysis. However, the table is not shown in the report. Therefore, review of the results by NICEATM could not be conducted.

system) were obtained from historical results for studies that were conducted as described by Draize et al. (1944).

For the analysis presented in this report, the classifications for each laboratory for each substance were determined. The irritation classification made by a majority of the laboratories was determined to be the *in vitro* classification call for the substance. The investigators then stated that if 75% of all the laboratories determined a correct classification (i.e., *in vitro* classification was the same as the *in vivo* classification), then the overall call by the testing laboratories was "correct."

A review of the data presented in the report indicates that nine of 12 substances that were classified as a corrosive or severe irritant based on the Draize test result were correctly classified by a majority of the testing laboratories when using the HET-CAM test method. For eight of the correctly identified severe irritants, between 80% and 100% of the testing laboratories classified the test substances as a strong irritant.[9] For the remaining three severe irritants, two were classified as inconclusive and one substance was classified as a negative (nonsevere irritant; i.e., false negative) by 90% of the testing laboratories.

There were 15 substances classified as a nonirritant, slight irritant, or moderate irritant based on a Draize test result. Of these substances, seven were correctly identified as nonsevere irritant substances by at least 75% of the testing laboratories. Of the remaining eight substances, five were identified as false positives and three were classified as inconclusive (i.e., a majority [75%] of the testing laboratories did not classify the test substances as an irritant or nonsevere irritant). The concordance between testing laboratories for the false positives ranged from 75% to 91% of the laboratories.

7.2.4.5 *Spielmann et al. (1993)*

Starting in 1988, a national validation study on two alternative ocular toxicity test methods was initiated by ZEBET. In this second report, 136 coded substances that represented a variety of chemical and toxicological properties were discussed. The substances tested were evaluated at 10% concentrations. The lowest concentration required to produce a slight reaction on the CAM also was determined. The studies were conducted in two laboratories with experience in the test method. In this report, the *in vitro* scores (calculated as IS) were calculated as described by Kalweit et al. (1990). The irritation classification scheme used in the evaluation used both the IS and ITC values. The *in vivo* scores (classified per the EU classification scheme [EU 1992]) were obtained from historical results that were conducted as described by Draize et al. (1944) and were conducted in accordance with GLP standards.

Of the 136 substances tested, 46 were classified as severe irritants (R41) based on *in vivo* studies. Of these 46 R41 substances, both laboratories correctly identified 22 of the substances (48%) as severe irritants. For the remaining 24 substances, 15 were classified as nonirritant or moderately irritant by both laboratories, two were classified as nonirritant or moderately irritant by one laboratory and irritant by the other, four were classified as irritant

[9] For one substances (SDS) it is stated, "since even the low concentration of 1% led to predominately high scores, the substance was classified as a correct positive." (Spielmann et al. 1991)

(R36) by both laboratories, one was classified as an irritant by one laboratory and severely irritant by the other, and two were identified as inconclusive.

The remaining 90 substances were classified as nonsevere irritants and nonirritants based *in vivo* results (10 substances were R36 and 80 were nonirritant). Of these substances, both laboratories classified 65 of the substances (72%) as nonsevere irritants (R36) and nonirritants.

7.3 Historical Positive and Negative Control Data

7.3.1 Historical Positive Control Data

Historical positive control data were obtained from two sources, in response to a request from NICEATM. For one set of data, positive control substances were dimethyl formamide (DMF) and imidazole. Studies were conducted with and without the use of a TSA. For a second set of data SDS and NaOH were used. For the negative control studies a TSA was not used.

7.3.1.1 *Positive Control Studies Using DMF and Imidazole*

Positive control studies were conducted with imidazole and DMF (see **Appendix H1-H2**). With the DMF studies that were conducted with the TSA, the hemorrhage endpoint was evaluated inside and outside the TSA. Of note, the time of development of the hemorrhage endpoint inside the TSA was lower than the time to development of the hemorrhage endpoint outside the TSA (**Table 7-15**). Two proposed reasons for the difference in time to development, according to Dr. Vanparys (submitter of the data), are (1) the vessels outside the TSA may open more easily than those under the TSA, or (2) once the liquid is applied it accumulates around the edge of the TSA rather than between the TSA and CAM.

Table 7-15 Comparison of Means and Standard Deviations for Positive Controls Tested With and Without Test Substance Applicator

Positive Control	N[1]	Hemorrhage[2] (mean ± SD[4])	Lysis[2] (mean ± SD)	Coagulation[2] (mean ± SD)	*In Vitro* Score[3] (mean ± SD)
DMF: With TSA[4]	69	0.02 ± 0.17	6.93 ± 0.03	8.82 ± 15.77	15.77 ± 0.19
DMF: With TSA[4]	10	3.36 ± 0.32	6.54 ± 0.19	8.81 ± 0.04	18.71 ± 0.38
DMF: Without TSA	2	4.00 ± 0.13	6.84 ± 0.05	8.76 ± 0.08	19.60 ± 0.15
Imidazole: Without TSA	15	4.50 ± 0.39	6.84 ± 0.08	8.66 ± 0.17	20.00 ± 0.45

Abbreviations: DMF = dimethylformamide; SD = standard deviation; TSA = test substance applicator (as described in Gilleron et al. 1996, 1997)
[1]N = number of tests.
[2]Mean values of time until development of identified endpoint.
[3]*In Vitro* irritation score calculated as IS(B).
[4]Hemorrhage endpoint in studies described in the first row were evaluated inside the TSA, while hemorrhage endpoint in studies described in the second row were evaluated outside the TSA.

Using the data provided, the intralaboratory reproducibility of the positive controls was evaluated. For the positive control imidazole, the %CV values were calculated for each endpoint as well as for the overall IS(B) score. The range of %CV values was 0.12 to 18.97

for the hemorrhage endpoint, 0.34 to 1.20 for the lysis endpoint, and 0.20 to 2.11 for the coagulation endpoint. The range of %CV values for the overall IS(B) score was 0.12 to 1.58. The average and median %CV values for the overall IS(B) score were 0.97 and 0.50, respectively (**Table 7-16**).

Table 7-16 Intralaboratory Reproducibility Results for Evaluation of Imidazole as a Positive Control

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean (SD)	4.5 (0.39)	6.84 (0.08)	8.66 (0.17)	20.00 (0.45)
Range of %CV	0.12 – 18.97	0.34-1.20	0.20-2.11	0.12-1.58
Overall %CV	8.6	1.10	1.99	2.23
Mean Total Score %CV	0.97			
Median Total Score %CV	0.50			

Abbreviations: %CV = percent coefficient of variation; SD = standard deviation.

For the positive control DMF, the data where hemorrhages develop inside the TSA was evaluated. The range of %CV values was 0.00 to 1.27 for the lysis endpoint and 0.00 to 1.76 for the coagulation endpoint. For the hemorrhage endpoint, a single test produced a result other than zero for the mean and the tested eggs and the standard deviation; the %CV value for the single test was 173.94. The range of %CV values for the overall IS(B) score was 0.04 to 14.07. The average and median %CV values for the overall IS(B) score were 0.59 and 0.29, respectively (**Table 7-17**).

Table 7-17 Intralaboratory Reproducibility Results for Evaluation of DMF as a Positive Control

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean (SD)	0.02 (0.17)	6.93 (0.03)	8.82 (0.09)	15.77 (0.19)
Range[1] of %CV values	173.94[1]	0.00-1.27	0.00-1.76	0.04-14.07
Overall %CV	850	0.49	1.05	1.20
Mean Total Score %CV	0.59			
Median Total Score %CV	0.29			

Abbreviations: %CV = percent coefficient of variation; SD = standard deviation/
[1]Range is representative of a single value since CV values for other experiments could not be calculated, since mean and SD values were zero.

7.3.1.2 *Positive Control Studies Using SDS and NaOH*

HET-CAM studies using 1% SDS and 0.1 N NaOH were provided in response to a request from NICEATM. Additional information on these data, as well as an alternative analysis conducted, is provided in **Appendix H3**. Using the mean values determined for these studies, the overall irritation score calculated (according to the method of Kalweit et al. 1987, 1990) for these substances classified them as irritants (**Table 7-18**).

Table 7-18 Means and Standard Deviations of Positive Control Test Substances

Positive Control	Hemorrhage[1] (mean ± SD[2])	Lysis[1] (mean ± SD)	Coagulation[1] (mean ± SD)
1% SDS (n=377)	14.69 ± 5.36	35.18 ± 17.15	---[2]
0.1 N NaOH (n=336)	8.96 ± 4.96	35.60 ± 24.71	48.04 ± 34.56

Abbreviations: NaOH = sodium hydroxide; SD = standard deviation; SDS = sodium dodecyl sulfate.
[1]Mean values of time until development of identified endpoint.
[2]It was indicated that 1% SDS does not produce coagulation in the CAM after application. However, in the studies conducted coagulation was identified in a single study. In these evaluations, the non-existing data was calculated with an arbitrary value of "0." Therefore, the calculation of a mean value for the coagulation endpoint was not meaningful.

7.3.2 Historical Negative Control Data

HET-CAM studies using 0.9% NaCl as a negative control were provided in response to a request from NICEATM. Studies were conducted with and without the use of a TSA (see **Appendix I**). The use of a TSA was described in Gilleron et al. (1996, 1997) (see **Section 2.2.4.3**).

Over 90 tests with 0.9% sodium chloride (NaCl) using the TSA and three tests with 0.9% NaCl without using TSA were provided. As shown in **Table 7-19**, time to development of endpoints and the overall irritation scores calculated were consistent and classified as nonirritants for all tests.

Table 7-19 Comparison of Means and Standard Deviations of 0.9% NaCl With and Without Use of the Test Substance Applicator

0.9% NaCl	N[1]	Hemorrhage[2] (mean ± SD)	Lysis[2] (mean ± SD)	Coagulation[2] (mean ± SD)	*In Vitro* Score[3] (mean ± SD)
With TSA	92	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Without TSA	3	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Abbreviations: NaCl = sodium chloride; SD = standard deviation; TSA = test substance applicator.
[1]N = number of tests
[2]Mean values of time until development of identified endpoint.
[3]*In Vitro* irritation score calculated as IS(B).

7.4 Summary

The analysis of intralaboratory repeatability was evaluated using data from two different publications (Gilleron et al. 1996, 1997) for the IS(B) analysis method. In both studies, the hemorrhage endpoint had a high %CV value (104-117). Additionally, the %CV values for the coagulation endpoint were the lowest of the three endpoints evaluated in the HET-CAM test method. However, the actual values were quite disparate between the two studies (e.g., Gilleron et al. 1996 coagulation %CV = 95.69, Gilleron et al. 1997 coagulation %CV = 41.78). The difference in the numbers may be due to several factors including test substances evaluated and differences in the test method protocols used between the two studies. The calculated variability for the endpoints and the overall test method may be exaggerated because of the relatively small values that are obtained from each of the

endpoints (5 for hemorrhage, 7 for lysis, and 9 for coagulation). Similar results were obtained from the analysis of intralaboratory reproducibility. The overall irritation score was generally reproducible (%CV values of 53 and 17.5 for the two studies evaluated).

A qualitative assessment of the data provided for multiple laboratories in three to four studies indicates the extent of interlaboratory reproducibility. Given the relatively homogeneous performance of the HET-CAM test method among the three classification systems, the discussions for the individual studies and analysis methods encompasses all three hazard classification systems, unless otherwise indicated.

In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the two to four participating laboratories for the Balls et al. (1995) study were in 100% agreement in regard to the ocular irritancy classification for 21 (45%) of the 47 substances analyzed using the Q-Score analysis method. The extent of agreement between testing laboratories for the Q-Score analysis method was greatest for substances correctly identified as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (60% to 71% [9/15 to 10/14] of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories, depending on the classification system). Comparatively, participating laboratories were in 100% agreement for 12 to 13 (66% to 68%) of the 18 to 19 substances analyzed using the S-Score analysis method, depending on the classification system used.

For the IS(B)-10 analysis methods (Spielmann et al. 1996), the participating laboratories were in 100% agreement for 84 to 85 (79% to 81%) of 104 to 107 substances evaluated. The extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM (94% to 97% [31/33 to 32/33]). Comparatively, greater disparity between individual substance classifications, for all hazard classifications, was observed for substances that were identified as false positives (52% to 58% false positive had less than 100% concordance between testing laboratories).

For the IS(B)-100 analysis method, the participating laboratories were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM and greater disparity between individual substance classifications was observed for substances that were identified as false positives.

For the report by Hagino et al. (1999), there was 100% agreement in regard to the GHS ocular irritancy classification for 11 (64% to 69%) of the 16 to 17 substances evaluated in five laboratories. Discordance in the classification results was present for substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants. Substances classified as false positives had the greatest extent of agreement among laboratories.

Sufficient *in vivo* information for the CEC (1991) study was only available to assess the interlaboratory reproducibility performance for the EU classification system. For the CEC

evaluation, the participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 6 (23%) of the 26 substances tested. The extent of agreement among laboratories was greatest for accurately identified EU corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (50% [3/6] of the identified EU corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives and those substances accurately classified as EU nonsevere irritants/nonirritants.

The overall reliability statistics, arranged by HET-CAM data analysis method, for the IS(B), IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-5, 7-6,** and **7-7**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For both of these analysis methods, the addition of the results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Hagino et al. (199) and Spielmann et al. (1996).

Substances with less than complete agreement in the testing laboratories depended upon the analysis method evaluated. For the IS(A) analysis method, chemical classes included amidine, ether, carboxylic acid, amine, and alcohol. For the IS(B)-10 analysis method, the most common chemical classes shown to be overpredicted, and where there were discordant results between testing laboratories, was alcohols. For the Q-Score analysis method, alcohols were shown to produce discordant results between testing laboratories.

A quantitative evaluation of interlaboratory reproducibility was conducted for four studies (CEC 1991; Balls et al. 1995; Spielmann et al. 1996; Hagino et al. 1999) by performing a %CV analysis of *in vitro* scores obtained for substances tested in multiple laboratories. For CEC (1991), two different evaluations were conducted based on the concentration tested *in vitro*. For 14 substances evaluated at 100% concentration, the mean and median %CV values were 31.86 and 33.04, respectively. For 12 substances evaluated at 10% concentration, the mean and median %CV values were 34.6 and 33.1, respectively. For the Balls et al. (1995) study, the average and median %CV values for substances evaluated with the Q-Score were 49.83 and 42.50, respectively. The average and median %CV values for the substances evaluated with the S-Score were 84.42 and 71.90, respectively. For the substances evaluated in Spielmann et al. (1996), the average and median %CV values for substances tested at 10% concentration were 60.17 and 42.65, respectively. For substances tested at 100% concentration in Spielmann et al. (1996), the average and median %CV values were lower: 35.21 and 26.22, respectively. When substances that were tested in three different testing laboratories were removed from the assessment, little change was seen in the mean and median %CV values for both concentrations tested. For Hagino et al. (1999), the average and median %CV for substances classified as GHS Category 1 (UN 2003) were 24.4 and 27.0, respectively. The average and median %CV for substances classified as EPA Category I (EPA [1996]) were 23.86 and 26.0, respectively.

