新規試験法提案書

ニワトリ摘出眼球を用いた眼刺激性試験法
(ICE法：Isolated Chicken Eye Test)

平成21年12月

国立医薬品食品衛生研究所

新規試験法提案書

平成 21 年 12 月 17 日
No.　2009-02

ニワトリ摘出眼球を用いた眼刺激性試験法（ICE 法：Isolated Chicken Eye Test）の提案

平成 21 年 12 月 17 日に東京、国立医薬品食品衛生研究所にて開催された新規試験法評価会議（通称：JaCVAM 評価会議）において以下の提案がなされた。

提案内容：ニワトリ摘出眼球を用いた眼刺激性試験法（ICE 法：Isolated Chicken Eye Test）を定められた方法で注意点を適切に守って利用すれば、化学物質の腐食性・強刺激性を科学的に評価できると結論した。

この提案書は ICCVAM(Interagency Coordinating Committee on theValidation of Alternative methods, USA)によりまとめられた BRD（Background Review Document）および評価資料をもとに、JaCVAM 眼刺激性評価委員会によりまとめられた文書を用いて JaCVAM 評価会議が OECD ガイダンス文書　No.34 に従って、評価および検討した結果、その有用性が確認されたことから作成された。

以上の理由により、行政当局の安全性評価方法として「**眼刺激性試験代替法（ICE 法）**」の使用を提案するものである。

添付資料一覧

1．JaCVAM 評価会議報告書
2．眼刺激性試験代替法（ICE 法）のための第三者評価委員会報告書

小島　肇

国立医薬品食品衛生研究所
安全性生物試験研究センター
薬理部　新規試験法評価室
室長

井上　達

JaCVAM 評価会議　議長
国立医薬品食品衛生研究所
安全性生物試験研究センター
センター長

JaCVAM 評価会議

井上　達（国立医薬品食品衛生研究所　安全性生物試験研究センター）
田中憲穂（食品薬品安全センター　秦野研究所）
吉田武美（昭和大学薬学部）
横関博雄（東京医科歯科大学）
吉村　功（東京理科大学）
中村和市（日本製薬工業協会）
岡本裕子（日本化粧品工業連合会）
大島健幸（日本化学工業協会）
小野寺博志（医薬品医療機器総合機構）
見田　活（医薬品医療機器総合機構）
吉田　緑（国立医薬品食品衛生研究所　安全性生物試験研究センター　病理部）
五十嵐良明（国立医薬品食品衛生研究所　環境衛生化学部）

オブザーバー：JaCVAM 運営委員

大野泰雄（国立医薬品食品衛生研究）
増田光輝（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
小島　肇（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
秋田正治（日本動物実験代替法学会）

JaCVAM statement
on *in vitro* ocular toxicity test methods for identifying ocular corrosive and severe irritants: Isolated Chicken Eye Test Method

At the meeting concerning the above method, held on 19 December 2009 at the National Institute of Health Sciences (NIHS), Tokyo, Japan, the members of the Japanese Center for the Validation of Alternative Methods (JaCVAM) Regulatory Acceptance Board [1] unanimously endorsed the following statement:

Following the review of the results of the ICCVAM(Interagency Coordinating Committee on theValidation of Alternative methods, USA) Background Review Document and Evaluation Report, it is concluded that the ***in vitro* ocular toxicity test methods: Isolated Chicken Eye Test Method** can be used **for identifying ocular corrosive and severe irritants**.

The JaCVAM Regulatory Acceptance Board has been regularly kept informed of the progress of the study, and this endorsement is based on an assessment of various documents, including, in particular, the report on the results from the study, and also on the evaluation supported by JSAAE of the study prepared for the JaCVAM ad hoc peer review panel.

Hajime Kojima,
Director,
JaCVAM,
National Centre for Biological Safety and Research (NCBSR)
NIHS,
Tokyo

Tohru Inoue,
Director,
NCBSR,
NIHS,
Tokyo

19 December, 2009

The JaCVAM Regulatory Acceptance Board was established by the JaCVAM Steering Committee, and is composed of nominees from the industry and academia.

This statement was endorsed by the following members of the JaCVAM Regulatory Acceptance Board:

Mr. Tohru Inoue (NIHS)
Mr. Noriho Tanaka (Food and Drug Safety Center)
Mr. Takemi Yoshida (Showa Univ.)
Mr. Hiroo Yokozeki (Tokyo Medical and Dental Univ.)
Mr. Isao Yoshimura (Tokyo Univ. of Science)
Mr. Kazuichi Nakamura (Japan Pharmaceutical Manufacturers Association)
Ms Yuko Okamoto (Japan Cosmetic Industry Association)
Mr. Takeyoshi Oshima (Japan Chemical Industry Association)
Mr. Hiroshi Onodera (Pharmaceuticals and Medical Devices Agency)
Mr. Iku Mitta (Pharmaceuticals and Medical Devices Agency)
Ms Midori Yoshida (NIHS)
Mr. Yoshiaki Ikarashi (NIHS)

The following members of the JaCVAM Steering Committee were involved as observers in the consultation process, but not in the endorsement process itself.