Finally, historical positive and negative control data were provided by two different sources. The negative control substance evaluated was 0.9% NaCl. The positive control substances were DMF, imidazole, 1% SDS, and 0.1 N NaOH. The studies showed that, between

experiments, the results for all control substances were reproducible. Additionally, studies indicated that all control substances consistently produced appropriate responses (e.g., negative control consistently produced a response that would be classified as nonirritant and positive controls consistently produced a response that would be classified as severe irritant).

[This Page Intentionally Left Blank]

8.0 TEST METHOD DATA QUALITY

8.1 Adherence to National and International GLP Guidelines

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

Based on the information provided in the publications, it appears that Gettings et al. (1991, 1994, 1996) and Hagino et al. (1999) conducted the HET-CAM study in compliance with GLP guidelines. It could not be determined, from the publications, whether any of the other HET-CAM studies considered in this BRD were GLP-compliant.

The *in vivo* reference studies used for Balls et al. (1995), Gilleron et al. (1997), and Spielmann et al. (1996) appear to have adhered to GLP guidelines. Balls et al. (1995) and Gilleron et al. (1997) used *in vivo* reference data from the ECETOC Eye Irritation Reference Data Bank (ECETOC 1992). These *in vivo* data were generated in GLP-compliant studies conducted according to OECD TG 405 (OECD 1987). Spielmann et al. (1996) used data obtained from German pharmaceutical and chemical companies. The *in vivo* data used in the evaluation were high-quality data that were carried out according to OECD TG 405 (OECD 1987). Additionally, Spielmann et al. (1996) noted that some chemicals were not used in the evaluation because the *in vivo* studies were not conducted according to GLP guidelines.

The coding procedures used in the studies considered in this BRD were evaluated only by the information provided in the published reports. No attempt was made to obtain original study records to assess these procedures. Based on the available information, the only reports that identified using coded chemicals were Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999).

8.2 Data Quality Audits

Formal assessments of data quality, such as a quality assurance (QA) audit, generally involve a systematic and critical comparison of the data provided in a study report to the laboratory records generated for a study. No attempt was made to formally assess the quality of the *in vitro* HET-CAM data included in this BRD or to obtain information about data quality audits from the authors of the HET-CAM study reports. The published data on the HET-CAM assay were limited to calculated *in vitro* scores and/or irritancy classifications. Data provided in response to two *FR* notices requesting data included average Q-Scores and S-Scores for each testing laboratory involved in a validation study (Balls et al. 1995), individual endpoint scores for each egg for each tested substance (Gilleron et al. 1996, 1997), and IS and ITC values for tested substances (Spielmann et al. 1996). Auditing these reported values would require obtaining the original data for each HET-CAM experiment, which were not obtained.

An informal assessment of the HET-CAM study reports revealed limitations that complicate interpretation of the HET-CAM data:

- *Incomplete substance information:* Some HET-CAM study reports provided limited information about the substances tested. The CASRN, purity, and supplier of the test substances were not consistently reported. Thus, comparisons of data from different studies that evaluated test substances of the same chemical name must be interpreted with caution because of possible differences in substance purity.
- *Data reporting:* A majority of the HET-CAM studies reported only the mean *in vitro* score with no accompanying standard deviation to indicate the variability of the data.
- *Methodology*: The methods were presented in varying levels of detail and completeness in the study reports.

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

8.3 Impact of Deviations from GLP Guidelines

The impact of deviations from GLP guidelines cannot be evaluated for the HET-CAM studies reviewed in this BRD, since no information on data quality audits was obtained.

8.4 Availability of Laboratory Notebooks or Other Records

As noted in **Section 5.2**, the availability of notebooks or other original records containing data from the reviewed HET-CAM studies was not determined. Therefore, the testing laboratory's summary judgment regarding the outcome of each study cannot be evaluated.

8.5 Need for Data Quality

Data quality is a critical component of the test method validation process. To ensure data quality, ICCVAM recommends that all of the data supporting validation of a test method be available with the detailed protocol under which the data were produced. Original data should be available for examination, as should supporting documentation, such as laboratory notebooks. Ideally, the data should adhere to national or international GLP guidelines (ICCVAM 2003).

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Reports in the Peer Reviewed Literature

In addition to the reports discussed in previous sections (**Sections 6.0** and **7.0**), additional HET-CAM study reports were identified during the literature review. In many of these reports, inadequate information on the substances tested (e.g., identity not specific) and/or the results obtained from the *in vitro* or *in vivo* studies (e.g., qualitative but not quantitative HET-CAM data, group mean but not individual *in vivo* animal scores) precluded their use in an assessment of the performance characteristics of HET-CAM as described in **Sections 6.0** and **7.0**. This section provides a summary of reports where sufficient information was not available to include them in the performance assessment as well as the summary conclusions of the reports used for the analyses described in **Sections 6.0** and **7.0**. In addition, where applicable, an explanation as why some data could or could not be used as part of the performance evaluation is provided.

9.1.1 Bagley et al. (1992)

Investigators from five chemical and pharmaceutical companies conducted an evaluation of five alternative ocular toxicity test methods, which had been used by these companies in a tiered-testing approach to evaluate eye irritation potential. The study evaluated 12 chemicals and 20 formulations (components of the formulations were not provided). In this study, the *in vitro* scores were calculated as IS(A) values. Comparative *in vivo* rabbit eye test results were obtained from concurrent studies conducted in accordance with the method described Draize et al. (1944), and *in vivo* test data was presented as MAS.

The correlation analyses described in the study compared IS(A) values with MAS values. This correlation yielded a Pearson's coefficient of 0.77 and Spearman's coefficient of 0.85. No additional analyses on the performance of the test method were provided.

Individual rabbit *in vivo* data was obtained for a subset of substances evaluated in this study. These data were used to assess the performance of the HET-CAM test method for detecting ocular corrosives and severe irritants based on the GHS (UN 2003), EPA (1996), or EU (2001) classification systems. The results of these analyses for the subset of substances are provided in **Section 6.0**.

9.1.2 Balls et al. (1995)

Under the auspices of the British Home Office and Directorate General XI of the European Commission, a validation study on proposed alternatives to the *in vivo* rabbit ocular toxicity test method was conducted. The goal of the evaluation was to identify at least one non-whole animal test method that could be proposed to regulatory authorities as a replacement for the currently accepted *in vivo* ocular toxicity test method. For the HET-CAM test method, a total of 52 substances were evaluated in 60 tests in two to four laboratories. Four test substances were evaluated at two different concentrations and two substances were evaluated at three different concentrations. The ocular irritancy potential of the test substances were ranked in terms of MMAS (which ranged from 0 to 108). The test substances evaluated in the validation study were classified as acids (4), acyl halide (1),

alcohols (9), aldehyde (1), alkali (1), esters (6), heterocyclics (3), hydrocarbons (2), inorganic chemicals (4), ketones (3), organophate (1), pesticides (5), surfactants (6), and miscellaneous (6). In this study, the *in vitro* scores were calculated as Q-Scores and S-Scores. *In vivo* data for 46 of the test substances, which were generated in compliance with OECD TG 405, were obtained from historical sources. *In vivo* rabbit eye data for 14 of the test substances were obtained from concurrent studies conducted in compliance with OECD TG 405.

The authors concluded that the correlations between HET-CAM *in vitro* and *in vivo* scores were generally poor to moderate, regardless of the physicochemical properties of the substances tested. A summary of the range of Pearson's and Spearman's correlation coefficients obtained in this study for the full set of substances as well as various subgroups are provided in **Table 9-1**.

Since the *in vivo* test results were expressed as MMAS, the data provided in this report could not be used to evaluate the accuracy of HET-CAM for detecting ocular corrosives and severe irritants based on the GHS (UN 2003), EPA (1996), or EU (2001) classification systems. However, in response to a request from NICEATM, ECVAM forwarded mean HET-CAM scores from each testing laboratory. Raw *in vivo* data were obtained from ECETOC (ECETOC 1998). These data were used in the performance assessment of the HET-CAM test method described in **Section 6.0** and **Section 7.0**.

9.1.3 Blein et al. (1991)

A multicenter study of alternative ocular toxicity test methods was conducted under Oeuvre Pour l'Assistance aux Animaux de Laboratoire (OPAL). The study evaluated 40 substances representing different chemical categories and ocular irritancies. In this study, the *in vitro* scores were calculated as IS(A) values. Comparative *in vivo* rabbit eye test results were obtained from concurrent studies conducted in accordance with Draize et al. (1944). The *in vivo* scores were segregated into three different irritancy classifications (mild, moderate, and extreme); the rationale for the *in vivo* decision criteria was not provided.

The investigators reported that the HET-CAM test method overpredicted the irritancy potential of test substances when they were tested undiluted, while *in vitro* studies conducted with 10-fold dilutions provided a better correlation with the *in vivo* rabbit ocular test results. Using a 10-fold dilution, the irritancy potentials of two substances (acetone and formaldehyde) were underestimated when compared to the *in vivo* classification.

HET-CAM data in this report were presented in graphical form and no attempt was made to extrapolate the graphically presented data to mean HET-CAM scores. Thus, the test substances could not be classified according to the classification system described in **Section 5.0** and were not used in the accuracy analysis described in **Section 6.0**.

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

Index Score	Pearson's Correlation	Spearman's Correlation
Full set of test substances (11-49 depending on endpoint)		
HET-CAM Q-Score	0.310-0.517	0.441-0.596
HET-CAM S-Score	0.060-0.332	0.018-0.340
HET-CAM Q-Score, with cut-off at 2	0.416-0.527	0.462-0.588
HET-CAM S-Score, with cut-off at 2	0.089-0.320	0.069-0.329
Chemicals soluble in water (5-25 depending on endpoint)		
HET-CAM Q-Score	0.314-0.758	0.327-0.681
HET-CAM S-Score	0.137-0.309	0.082-0.357
HET-CAM Q-Score, with cut-off at 2	0.185-0.364	0.309-0.480
HET-CAM S-Score, with cut-off at 2	Not Evaluated	Not Evaluated
Chemicals insoluble in water (4-12 depending on endpoint)		
HET-CAM Q-Score	0.232-0.445	0.345-0.688
HET-CAM S-Score	-0.922-0.716	-0.971-0.738
HET-CAM Q-Score, with cut-off at 2	0.370-0.609	0.396-0.651
HET-CAM S-Score, with cut-off at 2	Not Evaluated	Not Evaluated
Surfactants (12)		
HET-CAM Q-Score	0.448-0.847	0.596-0.839
HET-CAM S-Score	Not Evaluated	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.476-0.701	0.570-0.780
HET-CAM S-Score, with cut-off at 2	Not Evaluated	Not Evaluated
Solids (7-17 depending on endpoint)		
HET-CAM Q-Score	0.578-0.808	0.694-0.875
HET-CAM S-Score	0.060-0.332	-0.009-0.326
HET-CAM Q-Score, with cut-off at 2	0.458-0.694	0.512-0.816
HET-CAM S-Score, with cut-off at 2	Not Evaluated	Not Evaluated
Solutions (14 depending on endpoint)		
HET-CAM Q-Score	0.125-0.678	0.268-0.658
HET-CAM S-Score	Not Evaluated	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.238-0.483	0.292-0.493
HET-CAM S-Score, with cut-off at 2	Not Evaluated	Not Evaluated
Liquids (26)		
HET-CAM Q-Score	0.328-0.481	0.489-0.616
HET-CAM S-Score	Not Evaluated	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.502-0.550	0.546-0.625
HET-CAM S-Score, with cut-off at 2	Not Evaluated	Not Evaluated

9.1.4 Brantner et al. (2002)

The investigators evaluated compounds and plant extracts for anti-inflammatory properties using the HET-CAM test method. Eggs were initially incubated three days. Then a small hole was drilled into the eggshell, 10 mL of the egg white was removed, and then the hole was sealed. On the opposite side of the egg, the shell was opened with forceps and then covered with parafilm. The egg was then re-incubated for another three days. At that time,

the eggs were treated with SDS to induce an irritant response on the CAM. The investigators evaluated the anti-inflammatory properties of eight steroidal and non-steroidal substances. The ability of the test substances to reduce inflammation was calculated by determining the percent reduction in SDS-induced inflammation of the treated samples. *In vivo* inflammatory and anti-inflammatory responses were determined using the Croton oil test. The investigators indicate that the HET-CAM test method was more sensitive than the *in vivo* test method in determining anti-inflammatory activity of the test substances. However, it is noted that the *in vivo* test method was able to provide dose-response correlations in the substances evaluated, while the HET-CAM test method could not provide clear correlations.

The data from this study was not used in an analysis of HET-CAM test method accuracy, because the response being evaluated was not irritation potential, but anti-inflammatory responses.

9.1.5 Brantom et al. (1997) and Steiling et al. (1999)

Under the auspices of the European Cosmetic, Toiletry, and Perfumery Association (COLIPA), a validation study on alternatives to the *in vivo* ocular toxicity test method was conducted. Using 23 substances that represented cosmetic ingredients (selected from the ECETOC database; ECETOC 1992) and 32 finished products, the validation status of several alternative test methods were evaluated. In this study, the *in vitro* HET-CAM scores (calculated as Q-Score or S-Score) were separated into four different irritancy classifications (slightly, moderately, irritating, and severely). *In vivo* rabbit eye scores were segregated into the same four irritant classes. Comparative *in vivo* rabbit eye test results were obtained from historical sources or concurrent studies conducted in accordance with OECD TG 405. MMAS values were provided for a subset of the tested substances.

Accuracy and interlaboratory reproducibility between *in vivo* classification and *in vitro* classification was determined by the statistic κ[1]. The study indicated that the HET-CAM test method classifications did not accurately predict the *in vivo* classification categories (κ values from 0.268 to 0.541 and κ_Q values from 0.428 to 0.731). The interlaboratory reproducibility (for four laboratories) ranged from 0.342 to 0.607. Analysis indicated that the interlaboratory reproducibility of the test method appeared to be moderately good at the extreme ranges of irritancy (Q-Score of less than 0.8 or greater than 2.0) but was a poor predictor of irritancy of substances with a Q-Score in the middle range (between 0.8 and 2.0).

The data could not be used in the accuracy analysis because individual sample or mean sample *in vitro* scores were not provided in the report. Thus, the test substances could not be classified according to the classification system described in **Section 5.0** and were not used in the accuracy analysis described in **Section 6.0**.

[1] The statistic κ value can either weight all factors equally or use different weightings. For the analysis, three versions of the κ statistic were used: (1) equal weighting for all factors (κ), (2) linear weighting where greater weight was given to the effect of disagreements of more that two classification categories (κ_L), and (3) quadratic weighting where very high weighting was given to the effect of disagreements of more than two classification categories (κ_Q).

9.1.6 Budai et al. (1997) and Budai and Várnagy (2000)

Comparative screening of six pesticides at three concentrations (1%, 10%, and 100%) was conducted to assess the usefulness of the HET-CAM test method when compared to the *in vivo* rabbit eye test method. The *in vitro* scores (calculated as IS(A)) were separated into four different irritancy classifications (no, weak, moderate, severe). Comparative *in vivo* rabbit eye test results were obtained from concurrently run studies conducted in accordance with OECD TG 405. The *in vivo* scores (reported as MAS) were separated into four different irritancy classifications (no or slight, moderate, severe, super). The rationale for the *in vivo* categories was not provided.

The reports indicate that the HET-CAM results showed good correlation to the *in vivo* results. Of the four test substances tested *in vitro* and *in vivo*, three substances were classified in similar categories by both test methods. One test substance was overclassified by the HET-CAM test method.

The data from this study could not be used in the accuracy analysis because individual sample or mean sample *in vitro* scores were not provided in the report. Therefore, the tested substances could not be classified according to the classification system described in **Section 5.0**.

9.1.7 CEC (1991)

A collaborative study on alternative methods to the *in vivo* rabbit eye test was commissioned by the Division Control of Chemicals, Industrial Risks and Biotechnologies of Directorate General Environment, Nuclear Safety, Civil Protection and the Health and Safety Directorate of Directorate General Employment Industrial Relations and Social Affairs. *In vitro* IS values were calculated according to the method of Kalweit et al. (1987) (IS[B] analysis method). A score of greater than 9 was defined as a severe irritant. *In vivo* data were classified according to the EU classification system based on chemical profiles developed for the evaluation.

The authors indicate that the HET-CAM test method performed well in identification of severe irritants (R41 classified substances). However, nonirritants were overclassified. The authors suggest that an improved evaluation may be obtained if dilutions of the test substances were evaluated.

A subset of the data present in this study was used in the BRD. For the accuracy and reliability evaluations described in **Section 6.0** and **Section 7.0**, substances where no *in vivo* rabbit studies were used in the irritancy classification were excluded from consideration.

9.1.8 Dannhardt et al. (1996)

The investigators evaluated whether the HET-CAM test method could be used as a screen for nonsteroidal anti-inflammatory agents. A unique test method protocol was used in which the eggs were incubated for nine days. A small hole was then drilled into the eggshell and the test substance was placed on the CAM using a syringe. The hole was sealed with cement. The eggs were then incubated for 2, 4, or 6 hours. After the incubation period, the eggshell was opened and SDS was placed on the CAM. The time of the start of the irritation response

was then noted. The time of the start of the irritation response of eggs treated with test substance was compared to those treated with negative controls and the relative delay of onset, if any, was determined. The investigators indicate that the method allows the classification of the substances according to their potency; however, correlation with the results of *in vivo* studies was limited.

The data from this study was not be used in an analysis of HET-CAM test method accuracy, because the response being evaluated was not irritation potential, but anti-inflammatory responses.

9.1.9 Demirci et al. (2003)

The investigators evaluated substances isolated from essential oils from aerial parts of *P. linearis* for antiangiogenic and anti-inflammatory properties. The test method used in this evaluation comprised forming an agarose pellet with the test substance and applying the pellet to the CAM surface. The severity of the observed effect was scored on a scale from 0.5 to 1. No comparative *in vivo* studies were conducted in this evaluation. The evaluation showed that application of the agarose pellets containing the test substance was not toxic and did not produce irritant effects.

The data from this study was not be used in an analysis of HET-CAM test method accuracy, because the response being evaluated was not irritation potential, but anti-inflammatory responses.

9.1.10 Demirci et al. (2004)

The investigators evaluated substances isolated from the essential oils from aerial parts of *Origanum onites* L for antiangiogenic and anti-inflammatory properties of the isolated substances. The test method used in this evaluation comprised forming an agarose pellet with the test substance and applying the pellet to the CAM surface. The severity of the observed effect was scored on a scale from 0.5 to 1. No comparative *in vivo* studies were conducted in this evaluation. The evaluation showed that application of the agarose pellets containing the test substance was not toxic and did not produce irritant effects.

The data from this study was not be used in an analysis of HET-CAM test method accuracy because the response being evaluated was not irritation potential, but anti-inflammatory responses.

9.1.11 de Silva et al. (1992)

The investigators evaluated 60 chemicals and 41 cosmetic formulations; the chemicals and components of the formulations tested were not provided in the report. The commercial products classes of the formulations were oils, make-up removal, emulsions, gels, shampoos, and creams and body milk. The chemicals tested were evaluated at 1% and 10% concentrations, while the formulations were tested neat. Of the 41 formulations tested, 20 were rinsed off the CAM 20-seconds after application because they were opaque or colored. In this study, *in vitro* scores were calculated as IS(A) values and classified as described in

Luepke (1985). The comparative *in vivo* rabbit eye scores (calculated as MAS and classified per the EEC classification scheme[2]) were obtained from published results.

The studies showed that interlaboratory reproducibility was high for test chemicals evaluated at 1% and 10% concentrations. The Spearman's coefficient for both concentrations was greater than 0.9. The results from the *in vitro* analysis were plotted against the EEC categories (tabular data were not provided) and relationship between the two was determined using the Jonchkeere-Terpstra test, followed by calculation of the Spearman's coefficient. This analysis yielded a moderate coefficient of 0.726 ($p < 0.0001$). For these substances, HET-CAM had an accuracy of 90%, a sensitivity of 91%, and a specificity of 88%[3].

The rank correlation between those formulations that were rinsed and those that were not rinsed were compared to determine the effect of protocol differences. The results obtained were compared to the MAS. The Spearman's coefficients were 0.77 for the non-rinsed formulations and 0.76 for the rinsed formulations.

The *in vitro* data from this study could not be used in the HET-CAM accuracy analysis, because sufficient information on the test substances and *in vitro* score were not provided in the report. The lack of information on the test substances did not allow for categorization of the substances into the irritancy categories defined by the GHS (UN 2003) or EPA (1996).

9.1.12 Djabari et al. (2002)

Investigators combined the HET-CAM test method with histological evaluation of the CAM in an attempt to increase the sensitivity of the test method. The test method was conducted and scored as described in Luepke (1985) (IS[A] analysis method). Immediately after CAM scoring was completed, the central part of the CAM was removed, fixed, and stained with trypan blue to evaluate the state of the blood vessels. Twenty water-soluble test substances (identified as active ingredients in cosmetics) were evaluated undiluted and at a 10% dilution. No comparative *in vivo* studies were conducted in this evaluation.

The report indicates that when the diluted forms of the test substances were evaluated there was no discrepancy between the results of the HET-CAM evaluation and histological evaluation of the CAM. At a 10% concentration, all the substances were classified as nonirritant by the HET-CAM method and no morphological changes were observed by histological or trypan-blue evaluation. When the substances were evaluated undiluted, seven of the substances displayed discrepancies between the results of the HET-CAM evaluation and the histological evaluation. For six of the substances, the HET-CAM evaluation indicated that the substances were nonirritants while the histological evaluation indicated that the substances produced irritation. In the last case, the histological evaluation indicated that the substance produced slight hemorrhages while the visual inspection of the CAM indicated the development of hyperemia. The investigators concluded that inclusion of histological

[2] No citation is provided in the study regarding the specific guideline used in classifying substances. However, the study indicates that the classifications used in the analysis (Class I, II, and III) correlate to nonirritant, R36, and R41, respectively (EU 1992, 2001).
[3] Numbers used to calculate these percentages are only provided in graphical form and no attempt was made to count the points.

examination of the CAM after conducting the HET-CAM assay could increase the sensitivity of the method and provide greater information about the effects produced by the test substance.

The data from this study could not be used in a HET-CAM accuracy analysis (**Section 6.0**) because comparative, *in vivo* data for the test substances was not provided in the report and such data were not located.

9.1.13 Doucet et al. (1999)

Comparative screening of 40 cosmetic formulations was conducted to assess the usefulness of *in vitro* ocular toxicity test methods, including the HET-CAM test method, when compared to the currently accepted *in vivo* rabbit eye test method. The formulations were classified as skin care products (10), suncreen products (10), surfactant based products (10), and alcoholic products (10). In this study, the *in vitro* scores were calculated as IS(A) values; value greater than five was defined as an irritant. Comparative *in vivo* results (calculated as MMAS) were calculated from concurrently conducted studies run according to the method described by Draize et al. (1944). A substance with an MMAS value greater than 15 was defined as an irritant. There was no rationale provided for the classification and cut-off values used.

Correlation between the HET-CAM IS(A) values and MMAS values yielded a κ value of 0.58. The linear correlation between these values was statistically significant ($p < 0.001$) and Pearson's coefficient was 0.72. The calculated residual standard deviation, however, was large. Evaluation of accuracy parameters yielded the following values: accuracy: 80%, sensitivity: 100%, specificity: 56%, false positive rate: 44%, false negative rate: 0%. Of the substances that were identified as false positives, four were skin care products and four were sunscreen products.

The data from this study could not be used in an analysis of HET-CAM test method accuracy because the *in vivo* data provided in the report was insufficient to classify the substances according to the GHS (UN 2003), EPA (1996), or EU (2001) classification systems.

9.1.14 Gettings et al. (1991, 1994, 1996) and Lordo et al. (1999)

The CTFA developed an Evaluation of Alternatives Program, with the intent to provide industry with sufficient information on the performance of a series of potential alternatives to the *in vivo* ocular toxicity test method. This effort was a multi-year, multi-phase effort, with different product-types tested in each phase. The evaluation focused on assessing the accuracy of alternative test methods when compared to the FHSA classification system (CPSC 1988).

The initial phase evaluated a set of ten generic hydroalcoholic formulations (Gettings et al. 1991). In this phase, *in vitro* IS values were calculated via two mathematical methods (Bartnik et al. 1987; Kalweit et al. 1987). A substance with an IS value greater than 300 or 10, respectively, was defined as an irritant. The *in vivo* results were expressed as irritants or nonirritants, based on the FHSA regulatory classification system. No *in vivo* scores (e.g., MAS, Draize scores, animal scores) were provided in the report. In this phase of the

evaluation, the HET-CAM test method displayed 100% (5/5) sensitivity and 100% (4/4) specificity when compared to the *in vivo* classification based on the FHSA regulatory classification system (CPSC 1988).

The data from this report were re-evaluated since the *in vivo* data was classified according to the FHSA classification system. Based on additional data obtained from CTFA and the FDA, the ability of the HET-CAM test method to accurately identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996) and EU (2001) classification systems, is described in **Section 6.0**.

The second phase of the evaluation focused on a set of 18 generic oil-water emulsion formulations (Gettings et al. 1994). In this study, *in vitro* IS(A) and IS(B) values were calculated. In this evaluation, a substance with an IS(A) value equal to or greater than 4.8 or an IS(B) value equal to or greater than 5 was defined as an irritant. As in the previous phase, test substances were classified as either irritants or nonirritants according to the FHSA classification system. In this phase, when the *in vitro* data were transformed using the IS(B) analysis method, the sensitivity was 100% (5/5) and the specificity was 85% (11/13). When the *in vitro* results were transformed using the IS(A) analysis method, the sensitivity was 80% (4/5) and the specificity was 77% (10/13).

The data from this report was re-evaluated since the *in vivo* data was classified according to the FHSA classification system. Based on additional data obtained from CTFA, the ability of the HET-CAM test method to accurately identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996) and EU (2001) classification systems, is described in **Section 6.0**.

The third phase of the evaluation focused on a set of 25 generic surfactant-based formulations (Gettings et al. 1996). In this study, *in vitro* IS(A) and IS(B) values were calculated. A substance with an IS value equal to or greater than 5.1 or 4.83, respectively, was defined as an irritant. A ratio of IS to ITC also was evaluated. Substances with an IS/ITC value of equal to or greater than 3.0 was defined as an irritant.

In this evaluation, the formulations were classified as irritants or nonirritants based on each of the models described. Accuracy assessments were then conducted for each model. Using the IS(B) analysis method, the sensitivity was 94% (17/18) and the specificity was 71% (5/7). Using the IS(A) analysis method, the sensitivity was 94% (17/18) and the specificity was 100% (7/7). Using the IS/ITC ratio model, the sensitivity was 100% (18/18) and the specificity was 71% (5/7).

Since the *in vivo* data was classified according to the FHSA classification system, all of the data from this report was re-evaluated. Based on additional data obtained from CTFA, the ability of the HET-CAM test method to accurately identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996) and EU (2001) classification systems, is described in **Section 6.0**.

In the report by Lordo et al. (1999), the investigators evaluated the precision and extent of random variations associated with the regression fits determined with the data described in the Gettings et al. reports. The sources of variation around each of the regression models were evaluated by estimating the components of total variation associated with predicting MAS for each phase of the CTFA evaluation.

From the evaluation, the greatest source of variability associated with predicting the MAS was due to random variations around the prediction models (70% to 90%) for each of the phases. Generally, variability between *in vitro* replicates and variability between MAS replicates contributed only a minor proportion to the total variability associated with the models for the test substances. The authors conclude that the contribution of the latter two variability components could be decreased by increasing the number of replicates performed for each test formulation. However, it would have little impact on the overall precision of the prediction models developed by Gettings et al. (1996).

9.1.15 Gilleron et al. (1996)

This report discusses an alternative test method protocol for the HET-CAM test method. In this method, the investigator used a TSA to confine the test substance to a section of the CAM. The report discusses the evaluation of 46 substances. The *in vitro* scores were calculated as IS(B) scores. A substance with an IS(B) value equal to or greater than 5.0 was defined as an irritant. The *in vivo* results were reported as MAS. Additionally, the irritancy potential of each test substance was classified based on the EU classification system (EU 1992).