Mr. Yasuo Ohno (NIHS)
Mr. Mitsuteru Masuda (JaCVAM)
Mr. Hajime Kojima (JaCVAM)
Mr. Masaharu Akita (JSAAE)

ニワトリ摘出眼球を用いた眼刺激性試験法（ICE 法：Isolated Chicken Eye Test）

目次

ニワトリ摘出眼球を用いた眼刺激性試験法

(ICE 法：Isolated Chicken Eye Test) の評価会議報告書

JaCVAM 評価会議

平成 21 年（2009 年）12 月 17 日
平成 23 年（2011 年）4 月 20 日改定

JaCVAM 評価会議

井上　達（国立医薬品食品衛生研究所　安全性生物試験研究センター）
田中憲穂（食品薬品安全センター　秦野研究所）
吉田武美（昭和大学薬学部）
横関博雄（東京医科歯科大学）
吉村　功（東京理科大学）
中村和市（日本製薬工業協会）
岡本裕子（日本化粧品工業連合会）
大島健幸（日本化学工業協会）
小野寺博志（医薬品医療機器総合機構）
見田　活（医薬品医療機器総合機構）
吉田　緑　（国立医薬品食品衛生研究所　安全性生物試験研究センター　病理部）
五十嵐良明（国立医薬品食品衛生研究所　環境衛生化学部）

任期：平成 21 年 1 月 1 日～平成 22 年 3 月 31 日

西川秋佳（国立医薬品食品衛生研究所　安全性生物試験研究センター）
田中憲穂（食品薬品安全センター　秦野研究所）
吉田武美（昭和大学薬学部）
横関博雄（東京医科歯科大学）
吉村　功（東京理科大学）
渡部一人（日本製薬工業協会）
岡本裕子（日本化粧品工業連合会）
大島健幸（日本化学工業協会）
小野寺博志（医薬品医療機器総合機構）
小笠原弘道（医薬品医療機器総合機構）
吉田　緑　（国立医薬品食品衛生研究所　安全性生物試験研究センター　病理部）
五十嵐良明（国立医薬品食品衛生研究所　生活衛生化学部）
長谷川隆一（独立行政法人　製品評価技術基盤機構）
浅野哲秀（元日東電工株式会社）

任期：平成 22 年 4 月 1 日～平成 24 年 3 月 31 日

オブザーバー：JaCVAM 運営委員

大野泰雄（国立医薬品食品衛生研究所）
増田光輝（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
小島　肇（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
秋田正治（日本動物実験代替法学会）
柴辻正喜（厚生労働省　医薬食品局　審査管理課　化学物質安全対策室）
実国慎一（経済産業省　製造産業局　化学物質安全対策室）

任期：平成 21 年 1 月 1 日～平成 22 年 3 月 31 日

大野泰雄（国立医薬品食品衛生研究所）
関野祐子（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）

増田光輝（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
小島　肇（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
秋田正治（日本動物実験代替法学会）
柴辻正喜（厚生労働省　医薬食品局　審査管理課　化学物質安全対策室）
実国慎一（経済産業省　製造産業局　化学物質安全対策室）

任期：平成 22 年 4 月 1 日～平成 23 年 4 月 30 日

眼刺激性試験代替法であるニワトリ摘出眼球を用いた眼刺激性試験法（ICE 法：Isolated Chicken Eye Test）について、第三者評価委員会からの報告を受け[1)]、以下の 8 項目について審議した。7 項目までは OECD ガイダンス文書　No. 34 に示された検討項目である[2)]。なお、本動物実験代替法の利用にあたっては、適用範囲を十分に配慮した上で使用されるべきである。

＜審議内容＞

１．検討対象の試験法とその妥当性を示すデータは、透明で独立な評価を受けているか。

Prinsen らの及び Balls らのグループによって GLP に準拠して行われ、論文として公開された 5 つのバリデーション試験の結果が[3-7)]、動物実験代替法に関する独立した機関である ICCVAM[注釈1]において評価された。

上記 5 試験のデータ以外にも ICE 法による報告がされているが、被験物質情報の無記載、数字データの無記載、個体データの欠如などがあり、検討対象として取り上げなかった。

この評価データをもとに、「JaCVAM 眼刺激性試験代替法評価委員会」によって評価された。

２．当該試験法で得られるデータは、対象毒性を十分に評価あるいは予測できるものであるか。データは、当該試験法と従来の試験法の、代替法としての繋がりを示しているか。あるいは（同時に）そのデータは、当該試験法と、対象としているあるいはモデルとしている動物種についての影響との繋がりを示しているか。

当該試験法は、ヒトの眼刺激性に対する毒性評価に用いられているウサギを用いた Draize 法のうち、腐食性・強刺激性評価の代替法として、ニワトリの摘出眼球に生じる角膜の傷害（腫脹，混濁、蛍光色素染色性）を判定するものである。

当該試験法は、Draize 法による化学物質の眼に対する腐食性・強刺激性分類との一致度が十分である。これらの指標の変化は眼に対する非可逆的影響を現わすものであり、化学物質の眼に対する腐食性・強刺激性の判定に利用できる。

なお、ウサギとニワトリの角膜は解剖学的及び生理学的な違いがあり、かつ、当該試験法では摘出眼球を用いているため、腐食性・強刺激性以外の予測には限界がある。

３．当該試験法は、ハザードあるいはリスク、あるいはその両方を評価するのに有用であるか。

化学物質の眼に対する腐食性及び強刺激性（ハザード）を評価するために有用である。リスク評価を目的とした判定基準でなく、濃度反応性の解析も行われていない。

４．当該試験法とその妥当性を示すデータは、その試験法で安全性を保証しようとする、行政上のプログラムあるいは関係官庁が対象としている化学物質や製品を、十分広く対象としたものとなっているか。当該試験法が適用できる条件及び適用できない条件が明確であるか。

当該試験法の妥当性を示すデータは、合計 175 の化学物質または製品が試験され、うち 90 は単一の化学物質で、85 は市販品あるいは製剤など混合物で行われている。様々な化学構造、性質、性状の物質、かつ種々の刺激性強度のものが対象となっており、適用できる物質の範囲が明確である。