The correlation between IS(B) and MAS values was moderate and statistically significant ($r = 0.58$, $p \leq 0.001$). The best correlation was obtained between the total IS(B) value and the *in vivo* conjunctival score ($r = 0.68$, $p \leq 0.001$). Correlation coefficients between *in vivo* and *in vitro* results, based on physical properties of the test substances, also were conducted ($r = 0.72$ for solids; $r = 0.78$ for liquids; and $r = 0.93$ for surfactants).

Accuracy analysis with the test substances indicated that the HET-CAM method (with the use of TSA) exhibited high sensitivity (92.3% [12/13]) but low specificity (54.5% [18/33]) in classifying substances as irritants or nonirritants. The results of an assessment of the accuracy of the test method for solids, liquids, and surfactants are provided in **Table 9-2**.

Table 9-2 Accuracy Statistics for Test Substances Evaluated in Gilleron et al. (1996)

Statistic	Solids	Liquids	Surfactants
Accuracy	88% (15/17)[1]	38% (8/21)	88% (7/8)
Specificity	92% (11/12)	24% (4/17)	75% (3/4)
Sensitivity	80% (4/5)	100% (4/4)	100% (4/4)
False Negative	20% (1/5)	0% (0/4)	0% (0/4)
False Positive	8% (1/12)	76% (13/17)	25% (1/4)

[1]Numbers in parentheses were used to calculate the percentages.

Data obtained from the report were reclassified based on chemical class and properties of interest. The average HET-CAM IS(B) values and EU irritancy classification provided in the report were used in the analyses described in **Section 6.0**. In response to a request from

NICEATM, Drs. Vanparys and Goethem forwarded raw *in vitro* data that was used for the reliability analysis described in **Section 7.0**.

9.1.16 Gilleron et al. (1997)

This report describes a follow up evaluation to the Gilleron et al. (1996) study. In this study, the investigators evaluated 60 substances. The substances included 28 liquids, 20 solids, and 12 surfactants. *In vitro* values were calculated as IS(B) scores, and a substance with an IS value equal to or greater than 5.0 was defined as an irritant. The *in vivo* scores (calculated as MMAS) were calculated from published data. A substance with a MMAS equal to or greater than 15.0 was defined as an irritant.

The total IS(B) and individual HET-CAM endpoints were compared to the total MMAS value. Correlation analyses indicated that no good correlation was observed. The relationship between MMAS and various physicochemical properties (e.g., solids, liquids, surfactants) also was low (r = 0.29 to 0.38)

An accuracy analysis of the data indicated that the HET-CAM method (with the use of TSA) exhibited moderate accuracy (80% [48/60]), high sensitivity (96% [45/47]), and low specificity (23% [3/13]). The results of an assessment of the accuracy of the test method for solids, liquids, and surfactants are provided in **Table 9-3**.

Table 9-3 Accuracy Statistics for Test Substances Evaluated in Gilleron et al. (1997)

Statistic	Solids	Liquids	Surfactants
Accuracy	90% (18/20)[a]	75% (21/28)	75% (9/12)
Specificity	100% (3/3)	0% (0/7)	0% (0/3)
Sensitivity	88% (15/17)	100% (21/21)	100% (9/9)
False Negative	12% (2/17)	0% (0/21)	0% (0/9)
False Positive	0% (0/3)	100% (7/7)	100% (3/3)

[a] Numbers in parentheses were used to calculate the percentages.

Data obtained from the report were reclassified based on chemical class and properties of interest. *In vivo* data for tested substances were obtained from ECETOC (1998). These *in vitro* and *in vivo* data were then used in the analyses described in **Section 6.0**. In response to a request from NICEATM, Drs. Vanparys and Goethem forwarded raw *in vitro* data that was used for the reliability analysis described in **Section 7.0**.

9.1.17 Hagino et al. (1991)

Investigators conducted a comparative screening of 12 surfactants (evaluated as 10% aqueous solutions) to assess the usefulness of the HET-CAM test method, when compared to the *in vivo* rabbit eye test method. The surfactants were classified as cationic (3), anionic (5), nonionic (2), and amphoteric (2). In this study, *in vitro* scores were calculated as IS(A) values. The *in vivo* rabbit eye study scores (presented as the maximum total Draize score) were calculated from concurrently run studies conducted according to the method described by Draize et al. (1944). The results from this study indicated that there was good correlation between the IS(A) value and the maximum total Draize score (r = 0.86).

The data in the report was presented in graphical form and no attempt was made to extrapolate the points to estimate mean HET-CAM IS(A) values. Since the *in vitro* scores were not provided, the tested substances could not be classified according to the classification system described in **Section 5.0**.

9.1.18 Hagino et al. (1993)

In this evaluation, the investigators compared the HET-CAM results of 12 substances to the *in vivo* rabbit eye test method. The 12 substances comprised a variety of physical forms (liquids, powders, and emulsions) and solubilities (seven of the 12 substances were not soluble in water). All but two substances were tested undiluted. No rationale was provided in the report as to the selection of the test substances, the number of substances tested, or the concentration tested. *In vitro* scores were calculated as IS(A) values. The *in vivo* scores (presented as the MAS) were obtained from published studies that used techniques that were similar to the method described by Draize et al. (1944).

There was good correlation between the IS(A) value and the maximum total Draize score (r = 0.90). Increasing concentrations of a test substance (ethanol) were shown to produce increased response in HET-CAM, suggesting that the method could assess dose response relationships. The data for this evaluation were combined with the data from a previous evaluation (Hagino et al. 1991) and then separated by solubility (water soluble and non-water soluble). The responses in the HET-CAM test method for these two classes were relatively similar. Overall, the correlation coefficient for all test substances was 0.80.

The data in the report was presented in graphical form and no attempt was made to extrapolate the points to estimate mean HET-CAM IS(A) values. Since the *in vitro* scores were not provided, the tested substances could not be classified according to the classification system described in **Section 5.0**.

9.1.19 Hagino et al. (1999) and Ohno et al. (1999)

Two types of CAM assays, HET-CAM and chorioallantoic membrane-trypan blue staining (CAM-TB), were evaluated by investigators as alternative methods to the *in vivo* rabbit eye test method. The validation effort was composed of three phases where a total of 39 test substances were evaluated. The test methods were evaluated in five different laboratories. In this study, *in vitro* scores were calculated as IS(A) values; value equal to or greater than 7.0 was defined as an irritant. The *in vivo* scores (calculated as the maximum total Draize score) were calculated from published studies that were conducted according to the method described by Draize et al. (1944). A substance with an MAS value greater than 15 was defined as an irritant. According to investigators, the *in vitro* cut-off was set arbitrarily based on the distribution pattern of the substances while the *in vivo* cut-off was set according to the classification system defined by Kay and Calandra (1962).

The results showed that HET-CAM correctly identified the irritancy potential of 46 of the 52 test substances. Five chemicals were classified as false positives and one chemical was classified as a false negative. Correlation analysis indicated that the rank correlation coefficient between the HET-CAM IS(A) values and MAS was 0.802. Spearman's rank correlation coefficient also was high for the relationship between the IS(A) value obtained

for each testing laboratory when compared the mean IS(A) value of all of the testing laboratories (0.856 to 0.950).

The data from this report was re-evaluated since the classification system described in the report did not have a severe irritant classification, which is the focus of this evaluation. In response to a request from NICEATM, Dr. Yasuo Ohno of NIHS forwarded raw *in vivo* data. Using this data, the ability of the HET-CAM test method to accurately identify severe irritants was provided in **Section 6.0** and reproducibility results were provided in **Section 7.0**.

9.1.20 Kalweit et al. (1987)

This report describes the results from the first preliminary trial of a national validation project to validate alternative methods to the *in vivo* ocular toxicity test method. During this phase, the HET-CAM test method was established in participating laboratories. In this evaluation, two substances (SDS and triethanolamine) were evaluated in six different laboratories. *In vitro* scores were calculated using the IS(B) analysis method. *In vivo* studies were not conducted for the analysis described in the report.

The report stated that there was close agreement of the results with a high concentration of SDS (1%). Five of the six laboratories classified the test substance as a strong irritant. At lower concentrations of SDS (0.1% to 0.5%), a clear classification of the irritancy potential of the test substance was not possible. Investigators stated that similar results were observed with triethanolamine.

The data from this study could not be used in the accuracy analysis because individual sample or mean IS(B) values were not provided in the report. Therefore, the tested substances could not be classified according to the system described in **Section 5.0**.

9.1.21 Kalweit et al. (1990)

This second report describes additional results from the preliminary phase of a national validation project to validate alternative methods to the *in vivo* ocular toxicity test method. During this phase, two HET-CAM test trials were conducted to test the protocols and software developed for the evaluation. Five substances were evaluated and interlaboratory reproducibility was determined. The substances tested were zinc pyridinethione, 2-butoxyethanol, dimethylsulfoxide, triethanolamine, and SDS. In this study, the *in vitro* scores were calculated as IS(B). *In vivo* studies were not conducted and *in vivo* data was not used in the analysis. The report stated that there were considerable differences between results obtained by trained and less experienced investigators. No additional statistical analyses or evaluations were provided in the report.

The data from this study could not be used in the accuracy analysis because individual sample or mean sample scores were not provided. Since the *in vitro* scores were not provided, the tested substances could not be classified according to the classification system described in **Section 5.0**.

9.1.22 Kojima et al. (1995)

Investigators evaluated seven alternative test methods to the *in vivo* ocular toxicity test method. Twenty-four test substances were evaluated; six substances were cationic surfactants, five substances were anionic surfactants, seven substances were nonionic surfactants, two substances were amphoteric surfactants, and four substances were solvents. *In vitro* HET-CAM test method data were calculated as IS(A) values. Concurrent *in vivo* studies were conducted similar to what was previously described in Draize et al. (1944). In this evaluation, three female rabbits were observed at 1, 3, 6, 24, 48, 72, 96, and 168 hours post application of the test substance. The maximal Draize rabbit eye irritation score (MDES; calculated in a manner similar to the MAS and MMAS) was then calculated. The investigators concluded that there was a moderate relationship between the HET-CAM IS(A) values and MDES (correlation coefficient = 0.824).

The analysis described in this report could not be used directly in an analysis of HET-CAM accuracy because the *in vivo* data was insufficient to classify the substances according to one of the three ocular irritation classification systems used in this analysis. However some of the data from this study was re-analyzed, using historical *in vivo* data from other sources. The results of this re-analysis are provided in **Section 6.0**.

9.1.23 Lawrence et al. (1990)

Investigators conducted comparative screening of 34 substances to assess the usefulness of the HET-CAM test method. The substances ranged from single chemicals to fully formulated products (e.g., shampoos and industrial detergent cleaners). Results from the HET-CAM test method were expressed as *in vitro* irritation classification categories described in Luepke (1985). The *in vivo* results were classified into eight irritancy categories. The data used for the *in vivo* classifications were based on published data from studies that were conducted according to the method described by Draize et al. (1944). The investigators reported that there was not a good correlation between the *in vitro* and *in vivo* results.

The data from this study could not be used in the HET-CAM accuracy analysis, because the *in vivo* data was insufficient to classify the substances according to one of the three classification systems reviewed. The test substances evaluated also were not identified; therefore the use of historical *in vivo* rabbit eye data to conduct an accuracy analysis was not possible.

9.1.24 Lönnroth et al. (1999)

The irritation potentials of eight dental polymer products were tested using the HET-CAM test method. *In vitro* data were evaluated using the IS(B) analysis method. The report did not evaluate the *in vivo* effects of these test substances or correlate *in vitro* results with *in vivo* results. The results showed that the liquid components of all the products had strong irritation potential but the powder suspensions and extracts had no effect.

The data from this study could not be used in the HET-CAM accuracy analysis because *in vivo* data and ocular irritancy classification information, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems, were not provided.

9.1.25 Luepke (1985)

This report provides the initial description of the HET-CAM test method. Chemicals and formulations (vehicles, antimicrobial agents, oxidation dyes, and commercial shampoos) were tested. *In vitro* irritancy classifications of tested substance, not IS values, were provided in the report. The *in vivo* irritancy classifications consisted of four categories; however, information on how the *in vivo* data was collected was not provided.

The author concluded that the HET-CAM test method was capable of demonstrating the mucous membrane irritating potencies of substances. The investigator indicates that the method was useful for screening large numbers of compounds.

The data from this study could not be used in the HET-CAM accuracy analysis because *in vivo* data for the test substances were not available and historical *in vivo* rabbit eye data could not be located.

9.1.26 Luepke and Kemper (1986)

In this study, the investigators evaluated the usefulness of the HET-CAM test method using about 190 substances and formulations. The investigators noted that there was good correlation between the *in vitro* and *in vivo* data and that there was a high level of reproducibility between laboratories.

The data from this study could not be used in the HET-CAM accuracy analysis because *in vivo* rabbit eye data for the test substances were not available. Additionally, the identities of the substances tested were not provided and, therefore, historical *in vivo* rabbit eye data could not be used to conduct an accuracy analysis.

9.1.27 Macián et al (1996)

The investigators report evaluated the toxic effects of a group of synthetic polyoxyethylene nonionic surfactants, which were developed by the investigators. Ocular toxicity potential was evaluated with the HET-CAM test method. In this study, the *in vitro* scores (reported as IS(B)) were calculated using a formula that evaluated the irritancy potential index. *In vivo* rabbit eye studies were not conducted and *in vivo* rabbit eye data were not used in the analysis. The report stated that the test substances were weak to moderate irritants based on the results from the HET-CAM test method.

Since the chemicals evaluated in this study were novel, historical data for the effects of these substances in ocular irritation tests could not be obtained. The lack of comparative *in vivo* data precluded the use of this study in the HET-CAM accuracy analysis.

9.1.28 Reinhardt et al. (1987)

The investigators conducted a comparative screening of 24 surfactants to assess the usefulness of the HET-CAM test method, when compared to the *in vivo* guinea pig eye test method. The selected surfactants all induced a similar range of *in vivo* eye irritation, which was defined as slight. All test materials were tested as a concentration of 300 mM or 10% mixtures. In this study, the *in vitro* scores were calculated as IS(A) values and classified as described Luepke (1985). The *in vivo* scores (reported as the maximum total Draize score

over a 24 hour period) were calculated in concurrent studies on guinea pigs. Eye irritation scoring was analogous to the rabbit *in vivo* eye ocular toxicity test method (Draize et al. 1944) and irritation severity was assessed according to Kay and Calandra (1962). The results showed that the HET-CAM test method was poor in predicting the eye irritation potential of anionic surfactants; the method overpredicted the severity of irritation produced by these test substances.

The data from this study could not be used in the HET-CAM accuracy analysis because comparative *in vivo* data were not provided in the report.