当該試験法は、対象とする物質の腐食性・強刺激性を多くの対象物質で評価できる。ただし、アルコール類、界面活性剤、性状が固体のものについての予測性能は不十分である。

当該試験法は、暴露直後の角膜の変化を評価する方法であり、その後の回復等の評価はできない。

５．当該試験法は、プロトコルの微細な変更に対して十分頑健で、適切な訓練経験を持つ担当者と適切な設備のある施設において、技術習得が容易なものであるか。

当該試験法は適切な訓練経験を持つ担当者と設備のある施設で実施可能である。しかし、機器の特殊性、汎用性及び試験技術の修得から、現時点で日本国内への試験法の移転は困難である。

本プロトコルでの判定は、十分頑健性が確保されている。

６．当該試験法は、時間的経費的に有用性があり、行政上で用いられやすいものであるか。

試験費用面では Draize 法と大きな違いはないが、試験期間は短縮される。

EU では ICE 法の陽性結果をもって化学物質を R41 に区分することを既に受け入れている。米国では、US EPA [注釈 2] が化学物質の眼刺激性評価において、腐食性・強刺激性物質の判断に本試験の受け入れを表明している。

当該試験法は眼刺激性試験法として、GHS [注釈 3] に準拠する腐食性・強刺激性物質の判定に有用と判断されており、上記状況から考えて、我が国でも法規制の試験法として受け入れられる。

現時点では日本国内での日常的な実施は困難であるが、国外には受託機関（1 施設）があり、委託が可能である。

７．当該試験法は、従来の試験法と比べて、科学的・倫理的・経済的に、新しい試験法あるいは改訂試験法であることが正当化されているか。

当該試験法は、化学物質による不可逆的角膜傷害を捉えており、眼の腐食性及び強刺激性を判定する試験法として科学的に正当である。

当該試験法は、Draize 法と比較して倫理的に優れている。

当該試験法は、経済的な動物実験代替法となる可能性がある。

８.安全性評価のための行政的資料として、 受け入れ可能な試験法であるか。

化学物質による直接的な腐食性・強眼刺激性を評価できる方法である。その範囲において、行政的な利用は可能である。

以上の審議の結果、JaCVAM 評価会議は、眼刺激性試験代替法(ICE 法)を定められた方法で注意点を適切に守って利用すれば、化学物質の眼刺激性を科学的に評価できると結論した。

注釈

1. ICCVAM: Interagency Coordinating Committee on the Validation of Alternative Methods, USA
2. US EPA: United States Environmental Protection Agency
3. GHS: Global Harmonized System of Classification and Labeling of Chemicals

参考文献

１）眼刺激性試験代替法の第三者評価報告書　ニワトリ摘出眼球を用いる眼刺激性試験

２）OECD (2005) OECD Series on testing and assessment Number 34, Guidance document on the

validation and international acceptance of new or updated test methods for hazard assessment, ENJ/JM/MONO(2005) 14
3) Prinsen and Koeter. (1993)
4) Balls et al. (1995)
5) Prinsen. (1996)
6) Prinsen. (2000)
7) Prinsen. (2005)

眼刺激性試験代替法の第三者評価報告書

評価対象試験：眼に対する腐食性および強刺激性評価のための ニワトリ摘出眼球を用いた眼刺激性試験法

Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants

Version 2

平成 21 年 10 月 14 日

眼刺激性試験代替法評価委員会

委員長　簾内　桃子（国立医薬品食品衛生研究所）
委　員　小坂　忠司（残留農薬研究所）
　　山本　直樹（藤田保健衛生大学）
　　増田　光輝（国立医薬品食品衛生研究所）
　　竹内　小苗（P&G イノベーション合同会社）
　　宮岡　悦良（東京理科大学）

略語

3Rs: Replacement, Reduction, and Refindment Alternative
BCOP: Bovine Corneal Opacity and Permeability
BRD: Background Review Document
CAS: Chemical Abstracts Service
CV: Coefficient of variation
ECVAM: European Center for the Validation of Alternative Methods
ESAC: ECVAM Scientific Advisory Committee
EU: European Union
GHS: Global Harmonized System
GLP: Good Laboratory Practices
HBSS: Hank's Balanced Salt Solution
ICCVAM: Interagency Coordinating Committee on the Validation of Alternative Methods
ICE: Isolated Chicken Eye Test Method
IRE:Isolated Rabbit Eye Test Method
IVIS: *In Vitro* Irritancy Score
MAS: Maximum Average Score
MMAS: Modified Maximum Average Score
NICEATM: Interagency Center for the Evaluation of Alternative Toxicological Methods
OECD: Organization for Economic Co-operation and Development
US EPA: United States Environmental Protection Agency

要旨

ニワトリ摘出眼球を用いた眼刺激性試験（ICE：Isolated Chicken Eye Test）は被験物質の眼刺激性を評価する試験法であり、ウサギを用いた眼刺激性試験（Draize 法）の代替法として開発された。本試験法を腐食性および強刺激性物質をスクリーニングする目的で使用するために行われた ICCVAM（Interagency Coordinating Committee on the Validation of Alternative Methods）におけるバリデーション試験の情報（BRD: Background Review Document）をもとに、JaCVAM 眼刺激性試験代替法評価委員会においても、本試験法についての Peer Review を実施した。

ICE 法は、ニワトリから摘出した眼球に被験物質を暴露し、その結果、眼球に生じる物理的特性の変化を、角膜の腫脹、角膜混濁および角膜のフルオレセイン染色性を経時的に評価することによって、眼の腐食性および強刺激性を判定する方法である。

本バリデーション試験には、様々な種類と十分な数の被験物質が用いられた。ICCVAM の BRD によると、ICE 法の正確性は、GHS 分類による腐食性・強刺激性の眼刺激性分類と比較して、その一致度は 83％であり、偽陽性率と偽陰性率はそれぞれ 8%および 50%であった。一方、被験物質には分類により、偽陽性率および偽陰性率とも高い区分（アルコール類）、また偽陰性率の高い区分（固体、界面活性剤）がみられた。これら特定の区分を除くと、一致度および偽陽性率は 92%と 6%であり、偽陰性率は 7 例中 2 例の 29％となり、腐食性・強刺激性の検出精度は良好と判断された。試験法の信頼性については、施設内変動の検討は十分といえないものの、施設間変動については良好な結果が得られていると判断された。

以上のような結果から、ある特定の化学物質（アルコール類、固体、界面活性剤）の特性を考慮に入れた上で、化学物質の眼刺激性の段階的評価の１つとして、腐食性および強刺激性物質を評価する目的のために、ICE 法を使用することに問題はないと判断された。我が国の GHS に準拠する化学物質に関わる法規制において、ICE 法により腐食性・強刺激性物質を評価することは可能であると考えられる。

1　試験法の科学的および規制上での妥当性

ニワトリ摘出眼球を用いた眼刺激性試験 ICE は、ニワトリから摘出した眼球の角膜に被験物質を暴露し、その結果、角膜に生じる物理的特性の変化から被験物質の眼刺激性を評価する試験法である。

角膜は偶発的な事故などにより刺激物に暴露される眼表面組織の広範囲を占めており、その損傷は視力障害を引き起こす可能性がある。したがって、従来の眼刺激性試験評価法である Draize 法では、角膜への影響に評価の重みをおいている。ニワトリから採取した眼球を用いる本試験法も、Draize 法と同様な考え方に基づいて化学物質の眼刺激性を評価していると判断される。

ヒト、ウサギおよびニワトリの角膜には解剖学的および生理学的な違いがあり、この違いが Draize 法や ICE 法を用いた場合のヒト眼刺激性の予測性におよぼす影響については明らかではない。しかしながら、現在、Draize 法を使用することで、化学物質がヒトの眼に対して重篤な損傷を与える可能性については十分予知できている。したがって、ICE 法の妥当性を検討するにあたっては、Draize 法との比較を行うことで、その目的が達成されると考えられる。

生体を用いた試験では、化学物質の眼の暴露に対する保護作用が働くが、ICE 法ではこの作用は組み入れられないため、より過酷な試験条件下で評価していると考えられる。また、暴露後の角膜の刺激性反応は評価できるが、回復性や角膜以外の眼組織については、評価できない。しかし、腐食性・強刺激性の判定が微妙な場合は、病理組織学的評価を参考とすることが可能である。

化学物質の危険有害性の情報については、ラベルや安全性データシートを通じて使用者に伝達されるような法律や規則が各国において定められている。眼刺激性については、現在、Draize 法による腐食性（非回復性）の有無、さらに、回復性の刺激についてはその重症度（強･中･弱刺激性）等をもとに、米国環境保護庁（US EPA）、欧州連合（EU）、および化学品の分類と表示に関する世界調和システム（GHS）において分類基準が示されている（EPA 1996, EU 2001, UN 2003）。

ICCVAM のバリデーション試験（ICCVAM 2006）では、US EPA、EU、GHS のそれぞれの分類基準に応じて、腐食性および強刺激性物質の判定法としての評価を行っている。わが国においては、化学物質に関する法律のうち、2009 年 1 月現在、労働安全衛生法が GHS を導入している。また、今後、化学物質の流通の更なるグローバル化を考えると、危険有害情報の共有化が必要となり、国内の他の化学物質に関する法律も GHS を導入していくことが推測される。よって、ICE 法の GHS 分類基準に対する評価は、国内法への適用にも対応するものである。

2　試験法のプロトコールの妥当性

本試験は、以下の３つの評価項目をもとに、眼の腐食性・強刺激性を評価する。

1. 角膜の腫脹(Corneal swelling)：光学的厚度計（Optical Pachymeter）を装着した細隙灯顕微鏡にて、角膜の厚さを測定し、その変化を定量的に評価する。
2. 角膜混濁（Corneal opacity）：細隙灯顕微鏡にて角膜の濁度すなわち変性を、評価する。
3. 角膜のフルオレセイン染色性（Fluorescein retention）：細隙灯顕微鏡にてフルオレセイン染色性を検査する。

この３つの評価値を総合し、眼の腐食性および強刺激性を判定する。ICCVAM のバリデーション試験で用いられたプロトコールの概要は、以下のとおりである。

１）眼球の入手

主に食用目的で屠場にて処分された若鳥（7 週齢程度で 2.5～3.0 kg、雌雄の区別無し）から頭部を入手する。屠殺後の鶏頭部は、生理食塩液で湿潤させた紙を敷いた上に置き、箱に入れて室温で運搬する。運搬は屠殺後 2 時間以内に行うことが指定されている。

２）眼球の準備

角膜の損傷をフルオレセイン染色により確認後、鶏頭部から眼球を取り出し、ステンレス製のクランプ（固定器）に固定して Corneal Swelling（角膜の腫脹）を検査し、腫脹が平均値の 10％を逸脱する眼球を除く。32±1.5℃の生理食塩水を用いて 1 分間に 0.1～0.15 mL (2～3 滴) の流速で眼球を湿潤させる。

３）試験群：

各試験群あたり３個の眼球を使用する。被験物質は、基本的に固体、液体とも原体および原液のままで使用する。液体被験物質の希釈が必要な場合は、生理用食塩水あるいは蒸留水を用いる。陰性対照として、生理食塩水あるいは蒸留水を用いる。

４）暴露法：

ICE 法検査装置の固定器に設置した眼球を上向きにし、被験物質が液体の場合には 30 mL、固体の場合には 30 mg を暴露させ、10 秒後に 20 mL の生理食塩水にて洗い流す。