9.1.29 Rougier et al. (1992)

The investigators conducted comparative screening of 41 surfactants and surfactant based formulations to assess the usefulness of several *in vitro* ocular toxicity test methods, including HET-CAM. *In vitro* results were calculated as IS(A) values. The *in vivo* scores (reported as MAS) were based on published data. Spearman Rank correlations were calculated across various data sets and *in vitro* and *in vivo* endpoints.

The analyses showed a high degree of correlation between the HET-CAM hemorrhage score and the MAS value for surfactants and surfactant-based formulations ($r_s = 0.98$ and 0.95, respectively). The overall rank correlation coefficient for all 41 substances was 0.96.

The data from this study could not be used in the HET-CAM accuracy analysis because the *in vivo* data was insufficient to classify the substances according to the GHS (UN 2003), EPA (1996), or EU (2001) classification system, and historical *in vivo* rabbit eye data for the substances tested could not be located.

9.1.30 Schlage et al. (1999)

The investigators evaluated the use of the HET-CAM test method to determine the irritant potential of cigarette mainstream and sidestream smoke. In this study, *in vitro* IS values were calculated and classified as described in Kalweit (1985). *In vivo* results were not evaluated for this analysis. The data from this study could not be used in the HET-CAM accuracy analysis because historical *in vivo* data were not located.

9.1.31 Spielmann et al. (1991, 1993, 1996)

Starting in 1988, a national validation study on two alternative ocular toxicity test methods was initiated by ZEBET. Spielmann et al. (1991) described the interlaboratory assessment and the database development. In this phase, 32 coded substances that represented a variety of chemical and toxicological properties were evaluated in 12 laboratories to assess interlaboratory reproducibility of HET-CAM. All but four test substances were evaluated at 10% concentrations; the four remaining substances were evaluated at concentrations ranging from 0.5% to 100%. Additionally, the lowest concentration required to produce a slight reaction on the CAM was determined. These studies were conducted in two laboratories with experience in the test method. In this study, the *in vitro* scores were calculated as IS(B) values. The irritation classification scheme used in the evaluation was performed according to Luepke (1985). The *in vivo* results, expressed as irritation classification categories (e.g.,

slight, moderate, severe) developed by the investigators, were obtained from published studies that were conducted as described by Draize et al. (1944).

The results indicated that of the 27 test substances that were evaluated by the HET-CAM test method, 16 (59%) were classified correctly (nine positives and seven negative) by 75% of the testing laboratories. There were five false positives and one false negative result. In addition, five of the test substances evaluated did not yield the same classification in at least 75% of the testing laboratories and thus could not be classified.

In the second phase of the national validation study conducted by ZEBET, 136 coded substances that represented a variety of chemical and toxicological properties were discussed (Spielmann et al. 1993). The substances tested were evaluated at 10% concentration. Additionally, the lowest concentration required to produce a slight reaction on the CAM also was determined. The studies were conducted in two laboratories (of seven possible laboratories) with experience using the test method. In this study the *in vitro* scores were calculated as IS(B) values. The mean value of three eggs was used for each test substance. The irritation classification scheme used in the evaluation used both the IS and ITC values. The *in vivo* scores (classified per the EU classification scheme [EU 1992]) were obtained from published studies that were conducted as described by Draize et al. (1944) in compliance with GLP guidelines.

Of the 136 substances tested, 46 were classified as severe irritants (R41) based on *in vivo* studies. Of these 46 substances, both test laboratories correctly identified 22 substances as R41. A majority of the remaining substances (15) were classified as nonirritant or moderately irritant by both test laboratories. Correct identification of the nonirritants was 80%, while identification of R36 labeled chemicals was 10%. The authors indicate that the HET-CAM test method could be incorporated into the OCEG TG 405 testing scheme and be used to reduce the suffering associated with the evaluation of ocular corrosives and severe irritants.

In the third phase of the national validation study conducted by ZEBET, 200 coded substances that represented a variety of chemical and toxicological properties were discussed (Spielmann et al. 1996). The chemicals tested were evaluated undiluted and at a 10% concentration. Additionally, the lowest concentration required to produce a slight reaction on the CAM also was determined. The studies were conducted in two laboratories (of seven possible laboratories) with experience in the test method. The *in vitro* scores were calculated as IS(B) values. The irritation classification scheme used in the evaluation considered both IS and ITC values. The *in vivo* results (expressed as irritation severity categories defined by the EU classification system [EU1992]) were obtained from published results and unpublished results provided by chemical and pharmaceutical companies; the studies were conducted as described by Draize et al. (1944) and in compliance with GLP guidelines.

Of the 200 substances tested, 118 were used in the evaluation of the ability of the HET-CAM test method to identify severe irritants (R41). An assessment of the accuracy statistics of the

test method, based on these 118 chemicals, was conducted by the authors who reported a sensitivity of 41% (19/45), specificity of 89% (65/73), and false positive rate of 7%[4].

Additional endpoints were derived from the calculated scores to conduct discriminant analyses. These analyses showed that use of the mtc10 endpoint (mean detection time for appearance of coagulation when using a 10% solution) correlated better with severe irritants than any other evaluated endpoints (sensitivity: 52.1% [25/48], specificity: 88.3% [83/84], false positive rate: 7.8%[5]). The power of this endpoint to discriminate between R41 and non-R41 chemicals was 10 times higher than that of the next best endpoint (mtc100 [mean detection time for appearance of coagulation when using 100%]). The authors note that the mtc10 endpoint was better suited to identifying R41 irritants than the original prediction model (using IS and ITC values).

The authors additionally proposed several sequential testing strategies to classify ocular corrosives and severe irritants. According to the authors, the best discrimination of R41 substances occurred when the solubility of the substance was determined in water and/or oil. Based on the level of solubility, one of three different procedures could be followed. The three procedures (described in the report) combined endpoints from the HET-CAM test method (mtc10 and/or mtc100) with endpoints of the Neutral Red Uptake test method. Additional details regarding the procedures are provided in the report.

Based on a request from NICEATM, *in vivo* and *in vitro* data were obtained from the authors. The obtained data were re-evaluated using the classification rules described in **Section 4.0** and **Section 5.0**. These data were then used in the analyses described in **Section 6.0** and **Section 7.0**.

9.1.32 Spielmann et al. (1997)

This report describes a retrospective study of the HET-CAM test method that was conducted by the U.S. Interagency Regulatory Alternatives Group (IRAG). In response to a request by IRAG to the scientific community, five sets of data using three different test method protocols were submitted. The substances represented a broad spectrum of industrial chemicals. Overall information about the solubility of the test substances, the pH ranges, chemical classes, and physical form were provided for each set of submitted data. Individual *in vitro* scores were not provided in the report. The *in vivo*/*in vitro* correlation between HET-CAM scores and five *in vivo* endpoints (cornea/opacity, iris, erythema, chemosis, discharge) were calculated and reported. In this evaluation *in vitro* scores were compared to the non-weighted mean of modified maximum individual score (∑MMMIS). To assess *in vitro*/*in vivo* correlations between different *in vitro* endpoints and *in vivo* scores, Pearson's single and Pearson's partial linear regressions were calculated.

Based on Pearson's correlation coefficients, the HET-CAM scores were highly correlated to the ∑MMMIS (r_p = 0.607-0.913) for four of the five HET-CAM data sets that were submitted. *In vitro*/*in vivo* correlations indicated that, overall, corneal opacity and iritis showed better correlation with *in vitro* endpoints than other adverse effects in the eye. When

[4] With a specificity rate of 89%, the false positive rate would be expected to be 11%.
[5] With a specificity rate of 88.3%, the false positive rate would be expected to be 10.7%

a single linear regression was used to correlate *in vitro* scores to *in vivo* endpioints, the prediction rates ranged from 16 to 36% for erythema and 84% to 88% for chemosis. However, when a partial regression was used, the prediction rates ranged from 92% to 100%. Additionally, when analyses were restricted to a protocol conducted by a single test laboratory and a limited chemical class, the tissue damage prediction was > 95%. The HET-CAM test method showed the best prediction with surfactants and surfactant-based formulations.

Data from this evaluation was encompassed by other studies that were used in evaluating accuracy and reliability of the HET-CAM test method in **Section 6.0** and **Section 7.0**,

9.1.33 Sterzel et al. (1990)

Comparative screening of 10 substances was conducted to assess the usefulness of the HET-CAM test method. In this study, the *in vitro* scores were calculated as IS(A) values. The *in vivo* rabbit eye scores were obtained from concurrent rabbit studies conducted in accordance with OECD TG 405. The authors concluded that the results indicated that the HET-CAM test method could identify irritating test substances. The study also indicated that the test method was highly sensitive and the authors concluded that, due to this heightened sensitivity, only substances that cause irritation *in vitro* over a 100-fold concentration range should be specified as potential eye irritants.

The data from this study could not be used in the HET-CAM accuracy analysis, because sufficient *in vitro* information was not available in the report. Additionally, sufficient *in vivo* data was not provided to allow for classification of the test substances into at least one of the three classification schemes (GHS [UN 2003], EPA [1996], or EU [2001]), used in this analysis.

9.1.34 van Erp et al. (1990)

The HET-CAM test method was combined with the use of bovine eyes to assess the irritancy potential of chemical substances towards the conjunctivae and the cornea, respectively. The screening method was referred to as BECAM. In this study, *in vitro* scores were calculated as IS(A) values. *In vivo* rabbit eye studies were performed concurrently in accordance with OECD TG 405. The *in vivo* effects were scored according to the Draize scoring system (Draize et al. 1944) and the scores were classified according to the classification scheme of Kay and Calandra (1962). *In vitro* classification of test substances was compared to the EEC classification system (EEC 1983, 1984).

The authors concluded that the combination of HET-CAM and BCOP *in vitro* results showed a good correlation with the *in vivo* classification results. The investigators noted three limitations with the BECAM screening method: (1) inability of the assay to determine effects on the iris, (2) substances that contained a carbamate group or adhere firmly to the bovine cornea or CAM might generate false results, and (3) it was not possible to determine if a severe effect *in vitro* would result in either reversible or irreversible injury of the eye.

The data from this study could not be used in the HET-CAM accuracy analysis because sufficient *in vitro* information was not available in the report. Additionally, sufficient *in vivo*

data was not provided to allow for classification of the test substances into at least one of the three classification schemes used in this analysis.

9.1.35 Vinardell and García (2000)

In this study, the HET-CAM test method was modified to include evaluation of damage to the membrane, which was quantified as the amount of trypan blue adsorbed into the CAM. The modified method was used to assess the potential ocular irritation caused by mixtures of liquid scintillation cocktails. Adsorbed trypan blue was quantified using a spectrophotometer. The test substances were evaluated at concentrations ranging from 12.5% to 100%. The *in vivo* rabbit eye scores (expressed as the Draize score) were obtained from concurrent studies conducted as described by Draize et al. (1944).

The analysis in the report showed that there was good correlation between test substance concentration and the amount of trypan blue adsorbed into the CAM. Additionally, there was good correlation observed between the amount of trypan blue adsorbed and *in vivo* ocular irritation ($r^2 = 0.9722$).

Individual sample or mean sample *in vitro* scores were not provided in the report. Therefore, classification of substances into standardized *in vitro* irritancy classification categories was not possible. Additionally, the *in vivo* rabbit study data was insufficient to classify the substances according to one of the three classification systems evaluated in this analysis.

9.1.36 Vinardell and Macián (1994)

The irritancy potential of substances used as vehicles (six chemicals) and disinfectant solutions (six solutions) were evaluated in the HET-CAM test method to assess ocular irritancy potential. In this study, the *in vitro* scores were calculated as IS(B) values. *In vivo* scores for the six solutions were obtained from concurrent *in vivo* rabbit eye studies that were conducted in accordance with Draize et al. (1944). For the rabbit studies, the ocular irritation index was calculated, which corresponded to the highest total value obtained after a single application. The results of the *in vivo* test were categorized according to Le Moult et al. (1976) and a previous classification scheme implemented by the EPA (1974).

In this study, four of the six vehicles were classified as nonirritants or weak irritants, while the remaining two vehicles (0.1 N NaOH and 1% SLS) were classified as severe irritants. These results were not compared with *in vivo* rabbit eye test results. The study indicated that four of the six tested disinfectant solutions gave similar results *in vitro* and *in vivo* (when classified by the Le Moult et al. or EPA [1974] classification systems).

The disinfectant solution test data from this study could not be used in the accuracy analysis because sufficient information on the test formulations was not provided for additional analysis. *In vivo* data for some of the vehicles tested were obtained from published sources (e.g., ECETOC); a HET-CAM accuracy analysis of these substances is provided in **Section 6.0**.

9.1.37 Vives et al. (1997)

The ocular irritation potentials of six anionic and nonionic surfactants, which were derived from lysine, were evaluated in the HET-CAM test method. The focus of this evaluation was to correlate irritation potential with structural characteristics of the surfactants in order to develop a less irritating surfactant. In this study, *in vitro* scores were calculated as IS(B) values. The ocular effects of these substances *in vivo* were not evaluated. This evaluation showed that anionic surfactants showed higher irritation potential than nonionic surfactants. However, the presence of lysine as a counterion reduced the irritancy potential of anionic surfactants.

The data from this study could not be used in the HET-CAM accuracy analysis because effects of these substances in the *in vivo* rabbit test were not evaluated and historical rabbit test method information on these surfactants could not be located.

9.1.38 Wilson and Steck (2000)

A modified HET-CAM test method protocol was used by the investigators to assess the anti-irritant properties of plant extracts. The investigators measured delays in the onset of vascular hemorrhage, membrane lysis, and membrane coagulation relative to the effect of the irritant (15% lactic acid) alone. In this study, *in vitro* scores were calculated as IS(B) values. An anti-irritation score (AIS) then was calculated which represented the time of onset of one of the measured endpoints with pretreatment of a test substance compared to the onset of the measured endpoint without pretreatment of the test substance. The three AIS values were used in describing the anti-irritant potential of the test substances. The *in vivo* results were obtained from studies on human volunteers and the effect was evaluated over a 24-hour period.

The data from this study could not be used in the HET-CAM accuracy analysis because the response being evaluated was not ocular irritation potential, but anti-irritant responses.

9.1.39 Worth and Cronin (2001)

The investigators developed prediction models to explore the possibility of distinguishing between eye irritants (as expressed by the EU classification system [EU 1993]) and nonirritants, by using *in vitro* endpoints of the HET-CAM test method and the neutral red uptake test. The investigators used the *in vitro* data published in the report by Spielmann et al. (1996) to develop the prediction models. The quality of each prediction model was determined by applying it to a training set of 129 chemicals and by expressing the goodness of fit in terms of the sensitivity, specificity, concordance, false negative rate, false positive rate, negative predictivity, and positive predictivity of the prediction models.

Four prediction models were developed by the authors. Using a training set of 129 chemicals, the investigators determined that a combination of three endpoints provided the best prediction of *in vivo* ocular irritation. The prediction model used was:

> If 3.63 log (TH10) + 2.10 log (TH10) + 0.94 log (IC50) < 11.87, predict Irritant; otherwise predict Nonirritant

where

TH10 = mean detection time for hemorrhage with a 10% solution
IC50 = concentration of test chemical (mg/mL) resulting in 50% inhibition of neutral red uptake in 3T3 cells[6]

Accuracy statistics indicate that this prediction model, using the training set, had an accuracy rate of 81%, specificity of 90%, sensitivity of 69%, false negative rate of 31%, and a false positive rate of 10%[7].

The HET-CAM data described in this report was initially described in Spielmann et al. (1996); these data are considered in **Section 6.0**.

9.2 Data Received in Response to the ICCVAM *Federal Register* Notice or from Study Authors

NICEATM staff made attempts to obtain original HET-CAM data for substances that also had been tested *in vivo* using the standard rabbit eye test. A *FR* notice (Vol. 69. No. 57, pp. 13589-12861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original HET-CAM (and comparative *in vivo* rabbit) data was published on March 24, 2004. A second *FR* notice for original HET-CAM (and comparative *in vivo* rabbit) was published on February 28, 2005 (Vol. 69, No. 38, pp. 9661-9662). In addition, NICEATM staff contacted authors of selected published HET-CAM studies to request the original HET-CAM data (Gettings et al. 1991, 1994, 1996; Gilleron et al. 1996, 1997; Spielmann et al. 1996; Hagino et al. 1999). In response to these efforts, the following data were obtained.

In vivo data was submitted by the CTFA for the studies described in Gettings et al. (1991, 1994, 1996). Individual animal responses for the days that the animals were observed were provided. This data was used to identify the ocular irritant potential (based on the GHS [UN 2003], EPA [1998], or EU [2001] classification systems) of the test substances for each formulation evaluated. Using this information, combined with the results provided in the published literature, the accuracy of each version of the HET-CAM test method used in these reports was determined and these results are provided in **Section 6.0**.

In vivo data was submitted by Dr. Yasuo Ohno, of the National Institute of Health Sciences in Japan, for the test substances and test concentrations used in the evaluation described in Hagino et al. (1999). Individual animal responses for the days that the animals were observed were provided. This data was used to identify the ocular irritant potential (based on the GHS [UN 2003], EPA [1998], and EU [2001] classification systems) of the test substances for each substance evaluated in this study. Using this information, combined with the results provided in the published literature, the accuracy and reliability of the tested

---

[6] The prediction model noted to have the best performance contained three variables, TH10, TC10 (mean detection time for coagulation with a 10% solution), and IC_{50}. However, the prediction model shown in the reference only indicated two variables, TH10 (repeated twice) and IC_{50}. According to the text, it appears that one of the TH10 variables in the equation should be TC10; it is unclear from the text which TH10 should be changed to TC10.

[7] The numbers used to generate these values were not provided in the literature study and no attempt was made to calculate the values from the prediction model provided in the reference.

version of the HET-CAM test method was determined and the results are provided in **Sections 6.0** and **7.0**.

In vivo and *in vitro* data was submitted by Dr. med. Horst Spielmann and Dr. Manfred Liebsch of ZEBET. Individual animal responses for the days that the animals were observed were provided. This data was used to identify the ocular irritant potential (based on the GHS [UN 2003], EPA [1998], and EU [2001] classification systems) of the test substances for each substance evaluated in this study. *In vitro* scores for each test substance for each individual testing laboratories was provided as were results using control substances. Using this information, the accuracy and reliability of the tested version of the HET-CAM test method was determined, and the results are provided in **Sections 6.0** and **7.0**. Results of the analyses conducted with control test substances are provided in **Section 2.0**.

In response to a request from NICEATM, *in vitro* HET-CAM scores were obtained for the substances evaluated in Balls et al. (1995). The data, provided by ECVAM, comprised of Q-Scores and S-Scores for all tested substances for each testing laboratory. The individual sample scores were not provided. Comparative *in vivo* individual rabbit data was obtained from the ECETOC database (ECETOC 1998). These data were used to identify the ocular irritant potential (based on the GHS [UN 2003], EPA [1998], and EU [2001] classification systems) of each substance evaluated. Using this information, combined with the results provided in the published literature, the accuracy and reliability of the tested version of the HET-CAM test method used in this study were determined, and the results are provided in **Sections 6.0** and **7.0**.

In vitro data was submitted by Dr. Philippe Vanparys and Dr. Freddy Van Goethem of Johnson & Johnson Pharmaceutical R&D (a division of Janssen Pharmaceutica N.V.). Times of development of endpoints for each egg tested for substances were provided for data presented in Gilleron et al. (1996, 1997). Furthermore, results from studies using control substances were provided upon request. Using this information, combined with the results provided in the published literature, the accuracy and reliability of this version of the HET-CAM test method used in this study were determined and the results of this re-analysis are provided in **Section 6.0** and **Section 7.0**. Results of the analyses conducted with control test substances are provided in **Section 2.0**.

[This Page Intentionally Left Blank]

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

10.1 How the HET-CAM Test Method Will Refine, Reduce, or Replace Animal Use

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

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

The HET-CAM test method has the potential to refine and reduce animal use in eye irritation testing. The HET-CAM test method refines animal use by the *in vitro* identification of ocular corrosives and severe irritants, when used in the tiered testing scheme. According to the GHS tiered testing scheme, substances are to be initially tested in a valid *in vitro* ocular test method, if available, that assesses severe eye damage (see **Figure 1-2**). If the substance is identified as a corrosive or severe irritant, it may be classified as such and no additional testing is required. If a negative result is obtained from this test, then the substance is to be tested first in a valid *in vitro* test method, if available, for eye irritation. If the substance is identified as an eye irritant, it may be classified as such and no additional testing is required. The next step in the testing scheme is to assess dermal corrosion potential (either by *in vitro* or *in vivo*) methods. If the substance is classified as noncorrosive, it is then evaluated in a single rabbit. The HET-CAM test method, used in the GHS tiered testing strategy, would be used prior to evaluation in any animals (to assess severe eye damage). This test method, therefore, will reduce the numbers of animals subjected to testing and reduce the pain and suffering of rabbits by their exclusion from the testing of corrosives and severe irritants.

10.2 Requirement for the Use of Animals

The HET-CAM test method has been designed so as not to require the use of animals. International regulations have provisions for the protection of animals used for experimental or other scientific purposes. Some provisions indicate the time in which a test method using an animal embryo or fetus is considered an animal, and therefore protected by the regulations. According to some of these regulations, a bird is considered a protected animal (and therefore the test is considered an *in vivo* and not *in vitro* test) when greater than half of the gestation or incubation period has elapsed (day 10.5 of the 21 day incubation period for a chicken embryo) (Animals [Scientific Procedures] Act 1986; EU 1986). The Public Health Service Policy, with which all National Institutes of Health (NIH)-funded research projects must comply, applies to all live vertebrate species. The NIH Office of Laboratory Animal Welfare has provided written guidance in this area, interpreting "live vertebrate animal" to apply to avians (e.g., chick embryos) only after hatching (Kulpa-Eddy J, personal communication; NIH 2000).

It has been proposed that at incubation day nine, the embryonic differentiation of the chicken central nervous system is sufficiently incomplete that suffering from pain perception is unlikely to occur (MSPCA 2005; Liebsch M, personal communication). Evaluations suggest that there are few sensory fibers present at day nine in the avian embryo and that there is significant development of the sensory nerve ending between incubation days 11 and 14 (Romanoff 1960). Studies also have suggested that the extraembryonal vascular systems (e.g., yolk sac, CAM) are not sensitive to pain (Rosenbruch 1997; Spielmann H, personal communication). Combined, these studies suggest that at incubation day nine there is little to no pain perceived by the developing embryo during the conduct of the HET-CAM test method.

11.0 PRACTICAL CONSIDERATIONS

Several issues are taken into account when assessing the practicality of using an *in vitro* test method in place of an *in vivo* test method. In addition to reliability and accuracy evaluations, assessments of the equipment and supplies needed for the *in vitro* test method, level of personnel training, costs of the *in vitro* test method, and time to complete the method are necessary. This information provides additional information as whether the time, personnel cost, and effort required to conduct the test method is considered reasonable.

11.1 Transferability of the HET-CAM Test Method

Test method transferability is defined as the ability of a test method to be accurately and reliably performed by different, competent laboratories (ICCVAM 2003). Issues of transferability include laboratories experienced in the particular type of procedure, and otherwise competent laboratories with less or no experience in the particular procedure. The degree of transferability of a test method affects its interlaboratory reproducibility.

11.1.1 Facilities and Major Fixed Equipment Required to Conduct a Study

The facilities needed to conduct the HET-CAM test method are widely available, and the necessary equipment is readily available from major suppliers. Major facilities necessary for performing the HET-CAM test method are standard toxicology, biochemistry, or molecular biology laboratory supplies. There are no specific requirements regarding the facility at which the test is conducted (e.g., sterile environment). However, it would seem appropriate to conduct the assay under ambient temperature and humidity conditions. To perform the test method an incubator that can rotate samples is needed. Depending upon the features of such an incubator, the cost of obtaining such an incubator through scientific vendors (e.g. Fisher Scientific, Phoenix Equipment Company) can range from about $130 to $4000.

The *in vivo* ocular toxicity test method requires that facilities develop and maintain an animal facility that adheres to pertinent State and Federal regulations. Personnel that are trained and skilled in dealing with such facilities also are needed for the *in vivo* test method. These facilities or personnel are not needed to conduct the HET-CAM test method. Similar to the *in vitro* test method, the *in vivo* test method uses equipment that is readily available from major suppliers.

11.1.2 General Availability of Other Necessary Equipment and Supplies

All other necessary equipment and supplies (e.g., candling light, rotating saw blade, pipettes, flasks, stop watch) are readily available from major scientific supply vendors.

11.2 Training Considerations

Training considerations are defined as the level of instruction needed for personnel to conduct the test methods accurately and reliably (ICCVAM 2003). Evaluation of the level of training and expertise needed to conduct the test method reliably and accurately, as well as the training requirements needed to ensure that personnel are competent in the test method, are discussed below.

11.2.1 Required Level of Training and Expertise Needed To Conduct the HET-CAM Test Method

An assessment of the protocols described in the reports reviewed in this BRD indicates that basic laboratory skills and training in embryo handling are necessary to conduct the HET-CAM test method. Some specialized training in removing the eggshell without damaging the inner membrane of the egg and removing the inner membrane from the CAM may be necessary to ensure competency in preparing the egg for the test method. Additionally, training in identifying the development of each of the three evaluated endpoints is necessary. A training video or other visual media to provide guidance on development of endpoints may be considered for use.

The level of training needed for the HET-CAM test method is not significantly greater than that required to conduct the *in vivo* ocular toxicity test method. Both the *in vivo* and *in vitro* test methods require developing competence in identifying endpoint development.

11.2.2 Training Requirements Needed to Demonstrate Proficiency

There are currently no known proficiency criteria used to ensure that personnel who are performing the test method are competent. Rather, this must be demonstrated through experience with the oversight of an experienced supervisor. Once the technician has demonstrated competence in identifying the study endpoints, it would seem appropriate for routine assessments of observations among trained personnel be conducted to ensure consistency.

11.3 Cost Considerations

A GLP-compliant EPA OPPTS Series 870 Acute Eye Irritation test (EPA 1998) or OECD TG 405 test (OECD 2002) at MB Research Laboratories (Spinnerstown, PA) ranges from $765 for a three day/three animal study up to $1665 for a 21 day/three animal study (MB Research laboratories, personal communication). The current costs of performing a GLP-compliant HET-CAM test have not yet been identified but are expected to be equivalent to or lower than the cost of an *in vivo* rabbit eye test.

11.4 Time Considerations

Use of the HET-CAM test method would significantly reduce the time needed to assess the ability of a test substance to induce ocular corrosivity or severe irritancy, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one hour to three days. Depending upon the severity of ocular effects produced by a test substance, the method can be extended for up to 21 days to assess reversibility of observed effects. Completion of the HET-CAM test method requires a nine-day pre-treatment incubation period, followed by approximately one hour for the treatment and observation/measurement period.

12.0 REFERENCES

Animals (Scientific Procedures) Act 1986. 1986. Chapter 14.

[Anonymous.] 1996. Breeds of Chicken. Oklahoma State University Board of Regents. Available: http://www.ansi.okstate.edu/poultry/ [accessed 27 April 2004].

[Anonymous.] 2002. Incubation & Embryology. Available: www.urbanext.uiuc.edu/eggs/res19-opincubator.html [accessed 30 August 2004].

[Anonymous.] 2004a. Incubation: Selection and Care of Hatching Eggs. Cornell Cooperative Extension. Available: www.cce.cornell.edu/~Allegany-Cattaraugus/hatchingeggs404.htm [accessed 30 August 2004].

[Anonymous.] 2004b. Incubation & Hatching. Available: http://www.feathersite.com/Poultry/BRKIncubation.html [accessed 30 August 2004].

ASTM. 1999. Standard practice for conducting an interlaboratory study to determine the precision of a test method. ASTM E691-92. In: Annual Book of ASTM Standards. Philadelphia:American Society for Testing and Materials.

Bagley D, Bruner L, DeSilva O, Cottin M, O'Brien K, Uttley M, Walker AP. 1992. An evaluation of five potential alternatives in vitro to the rabbit eye irritation test in vivo. Toxic In vitro 6:275-284.

Balls M, Botham P, Bruner L, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxic In Vitro 9:871-929.

Barkman R, Germanis M, Karpe G, Malmborg AS. 1969. Preservatives in drops. Acta Opthalmol. 47:461-475.

Bartnik FG, Holtmann W, Lupke NP, Sterzel W, Wallat S, Kunstler K. 1987. Final report of the BMFT-project: Validierung von Ersatmethoden fur Tierversuche zur Prifung auf localeVertraglichkeit, Henkel KgaA, Ressort Forschung, Dusseldorf.

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

Berta A. 1992. Chapter 17. In: Enzymology of the tears. (Berta A. ed.). Boca Raton:CRC Press, 285-295.

Blein O, Adolphe M, Lakhdar B, Cambar J, Gubanski G, Castelli D, Contie C, Hubert F, Latrille F, Masson P, Clouzeau J, LeBigot JF, DeSilva O, Dossou KG. 1991. Correlation and validation of alternative methods to the Draize eye irritation test (Opal Project). Toxic In Vitro 5:555-557.