５）測定：

5-1）角膜の腫脹(Corneal swelling)

暴露後 30 分、75 分、120 分、180 分、240 分に、光学的厚度計(Optical Pachymeter)を装着した細隙灯顕微鏡にて角膜の厚さを測定し、変化した角膜の厚さをパーセントで表示し、以下に示すⅠ～Ⅳ段階のカテゴリーに分類する。

平均最大角膜腫脹（%）#	カテゴリー
0-5	Ⅰ
6-12	Ⅱ
13-18（75分以降）	Ⅱ
13-18（75分以内）	Ⅲ
19-26	Ⅲ
27-32（75分以降）	Ⅲ
27-32（75分以内）	Ⅳ
33以上	Ⅳ

#：各観察時の3個の平均値の最大値

5-2）角膜混濁（Corneal opacity）

暴露前（0分）、暴露後30分、75分、120分、180分、240分に、細隙灯顕微鏡にて角膜混濁の程度を評点0～4の4段階に採点し、以下に示すⅠ～Ⅳ段階のカテゴリーに分類する。

平均最大角膜混濁#	カテゴリー
0-0.5	Ⅰ
0.6-1.5	Ⅱ
1.6-2.5	Ⅲ
2.6-4.0	Ⅳ

#：各観察時の3個の平均値の最大値

5-3）角膜のフルオレセイン染色性（Fluorescein retention）

暴露後30に、細隙灯顕微鏡にてフルオレセイン染色性を0～3の3段階に採点し、以下に示すⅠ～Ⅳ段階のカテゴリーに分類する。

平均最大フルオレスセイン染色による透過性（投与後30分）	カテゴリー
0-0.5	Ⅰ
0.6-1.5	Ⅱ
1.6-2.5	Ⅲ
2.6-3.0	Ⅳ

6）判定：

角膜腫脹の変化、また、角膜混濁スコアおよび角膜のフルオレセイン染色性スコアからカテゴリー化を行い、3種の測定項目の結果を総合して腐食性・強刺激性を判定する。

すなわち、カテゴリーⅣが2項目以上の場合、腐食性・強刺激性と判定される。

また、角膜混濁の採点が、暴露後30分の観察で3以上（2眼球）あるいはいずれの観察時期でも4（2眼球）の場合、そしてフルオレセイン染色性が重度の角膜上皮欠損（1眼球）となった場合、腐食性・強刺激性と判定される。

ICE法のプロトコールの詳細は、BRDに提示されている。1993から2005年までに実施されたICE法に関する5つのバリデーション試験では、統一プロトコールを用いていないが、これらプロトコールの主な違いは使用眼球数が異なることであり、類似したプロトコールが用いられている。これらのプロトコールでは眼刺激性試験を評価するにあたっての必要な項目を網羅している。

改良点、検討事項として以下の点が考えられる。

1）ニワトリ眼球の運搬条件

鶏頭部の運搬中の保存温度は室温とされているが、適切な温度範囲は指定されていない。氷上保存では角膜に白濁の恐れがあるので、冷蔵条件下（目安として摂氏4～10度程度）での運搬が好ましいと思われる。

2）被験物質の希釈溶媒

被験物質の希釈が必要な場合、浸透圧などの細胞への影響を考慮すると、希釈溶媒は蒸留水ではなく、生理用食塩水が適切であると考えられる。

3）陽性（腐食性・強刺激性）対照物質

Draize 法の結果と一致し、かつ、再現性の高い適切な陽性対照物質の選定が必要である。

3　バリデーション試験に用いられた物質の分類と妥当性

5つのバリデーション試験において、合計175の化学物質または製品が評価された。これらのうち、90物質は単一の化学物質であり、85物質は市販品あるいは製剤などの混合物であった。化学物質区分でまとめると、アルコール類、酸、アルカリ、ハロゲン化アシル類、アミド/アミジン類、カルボン酸類、エステル類、複素環化合物、炭化水素、無機塩、ケトン類、オニウム化合物と有機リン酸化合物などであった。製品別に分類すると、洗剤、農薬、粉末シリコーン、インク、染料、溶媒、界面活性剤、トイレ用クリーナー、感熱紙用コーティング剤などであった。

被験物質の化学物質区分および供試数：

アセテート類 1、無機塩化物 1、酸 5、無機塩 3、ハロゲン化アシル 1、無機銀/窒素化合物 1、アルコール類 15、ケトン類 4、アルデヒド類 2、ラクトン 1、アルカリ 3、脂質 1、アミド/アミジン類 7、ニトリル 1、アミノ酸 1、ニトロ化合物 1、ホウ素化

合物 1、炭水化物 2、オニウム化合物 8、カルボン酸類 12、有機シリコーン化合物 2、エステル類 10、有機硫黄化合物 3、エーテル 1、有機金属化合物 2、複素環類 9、有機リン 1、炭化水素 5、多環式化合物 4、イミド類 2、ポリエーテル類 3、無機化合物 1、尿素化合物 1、無分類 85。

被験物質の製品分類および供試数：

接着剤 2、肥料 1、抗真菌薬 2、食品添加物 1、抗ヒスタミン剤 1、殺菌剤 1、抗感染薬 3、消毒剤 2、腐食性剤 4、光学的分割剤 1、塩素化副産物 1、 塗料 4、クリーナー 8、殺虫剤/除草剤 15、共重合体 3、保存剤 6、化粧品用成分 1、医薬品 5、洗剤 8、原料 9、現像液 1、試薬 4、殺菌剤 5、樹脂 2、染料 10、シリコーン樹脂 1、エラストマー類 2、石けん 9、酵素阻害物質 1、界面活性剤 25、酵素溶液 3、溶媒 37、工業用化学物質・中間体・製剤 20、無分類 23。

これらのバリデーション試験に用いられた被験物質の数と種類（物質区分、製品区分、液体・固体区分、刺激性の程度など）は、十分であると判断される。

4　試験法の正確性を評価するために用いられた化学物質の *in vivo* および参照データ

現在入手できる化学物質のヒトに対する眼刺激性のデータは十分とは言いがたい。試験データとしてあるものは、そのほとんどが弱刺激性物質である。事故により強刺激性物質に暴露された報告はあるが、詳細については不明である。したがって、現段階においてヒトに対する眼刺激性のデータは、参照データとしては適切ではないと考えられる。

今回のバリデーション試験では、ウサギを用いた眼刺激性試験（Draize 法）のデータが参照として用いられた。Draize 法は、OECD テストガイドラインが作成されており（OECD TG 405）、我が国でも使用されている方法である。Draize 法では、ウサギの眼粘膜に化学物質を暴露させ、細隙灯顕微鏡などを用いて暴露後少なくとも 72 時間まで肉眼的に観察し、角膜、虹彩および結膜の刺激性程度を採点する。角膜混濁の採点に重みづけをしており、観察時間ごとに Maximum Average Score（MAS）や Modified Maximum Average Score（MMAS）を算出し、眼刺激性程度を評価する。

Draize 法のデータについては、既存の試験結果、または ICE 試験と並行して実施された結果が用いられている。その多くは GLP に準拠した試験であったため、GHS、EPA、EU のいずれかの分類基準による評価が可能なデータをバリデーションに用いた。この中で GHS の分類基準は、我が国において化学物質の眼刺激性・腐食性の評価・分類基準として採用されているため、分類基準において強度の刺激性に相当する区分 1 については以下に示す。

Draize 法については、ヒトと比較した場合の正確性や試験法の信頼性について検討されている。弱刺激性から中刺激性の物質に対しては、ヒトと同様の反応が確認されており（McDonald et al. 1987）、強刺激性物質については、チオグリコール酸で同様な反応が報告されている（Grant 1974, Butscher 1953）。一方、ヒトとウサギでは異なる反応を示したケースも報告されている（McDonald et al. 1987）。しかしながら、現在、Draize 法により、化学物質がヒトの眼に対して重篤な損傷を与える可能性について十分予知できており、腐食性・強刺激性を判定することを目的とした ICE 法の評価において、Draize 法を参照データとして用いることに問題はないと判断できる。

GHS の分類基準（眼に対する刺激性区分）

区分 1（眼に対する非可逆的影響）：

少なくとも 1 匹の動物で角膜、虹彩または結膜に対して、可逆的であると予測されない作用が認められる、または通常 21 日間の観察期間中に完全には回復しない作用が認められる、または試験動物 3 匹中少なくとも 2 匹で、試験物質滴下後 24、48 および 72 時間における評価の平均スコア計算値が角膜混濁≧3 または虹彩炎＞1.5 で陽性反応が得られる。

5　試験法のデータと結果の利用性

以下の 5 つのバリデーション試験から角膜の厚み・腫脹、角膜混濁およびフルオレセイン染色性のデータを得ている。

Prinsen and Koeter. (1993)	21 物質供試
Balls et al. (1995)	59 物質供試
Prinsen. (1996)	44 物質供試
Prinsen. (2000)	4 物質供試
Prinsen. (2005)	50 物質供試

全てのバリデーション試験は GLP に準拠して行われたことが確認されており、また、被験物質のコード化（Coded substances）が Balls et al. (1995) および Prinsen. (2005) のバリデーション試験で実施されている。

6　試験法の正確性

正確性の評価は、GHS（UN 2003）、EPA（EPA 1996）および EU（EU 2001）各法規制の分類基準ごとに行われている。Draize 法のデータと比較した場合、ICE 法は眼に対する腐食性・強刺激性の判定において、表 1 に示したような結果を得た。

表１　腐食性・強刺激性物質の予見に関しての ICE 法の精度結果

（ウサギを用いた試験データを基に GHS、EPA、EU の各基準で分類した場合との比較）

	一致度		感度		特異度		偽陽性率		偽陰性率	
	%	n	%	n	%	n	%	n	%	n
GHS分類	83	120/144	50	15/30	92	105/114	8	9/114	50	15/30
EPA分類	84	122/145	52	15/29	92	107/116	8	9/116	48	14/29
EU分 類	87	134/154	59	19/32	94	115/122	6	7/122	41	13/32

（個々のバリデーション試験の結果を、化学物質ごとでまとめ、最も多く分類された刺激カテゴリー、あるいは最も重度の刺激カテゴリーを判定結果として選択し、各法規制の分類基準と比較した。）

一　致　度：正確な結果（陽性・陰性）の比率
感　　度：全陽性物質中の陽性結果の比率
特　異　度：全陰性物質中の陰性結果の比率
偽 陽 性 率：全陰性物質中の陽性結果（偽陽性）の比率
偽 陰 性 率：全陽性物質中の陰性結果（偽陰性）の比率

また、化学物質、物質の特性（形状）、製品の区分（5 例以上の被験物質がある区分のみ）ごとにみると、全般的に一致度は高く、偽陽性率は低いが、分類により偽陽性率および偽陰性率の高い区分（アルコール類）と偽陰性率の高い区分（固体、界面活性剤、農薬）が確認された。その他、液体やカルボン酸類でも比較的高い偽陰性率が示された。GHS の分類基準と比較した被験物質の分類区分ごとの一致度、偽陽性率および偽陰性率を表２にまとめた。

そこで、偽陽性率および偽陰性率の高いアルコール類、偽陰性率の高い固体物質と界面活性剤を除いた場合の ICE 法の評価結果は、以下に示すとおり、一致度が上がり偽陽性率と偽陰性率はさらに低下した。