BLS. 2004. US Dept. of Labor, Bureau of Labor Statistics, Case and Demographic Characteristics for Work-related Injuries and Illnesses Involving Days Away From Work Available: http://www.bls.gov/iif/oshcdnew.htm [accessed 26 October 2004].

Branter A, Quehenberger F, Chakraborty A, Polligger J, Sisa S, DellaLoggia R. 2002. HET-CAM bioassay as in vitro alternative to the croton oil test for investigating steroidal and non-steroidal compounds. ALTEX 19:51-56.

Brantom P, Bruner L, Chamberlain M, DeSilva O, Dupuis J, Earl L, Lovell DP, Pape WJW, Uttley M, Bagley DM, Baker FW, Bracher M, Courtellemont P, Declercq L, Freeman S, Steiling W, Walker AP, Carr GJ, Dami N, Thomas G, Harbell J, Jones PA, Pfannenbecker U, Southee JA, Tcheng M, Argembeaux H, Castelli D, Clothier R, Esdaile DJ, Itigaki H, Jung K, Kasai Y, Kojima H, Kristen U, Larnicol M, Lewis RW, Marenus K, Moreno O, Peterson A, Rasmussen ES, Robles C, Stern M. 1997. A summary report of the COLIPA International validation study on alternatives to the Draize rabbit eye irritation test. Toxic In Vitro 11:141-179.

Bruner L. 1992. Chapter 7: Ocular irritation. In: In Vitro Toxicology Testing (Frazier J ed.). New York:Marcel Dekker, Inc., 149-190.

Bruner LH, Evans MG, McPherson JP, Southee JA, Williamson PS. 1998. Investigation of ingredient interactions in cosmetic formulations using isolated bovine corneas. Toxicology In Vitro 12:669-690.

Budai P, Várnagy L, Somlyay I, Varga T. 1997. Ocular irritation study of some pesticides in HET-CAM test. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische 62:265-268.

Budai P, Várnagy L. 2000. In vitro ocular irritation toxicity study of some pesticides. Acta Veternaria Hungarica 48:221-228.

Butscher P. 1953. Beitrag zur therapie von augenschadigunen durch thioglykolsaur bei der herstellung der sogenannten kaltwelle. Klin Monatsbl Augenheilkd. 122:349-350.

Calabrese EJ. 1983. Dermatotoxicity: Predictive Models. In: Principles of animal extrapolation. Boca Raton:CRC Press, 391-404.

Carpenter CP, Smyth HF. 1946. Chemical burns of the rabbit cornea. Am J Opthalmol. 29:1363-1372.

Cater K, Patrick E, Harbell J, Merrill J, Schilcher S. 2004. Comparison of in vitro eye irritation potential by BCOP assay to erythema scores in human eye sting test of surfactant-based formulations. [Abstract]. Toxicol Sci 78:1307.

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.

Clauer PJ. 2000. Incubating eggs. Available:
http://www.ext.vt.edu/pubs/poultry/factsheets/8.html[accessed 20 August 2004].

Clauer PJ. 2002. Embryology in the Classroom. Physiological Processes Within the Egg. The Pennsylvania State University. Available:
http://ulisse.cas.psu.edu/4hembryo/physiology.html [accessed 20 January 2004].

Clean Air Act Amendments of 1990. 1990. Public Law 101-549.

Cooper KJ, Earl LK, Harbell J, Raabe H. 2001. Prediction of ocular irritancy of prototype shampoo formulations by the isolated rabbit eye (IRE) test and bovine corneal opacity and permeability (BCOP) assay. Toxic In Vitro 15:95-103.

CPSC. 1984. Animal Testing Policy. Fed Reg 49:22522-22523.

CPSC. 1988. Commercial practices. Hazardous Substances and Articles; Administration and Enforcement Regulations. 16 CFR 1500.42.

CPSC. 1995. Test for eye irritants. 16CFR1500.42.

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

CPSC. 2004. Definitions. Hazardous Substances and Articles; Administration and Enforcement Regulations 16 CFR 1500.3.

Curren RD, Harbell JW. 1998. In vitro alternatives for ocular irritation. Environ Health Persp 106:485-492.

Dalbey W, Rodriguez S, Wilkins K, Cope C. 1993. Reducing the number of rabbits in eye and skin irritancy tests. Journal of the American College of Toxicology 12:347-357.

Dannhardt G, Kreher M, Nowe U, Pies A. 1996. Method for testing non-steroidal anti-inflammatories: the modified hen's egg chorioallantoic membrane test (HET-CAM test) compared to other procedures. Arch Pharm 329:301-310.

Demirci B, Dadandi M, Paper D, Franz G, Baser K. 2003. Chemical composition of the essential oil of Phlomis linearis Boiss. & Bal., and biological effects on the CAM-assay: A safety evaluation. Z Naturforsch 58:826-829.

Demirci F, Paper D, Franz G, Baser K. 2004. Investigation of the Origanum onites L. essential oil using the chorioallantoic membrane (CAM) assay. J Agric Food Chem 52:251-254.

de Silva O, Rougier A, Dossou K. 1992. The HET-CAM Test: A study of the irritation potential of chemicals and formulations. ATLA 20:432-437.

DeSousa D, Rouse A, Smolon W. 1984. Statistical consequences of reducing the number of rabbits utilized in eye irritation testing: Data on 67 petrochemicals. Tox Applied Pharm 76:234-242.

Djabari Z, Bauza E, Farra CD, Domloge N. 2002. The HET-CAM test combined with histological studies for better evaluation of active ingredient innocuity. Int J Tissue React 24:117-121.

Doucet O, Lanvin M, Zastrow L. 1999. Comparison of three in vitro methods for the assessment of the eye-irritating potential of formulated cosmetic products. In Vitro Mol Toxic 12:63-76.

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.

Dua HS, Azuara-Blanco A. 2000. Limbal stem cells of the corneal epithelium. Survey of Ophth 44(5):415-425.

Duane TD. 1949. The steady state of corneal hydration. Am J Ophth 32:203-207.

ECETOC 1992. Eye Irritation – Reference Chemicals Data Bank. Technical Report No. 48. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

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

EEC. 1983. Annex VI of the EEC Directive 84/479, Annex, Part B, Methods for the determination of toxicity, B.5 "Acute toxicity – Eye irritation". EEC Publication No. L251, September 1983.

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

EEC. 1991. Classification of irritant substances and preparations. Official J European Community. L180:52.

EPA. 1974. Fed Reg 39:801.

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

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

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

EPA. 2003b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and Rodenticide Act. 40 CFR 160.

Eskes C, Bessou S, Bruner L, Curren R, Harbell J, Jones P, KreilingR, Liebsch M, McNamee, Pape W, Prinsen M, Seidle T, Vanparys P, Worth A, Zuang V. 2005. Eye Irritation. In: Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Perspectives (Eskes C, Zuang V, eds.). ATLA 33 (Supplement 1): 47-81.

Estable JL. 1948. The ocular effect of several irritant drugs applied directly to the conjunctiva. Am J Opthalmol. 31:837-844.

EU. 1986. Council Directive 86/69/EEC. Official Journal of the European Economic Community. L358:1-28.

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

EU. 1993. Commission Directive 93/21/EEC of 27 April 1993 adapting to technical progress for the eighteenth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Union L110A:1-86.

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L255:1-333.

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

FDA. 2003. Good laboratory practice for nonclinical laboratory studies. 21 CFR 58.

Federal Environmental Pesticide Control Act of 1972. 1972. Public Law 92-516.
FSHA. 1964. Federal Hazardous Substances Act. Public Law 86-613.

Federal Insecticide, Fungicide and Rodenticide Act of 1947. 1947. Public Law 80-102.

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

FDA. 1938. Food, Drug and Cosmetic Act of 1938. Public Law 75-717.

Fox DA, Boyes WK. 2001. Toxic responses of the ocular and visual system. In: Casarett and Doull's Toxicology, the Basic Science of Poisons (Klaassen CD, ed.). 3rd ed. New York:McGraw-Hill, 565-595.

Friedenwald JS, Hughes WF, Hermann H. 1946. Acid burns of the eye. Arch Ophth. 35:98-108.

Gartner S. 1944. Blood vessels of the conjunctiva. Arch Opthalmol. 36:464-471.

Gautheron P, 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, Teal J, Bagley D, Demetrulias J, DiPasquale L, Hintze K, Rozen MG, Weise SL, Chudkowski M, Marenus KD, Pape WJW, Roddy MT, Schenetzinger R, Silber PM, Glaza SM, Kurtz PJ. 1991. The CTFA evaluation of alternatives program: An evaluation of in vitro alternatives to the Draize primary eye irritation test (Phase 1) Hydro-alcoholic formulations; (Part 2) Data analysis and biological significance. In Vitro Toxic 4:247-288.

Gettings S, Dipasquale L, Bagley D, Casterton P, Chudkowski M, Curren R, Demetrulias JL, Feder PI, Galli CL, Gay R, Glaza SM, Hintze KL, Janus J, Kurtz, Lordo RA, Marenus KD, Moral J, Muscatuello M, Pape WJW, Renskers KJ, Roddy MT, Rozen MG, Tedeschi JP, Zyracki J.1994. The CTFA evaluation of alternatives program: an evaluation of in vitro alternatives to the Draize primary eye irritation test. (Phase II) oil/water emulsions. Food Chem Toxic 32:943-976.

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.

Gilleron L, Coecke S, Sysmans M, Hansen E, van Oproy S, Marzin D, van Cauteren H, Vanparys P. 1996. Evaluation of a modified HET-CAM assay as a screening test for eye irritancy. Toxic In Vitro 10:431-446.

Gilleron L, Coecke S, Sysmans M, Hansen E, van Oproy S, Marzin D, van Cauteren H, Vanparys P. 1997. Evaluation of the HET-CAM-TSA method as an alternative to the Draize eye irritation test. Toxic In Vitro 11:641-644.

Grant WM. 1974. Toxicology of the eye. In: Toxicology of the eye. 2nd ed. (Grant WM, ed.) Springfield: Charles C Thomas.

Hagino S, Itagaki H, Kato S, Kobayashi T, Tanaka M. 1991. Quantitative evaluation to predict the eye irritancy of chemicals: modification of chorioallantoic membrane test by using trypan blue. Toxic In Vitro 5:301-304.

Hagino S, Itagaki H, Kato S, Kobayashi T. 1993. Further evaluation of the quantitative chorioallantoic membrane test using trypan blue stain to predict the eye irritancy of chemicals. Toxic In Vitro 7:35-39.

Hagino S, Kinoshita S, Tani N, Nakamura T, Ono N, Konishi K, Iimura H, Kojima H, Ohno Y. 1999. Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients. (2) Chorioallantoic membrane (CAM) test. Toxic In Vitro 13:99-113.

Harbell JW, Curren RD. 2002. In vitro methods for the prediction of ocular and dermal toxicity. In: Handbook of Toxicology (Eds. MJ Derelanko and MA Hollinger). CRC Press: Boca Raton, 835-866.

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjöström M. 1996. Recommendations of the application of biostatistical methods during the development and validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. ATLA, 24:511-530.

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

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

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

INVITTOX 1992. HET-CAM Test. ECVAM. Available: https://ecvam-sis.jrc.it/invittox/published/indexed_47.html [accessed 18 February 2004].

Kalweit S, Gerner I, Spielmann H. 1987. Validation project of alternatives for the Draize eye test. Mol Toxic 1:597-603.

Kalweit S, Besoke R, Gerner I, Spielmann H. 1990. A national validation project of alternative methods to the Draize rabbit eye test. Toxic In Vitro 4:702-706.

Kay JH, Calandra JC. 1962. Interpretation of Eye Irritation Tests. J Soc Cosmetic Chemists 12:281-289.

Kinoshita J. 1962. Some aspects of the carbohydrate metabolism of the cornea. Inves Ophth. 1:178-186.

Kojima H, Hanamura A, Miyamoto S, Sato A, Konishi H, Yoshimura I. 1995. Evaluation of seven alternative assays on the main ingredients in cosmetics as predictors of Draize eye irritation scores. Toxic In Vitro 9:333-340.

Kuckelkorn R, Schrage N, Keller G, REdbrake C. 2002. Emergency treatment of chemical and thermal eye burns. Acta Ophth Scand 80:4-10.

Lawrence R, MH G, Ackroyd D, Parish W. 1986. The chorioallantoic membrane in irritation testing. Food Chem Toxic 24:497-502.

Lawrence R, Ackroyd D, Williams D. 1990. The chorioallantoic membrane in the prediction of eye irritation potential. Toxic In Vitro 4:321-323.

Le Moult CH, Cotte J, Martini MC. 1976. Etude critique des tests d'innocuite applicables aux produits cosmetiques. Mise en evidence des reactions toxiques generales et locales. Parfums, Comsetiques, Aromes. 7: 55-71.

Leopold IH. 1945. Local toxic effect of detergents on ocular structures. Arch Opthalmol. 34:99-102.

Lewin L, Guillery H. 1913. Die Wirkungen von Arzneimitteln und Giften auf das Auge. 2nd edition. Hirschwald, Berlin.

Liebsch L, Spielmann H. 2002. Currently available in vitro methods used in the regulatory toxicology. Toxic Lett 127:127-134.

Lönnroth E, Dahl J, Shahnavaz H. 1999. Evaluating the potential occupational hazard of handling dental polymer products using the HET-CAM technique. Int J Occupational Safety Ergonomics 5:43-57.

Lordo R, Feder P, Greetings S. 1999. Comparing and evaluating alternative (in vitro) tests on their ability to predict the Draize maximum average score. Toxic In Vitro 13:45-72.

Luepke N. 1985. Hen's egg chorioallantoic membrane test for irritation potential. Food Chem Toxic 23:287-291.

Luepke N, Kemper F. 1986. The HET-CAM test: an alternative to the Draize eye test. Food Chem Toxic 24:495-496.

Macián M, Seguer J, Infante M, Selve C, Vinardell M. 1996. Preliminary studies of the toxic effects of non-ionic surfactants derived from lysine. Toxic 106:1-9.

Marsh RJ, Maurice DM. 1971. The influence of non-ionic detergents and other surfactants on human corneal permeability. Exp Eye Res. 11:43-48.

Marzulli F, Ruggles D. 1973. Rabbit eye irritation test: collaborative study. Journal of the Association of Analytical Chemists. 56:905-914.

Maurice DM. 1957. The structure and transparency of the cornea. J Physiol 136:263-286.

McCulley JP. 1987. Chemical injuries. In: Cornea; scientific foundations and clinical practice (Smolin G, Thoft RA, eds.). Boston:Little, Brown, 527-542.

McDonald TO, Seabaugh V, Shadduck JA, Edelhauser HF. 1987. Eye irritation. In: Dermatotoxicology. (Marzulli FN, Maibach HI, eds). 3rd edition. Washington:Hemisphere Publishing Corporation, 641-696.

McLaughlin RS. 1946. Chemical burns of the human cornea. Am J Opthalmol. 29:1355-1362.

Mishima S, Hedbys BO. 1968a. Physiology of the cornea. Int Opt Clin. 8(3):527-560.

Mishima S, Hedbys BO. 1968b. Measurement of corneal thickness with the Haag-Streit pachymeter. Arch Ophthal. 80:710-713.

MSPCA. 2005. Animals in Education: Laws and Regulations. Available: http://www.mspca.org/site/pp,asp?c=gtIUK4Osg&b=127066 [accessed 27 October 2005].

Nakano M. 1958. Effect of various antifungal preparations on the conjunctiva and cornea of rabbits. Yakuzaigaku. 18:94-99.

NIH. 2000. The Public Health Service Responds to Commonly Asked Questions. Available: http://grants.nih.gov/grants/olaw/references/ilar91.htm - 1 [accessed 7 July 2004].

NIOSH. 2004. Work-Related Injury Statistics Query System. Available: http://www2a.cdc.gov/risqs/ [accessed 26 October 2004].

Nourse WL, Tyson, CA, Bednarz RM. 1995. Mechanisms of Mild Ocular Irritation. Toxicol In Vitro. 9:967-976.
Occupational Safety and Health Act of 1970. 1970. Public Law 91-596.

OECD. 1987. Test guideline 405, Acute eye irritation/corrosion, adopted February 24, 1987. In OECD Guidelines for Testing of Chemicals. OECD, Paris.

OECD. 1996. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Methods. ENV/MC/CHEM/TG (96)9. OECD, Paris.

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

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

Ohno Y, Kaneko T, Inoue T, Moriakawa Y, Yoshida T, Fujii A, Masuda M, Ohno T, Hayashi M, Momma J, Uchiyama T, Chiba K, Ikeda N, Imanishi Y, Itakagaki H, Kakishima H, Kasai Y, Kurishita A, Kojima H, Matsukawa K, Nakamura T, Ohkoshi K, Okumura H, Saijo K, Sakamoto K, Suzuki T, Takano K, Tatsumi H, Tani N, Usami M, Watanabe R. 1999. Interlaboratory Validation of the in vitro eye irritation testes for cosmetic ingredients. (1) Overview of the validation study and Draize scores for the evaluation of the tests. Toxic In Vitro. 13:73-98.

Oksala O, Stjernschantz J. 1988. Effects of calcitonin gene-related peptide in the eye. Inv Opth Vis Sci. 29(7):1006-1011.

Parish W. 1985. Ability of in vitro (corneal injury-eye organ-and chorioallantoic membrane) tests to represent histopathological features of acute eye inflammation. Food Chem Toxic 23:215-227.

Pasquale LC, Hayes AW. 2001. Chapter 18. Acute Toxicity and Eye Irritancy. In: Principles and Methods of Toxicology 4th Ed. (Hayes AW, ed.). Philadelphia:Taylor and Francis, 853-915.

Pfister R. 2005. Chemical Trauma. In: Smolin and Thoft's, The Cornea: Scientific Foundations and Clinical Practice (Foster CS, Azar DT, and Dohlman CH eds). New York:Lippincort Williams and Wilkins, 781-796.

Reinhardt C, Aeschbacher M, Bracher M, Spengler J. 1987. Validation of three cell toxicity tests and the hen's egg test with guinea pig eye and human skin irritation data. Methods Toxic 5:463-469.

Romanoff AL. 1960. The avian embryo: Structural and functional development. New York: The Macmillan Company.

Rosenbruch M. 1997. Zur Sensitivitat des Embryos im bebruteten Huhnerei. ALTEX 111-113.

Rougier A, Cottin M, Silva O, Roguet R, Catroux P, Toufic A, Dossou KG. 1992. In vitro methods: their relevance and complementarity in ocular safety assessment. Lens Eye Tox Res 9:229-245.

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

Schlage W, Bülles H, Kurkowsky B. 1999. Use of the HET-CAM test for the determination of the irritant potential of cigarette sidestream smoke. Toxic In Vitro 13:829-835.

Sinn-Harlon J. 1998a. Chickscope. University of Illinois-Urbana-Champagne. Available: http://chickscope.itg.uiuc.edu/explore/embryology/day04/4Focuson_1.html [accessed 20 January 2004].

Sinn-Harlon J. 1998b. Chickscope. University of Illinois-Urbana-Champagne. Available: http://chickscope.itg.uiuc.edu/explore/embryology/day06/ [accessed 28 April 2004].

Smith TW. 2000. Care and Incubation of Hatching Eggs. Mississippi State University Extension Service. Available: http://www.msstate.edu/dept/poultry/hatch.htm#cond [accessed 28 April 2004].

Solti J, Freeman JJ. 1988. Effect of reducing the number of animals in acute toxicity/irritation tests on U.S. and European labeling requirements. The Toxicologist 8:263.

Spielmann H, Gerner I, Kalweit S, Moog R, Wirnsberger T, Krauser L, Kreiling R,Kreuzer H, Luepke NP, Miltenbrger HG, Müller N, Mürmann P, Paper W, Siegemund B, Spengler , Steiling W, Wiebel FJ. 1991. Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. Toxic In Vitro 5:539-542.

Spielmann H, Kalweit S, Liebsch M, Wirnsberger T, Gerner I, Bertram-Neiss E, Krauser K, Kreiling R, Miltenburger HG, Pape W, Steiling W. 1993. Validation study of alternatives to the Draize eye irritation test in Germany: cytotoxicity testing and HET-CAM test with 136 industrial chemicals. Toxic In Vitro 7:505-510.

Spielmann H. 1995. HET-CAM Test. In: Methods in Molecular Biology Vol 43. In Vitro Toxicity Testing Protocols. (O'Hare S, Atterwill C, eds). Totowa:Hunana Press, 199-204.
Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhütter H, Schneider B, Glaser S, Gerner I, Pape WJW, Kreiling R, Krauser K, Miltenburger HG, Steiling W, Luepke NP Müller N, Kreuzer H, Mürmann P, Spengler J, Bertram-Neis, E, Siegemund B, Wiebel F. 1996. Results of a validation study in Germany on two in vitro alternatives to the Draize eye irritation test, HET-CAM test and the 3T3 NRU cytotoxicity test. ATLA 24:741-858.

Spielmann H, Liebsch M, Moldenhauer F, Holzhütter H, Bagley D, Lipman J, Pape WJ, Miltenburger H, de Silva O, Hofer H, Steiling W. 1997. IRAG Working Group 2 CAM-based assays. Food Chem Toxic 35:39-66.

Spielmann H, Liebsch M. 2005. In vitro reproducibility and repeatability data for control chemicals tested in the HET-CAM test method. Unpublished data provided directly to NICEATM by H. Spielmann and M. Liebsch, ZEBET.

Springer J, Chambers W, Green S, Gupta K, Hill R, Hurley P, Lambert L, Lee C, Lee J, Liu P, Lowther D, Roberts C, Seabaugh V, Wilcox N. 1993. Number of animals for sequential testing. Food Chem Toxicol 31:105-109.

Steiling W, Bracher M, Courtellemont P, deSilva O. 1999. The HET-CAM, a useful in vitro assay for assessing the eye irritation properties of cosmetic formulations and ingredients. Toxic In Vitro 13:375-384.

Sterzel W, Bartnik F, Matthies W, Kästner W, Künstler W. 1990. Comparison of two in vitro and two in vivo methods for the measurement of irritancy. Toxic In Vitro 4:698-701.

Suker GF. 1913. Injury to cornea from oxalic acid. Opthalmol Rec. 23:40-47.

Swanston DW. 1985. Assessment of the validity of animal techniques in eye irritation testing. Food Chem Toxic 23:169-173.

Talsma D, Leach C, Hatoum N, Gibbons R, Roger J-C, Garvin P. 1988. Reducing the number of rabbits in the Draize eye irritancy test: A statistical analysis of 155 studies conducted over 6 years. Fundamental and Applied Toxicology 10:146-153.

Thoft RA. 1979. Chemical and thermal injury. Int Ophthalmol Clin Summer 19:243-256.

Toxic Substances Control Act of 1976. 1976. Public Law 94-469.

Tufan AC, Satiroglu-Tufan NL. 2005. A chick embryo chorioallatoic membrane as a model system for the study of tumor angiogenesis, invasion, and development of anti-angiogenic agents. Current Cancer Drug Targets. 5:249-266.

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

Unger WG. 1990. Review: Mediation of the Ocular Response to Injury. J Ocular Pharmacol 6(4):337-353.

Unger WG. 1992. Pharmacological and neural bases for eye irritation. Ann NY Acad Sci 641:176-186.

van Erp Y, Weterings P. 1990. Eye irritancy screening for classification of chemicals. Toxic In Vitro 4:267-269.

Vanparys P, Van Goethem F. 2005. In vitro reproducibility and repeatability data for control chemicals tested in the HET-CAM test method. Unpublished data provided directly to NICEATM by P. Vanparys and F. Van Goethem, Johnson & Johnson.

Vaughn D, Asbury T, Riordan-Eva P. 1999. General Ophthalmology. 15th ed. New York:McGraw-Hill.

Vinardell M, García L. 2000. The quantitative chlorioallantoic membrane test using trypan blue stain to predict the eye irritancy of liquid scintillation cocktails. Toxic In Vitro 14:551-555.

Vinardell M, Macián M. 1994. Comparative study of the HET-CAM test and the Draize eye test for assessment of irritancy potential. Toxic In Vitro 8:467-470.

Vives M, Macián M, Seguer J, Infante M, Vinardell M. 1997. Irritancy potential induced by surfactants derived from lysine. Toxic In Vitro 11:779-783.

Wilhelmus K. 2001. The Draize eye test. Survey of Ophthalmology 45:493-515.

Wilkie DA, Wyman M. 1991. Comparative anatomy and physiology of the mammalian eye. In: Dermal and Ocular Toxicology, Fundamentals and Methods (Hobson DW, ed). Boca Raton:CRC Press, Inc, 433-491.

Williams SJ, Graepel GJ, Kennedy GL. 1982. Evaluation of ocular irritancy potential: Interlaboratory variability and effect of dosage volume. Toxicol Letters 2:235-241.

Wilson SE, Mohan RR, Mohan RR, Ambrósio, Jr. R, Hong JW, Lee JS. 2001. The corneal wound healing response: Cytokine-mediated interaction of the epithelium, stroma, and inflammatory cells. Prog Retin and Eye Res 20(5): 625-637.

Wilson T, Steck W. 2000. A modified HET-CAM assay approach to the assessment of anti-irritant properties of plant extracts. Food Chem Toxic 38:867-872.

Worth A, Cronin M. 2001. Prediction models for eye irritation potential based on endpoints of the HETCAM and neutral red uptake tests. In Vitro Mol Toxic 14:143-156.

York M, Lawrence RS, Gibson GB. 1982. An in vitro test for the assessment of eye irritancy in consumer products – preliminary findings. Int J of Cosm Sci 4:223-234.

[This Page Intentionally Left Blank]

13.0 GLOSSARY[1]

Accuracy[2]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance" (see also "two-by-two" table). Accuracy is highly dependent on the prevalence of positives in the population being examined.

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

Benchmark substance: A substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties:

- a consistent and reliable source(s)
- structural and functional similarity to the class of substances being tested
- known physical/chemical characteristics
- supporting data on known effects
- known potency in the range of the desired response

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

Blepharitis: Inflammation of the eyelids.

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

Chorioallantoic membrane (CAM): A vascularized respiratory fetal membrane that is composed of the chorion and allantois.

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

Coagulation: The process of a liquid becoming viscous, jellylike, or solid by chemical reaction.

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

---

[1] The definitions in this Glossary are restricted to their uses with respect to the Draize rabbit eye test method and the HET-CAM test method.

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

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

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

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

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

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

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

Corneal opacity: Measurement of the extent of opaqueness of the cornea following exposure to a test substance. Increased corneal opacity is indicative of damage to the cornea. Opacity can be evaluated subjectively, as done in the Draize rabbit eye test, or objectively with an instrument such as an "opacitometer".

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

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

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

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

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

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

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

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

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

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

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

Hemorrhage: Discharge of blood from a vessel.

Hyperemia: Excess of blood in a body part.

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

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

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

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

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

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

Irritation Score: Value calculated by different analysis methods, which is used to classify the irritancy potential of a test substance. Also referred to as IS.

Irritation Threshold Concentration: The lowest concentration of a test substance required to produce a weak or slight irritant response on the CAM. Also referred to as ITC.

IS(A) analysis method: HET-CAM analysis method where endpoints are observed at specified time points after application of the test substance (typically 0.5, 2, and 5 minutes post exposure). At the time points, presence of an endpoint is determined and a score assigned, if it is present. The scores are totaled to yield an overall irritation score.

IS(B) analysis method: HET-CAM analysis method where endpoints are observed over the entire observation period after application of the test substance (typically 5 minutes). The time (in seconds) when an endpoint develops is noted and the times are used to yield an overall irritation score using a mathematical formula.

Lysis: The disintegration of blood vessels.

Mean Time to Coagulation (mtc): Mean detection time for appearance of coagulation endpoint.

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

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

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

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

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

Nonsevere irritant: (a) A substance that causes tissue damage in the eye following application to the anterior surface of the eye; the tissue damage is reversible within 21 days of application and the observed adverse effects in the eye are less severe than observed for a

severe irritant. (b) Substances that are classified as GHS Category 2A or 2B; EPA Category II, III, or IV; or EU R36 ocular irritants.

Ocular: Of or relating to the eye.

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

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

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

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

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

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

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

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

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

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

Q-Score: HET-CAM analysis method that calculates the ratio from the irritation score of a test substance compared to the irritation score of a reference substance. This HET-CAM analysis method is typically used with transparent test substances.

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

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

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

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

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

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

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

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

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

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

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

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

Solvent control: An untreated sample containing all components of a test system, including the solvent that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the same solvent. When tested with a concurrent negative control, this sample also demonstrates whether the solvent interacts with the test system.

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

S-Score: HET-CAM analysis method that totals the severity scores for each endpoint evaluated. The highest total score is used as the S-Score. This HET-CAM analysis method is typically used with non-transparent test substances.

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

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

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

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

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

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

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

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

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

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

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

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

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