一致度　：　92%（69/75）
偽陽性率：　 6%（ 4/68）
偽陰性率：　29%（ 2/ 7）（GHS 分類基準における結果）

以上の結果から、ICE 法においてアルコール類・固体・界面活性剤を除いた場合、眼に対する腐食性・強刺激性の検出精度は十分であると判断できる。

表２　被験物質分類ごとのGHS分類基準と比較した場合の一致度・偽陽性・偽陰性率

	一致度	偽陽性率	偽陰性率
全体	83%　(120/144)	8%　(9/114)	50%　(15/30)
化学物質区分[注1)]			
アルコール類	50%　(6/12)	50%　(5/10)	50%　(1/ 2)
アミド・アミジン類	80%　(4/ 5)	0%　(0/ 2)	33%　(1/ 3)
カルボン酸類	70%　(7/10)	0%　(0/ 3)	43%　(3/ 7)
エステル類	89%　(8/ 9)	13%　(1/ 8)	0%　(0/ 1)
複素環類	78%　(7/ 9)	0%　(0/ 3)	33%　(2/ 6)
オニウム化合物	75%　(6/ 8)	0%　(0/ 2)	33%　(2/ 6)
物質の特性(形状)			
液体	84%　(91/108)	10%　(9/90)	44%　(8/18)
固体	81%　(29/36)	0%　(0/24)	58%　(7/12)
製品区分[注2]			
界面活性剤	76%　(16/21)	0%　(0/12)	56%　(5/ 9)
農薬	73%　(8/11)	0%　(0/ 6)	60%　(3/ 5)

注１）　５例以上の被験物質がある化学物質区分のみのデータを抽出した

注２）　主な製品のみ

７　試験法の信頼性

施設内変動（Intralaboratory repeatability, Intralaboratory reproducibility）：

施設内変動では、４物質（界面活性剤２物質、Siloxane２物質）について、同一施設内で4～5回試験を繰り返し、角膜の厚さ、角膜の腫脹、角膜混濁および角膜のフルオレセイン染色性の各項目において、施設内の反復性と再現性が検討された（Prinsen, 2000）。供試物質は、EU分類で非刺激性１物質、刺激性（R36）２物質および強度刺激性（R41）１物質の４物質であった。

反復性解析において実験内で比較した場合、角膜の厚さのCV値は0.9～6.1%の範囲であった。角膜混濁のCV値は0～86.6%（最高値は非刺激性物質データ）、その他の角膜の腫脹およびフルオレセイン染色性においても非刺激性物質データの変動が大きかったが、非刺激性物質を除いた場合のCV値はいずれにおいても低下した（表３）。

表３　ICE法の施設内変動（反復性：Intralaboratory repeatability）

被験物質	４物質でのCV値	３物質でのCV値 （非刺激物質を除く）
角膜の厚さ	0.9 ～ 6.1	1.5 ～ 6.1
角膜の腫脹	-86.6 ～ 346.4	9.5 ～ 49.5
角膜の混濁	0 ～ 86.6	0 ～ 43.3
角膜の染色性	0 または 86.6	0

実験間での再現性について比較した場合、角膜の厚さの CV 値は 1.8～6.3%の範囲であった。角膜混濁の CV 値は 8.7～95.8%（最高値は非刺激性物質データ）の範囲であり、その他の角膜の腫脹およびフルオレセイン染色性においても非刺激性物質データの変動が大きく、非刺激性物質を除いた場合の CV 値はいずれにおいても低下した（表４）。

表４　ICE 法の施設内変動（再現性：Intralaboratory reproducibility）

被験物質	4 物質での CV 値	3 物質での CV 値 （非刺激物質を除く）
角膜の厚さ	1.8 ～ 6.3	4.0 ～ 6.3
角膜の腫脹	15.2 ～ 138.7	15.2 ～ 22.4
角膜の混濁	8.7 ～ 95.8	8.7 ～ 18.1
角膜の染色性	0 または 141.4	0

施設間変動（Interlaboratory reproducibility）：

59 の被験物質について、4 施設による施設間バリデーション試験を実施した（Balls et al., 1995）。Draize 法にて陽性と判定された物質は 22 物質であり、そのうち 11 物質は ICE 法で陽性と判断され、4 施設で一致した陽性物質は 7 物質（64%）、3 施設以上で一致した物質は 10 物質（91%）となった。同様に、偽陰性において 4 施設で一致した偽陰性物質は 11 物質中 9 物質（82%）、3 施設以上で一致した偽陰性物質は 11 物質中 11 物質（100%）であった。次に、Draize 法にて陰性と判定された物質において、4 施設で一致した陰性物質は 26 物質中 22 物質（85%）であり、3 施設以上で一致した陰性物質は 26 物質（100%）であった。

表５　ICE 法の施設間検証

（GHS 分類を用いた眼に対する腐食性・強刺激性あるいは非刺激性の *in vivo* データの比較）

眼刺激性分類 （*in vivo* / *in vitro*）	被験物質数	試験施設数	4 施設で一致した物質数 (%)	3 施設以上で一致した物質数 (%)
＋ / ＋	11	4	7 (64%)	10 (91%)
＋ / －	11	4	9 (82%)	11 (100%)
－ / ＋	6	4	1 (17%)	1 (17%)
－ / －	26	4	22 (85%)	26 (100%)
n.d. / ＋	3	4	3 (100%)	(－)
n.d. / －	2	4	2 (100%)	(－)
総　数	59	4	44 (75%)	53 (90%)

（n.d. ：データ無し）

ICE 試験データのみを用いて陽性（腐食性・強刺激性）および陰性（腐食性・強刺激性以外）判定の施設間における一致度の比較を行った（表６）。75～76%の被験物質が、

4 施設間で一致した。また、88％の被験物質が、4 施設中 3 施設以上で一致した。一致度が低いものは、アルコール、エステルおよびケトン類であった。

表６　腐食性・強刺激性物質の予見に関しての ICE 法の施設間一致度のばらつき

(GHS、EPA、EU 分類を用いた眼に対する腐食性・強刺激性判定の一致度比較)

	試験施設数	4 施設で一致 (一致度；100％)		3 施設で一致 (一致度：75％)		2施設で一致 (一致度：50％)	
		n	%	n	%	n	%
GHS分類	4	44	75	8	14	7	12
EU 分類	4	44	75	8	14	7	12
EPA分類	4	45	76	7	12	7	12

（被験物質総数；59、　n ；該当する被験物質数）

以上の結果から、施設内変動では角膜腫脹の指標（角膜の厚さ）がいずれの物質でも CV 値が低く、再現性が認められた。その他の指標では非刺激性物質データに変動が大きかった。非刺激性物質のデータを除いた場合の CV 値は低下したが、わずか 4 被験物質で実施された施設内変動についての検討は不十分であった。

施設間変動では、4 施設中 3 施設以上で陽性、偽陰性、陰性の再現性が確認され、Draize 法との比較で 91％、ICE 法のみの比較でも 88％を示し、良好な結果が得られていると判断された。

8　試験法のデータの質

バリデーション試験に用いられた以下の報告について、ICE 法と Draize 法の試験の GLP 準拠を示す。

- Prinsen and Koeter. (1993)

  ICE 法は GLP 準拠、Draize 法は GLP 準拠の宣誓は無い。被験物質はコード化されていない。

- Balls et al. (1995)

  ICE 法は GLP 準拠、Draize 法は GLP 準拠で OECD の試験ガイドライン TG405 に従い実施された。被験物質はコード化された。

- Prinsen. (1996)

  ICE 法および Draize 法とも GLP 準拠で実施された。被験物質はコード化されていない。

- Prinsen. (2000)

  ICE 法および Draize 法とも GLP 準拠で実施された。被験物質はコード化されていな

い。

- Prinsen. (2005)

  ICE 法および Draize 法とも GLP 準拠で実施された。被験物質はコード化された。

9　試験法の科学的な報告

バリデーション試験に用いられた 5 試験以外にも ICE 法による報告がされているが、被験物質情報の無記載、数字データの無記載、個体データの欠如など、バリデーション試験としては不適当な報告も見られた。

以下の 3 試験では *in vitro* の ICE 法と *in vivo* の Draize 法との関連性を検討し、ピアソンの相関解析法にて相関係数 (r) が評価されている。

- Balls et al. (1995)

  ICE 法の Index Score において、全被験物質で相関係数は 0.490～0.599 であり、特に界面活性剤では相関性は 0.724～0.833 と良好でであった。

- Chamberlain et al. (1997)

  角膜、虹彩、結膜などの採点における相関係数は 0.89～0.97 であった。

- Prinsen. (1996)

  角膜、虹彩、結膜などの採点で相関係数は 0.86～0.92 であった。

Procter and Gamble (P&G) 社より、28 物質について TNO Nutrition and Food Institute で実施された資料が提供され、*in vitro* の ICE 法と *in vivo* LVET (Low Volume Eye Irritation Test: 少容量法) の Draize 法において、EU 分類での評価が検討された。5% SLS（ラウリル硫酸ナトリウム）ではいずれも非刺激性、10％塩化ベンザルコニウムではいずれも R41（強刺激性）と判定され、その他の物質（界面活性剤、洗剤など）では 10 物質で ICE 法と Draize 法のデータが一致、5 物質で不一致であった。

10　3Rs への関与（動物福祉面としての妥当性）

ICE 法で使用されている眼球試料は、試験目的ではなく食用として屠殺されたニワトリの眼球を用いているため、試験目的だけの実験動物の使用を抑えることができる（reduction）。また、従来の Draize 法と比較して、試験操作による動物への苦痛は無い（refinement）。加えて、ICE 法で陽性と判断された場合には追加の動物試験を行う必要がなくなることから、化学物質の眼刺激性評価全体において不必要な動物試験を回避できる（reduction）など、眼刺激試験のためにウサギを用いる従来の Draize 法と比較して動物福祉面からも評価できる。この意味では、動物が受ける苦痛の除去（refinement）、使用動物数の削減（reduction）および不必要な動物試験の回避（reduction）が達成できる。

１１　試験法の有用性と限界

ICE 法の評価項目において、Corneal Swelling（角膜の腫脹）については光学的厚度計を用いた客観的な指標であることから、Draize 法と比較しても、試験実施施設間での誤差を軽減することができると考えられる。眼刺激性強度カテゴリーの境界線上の判断が要求される際には、Morphological effects（組織学的変化）を参考にすることも可能である。

試験法の移転性については、機器の特殊性、汎用性および試験技術の修得から、現時点では日本国内での日常的な実施は困難である。ただし、国外には ICE 法による眼刺激性試験を実施する受託機関（1 施設）があり、日本の科学者や企業からの委託が可能である。

ICE 法の限界として、アルコール類に対する偽陽性率と偽陰性率が高いこと、また固体、界面活性剤および農薬に対する偽陰性率も高いことから、これらの物質を評価する際には注意が必要である。

費用面では、従来の Draize 法と大きな違いはない。しかし、病理組織学的観察を考えた場合は、更なる出費が予想される。

試験期間については観察期間は短縮されるが、病理組織学的観察を行う場合は、従来の Draize 法と大きな違いはない。

１２　結　論

ICCVAM で実施された ICE 法の第三者評価はバリデーション試験に必要な項目、プロセス、データが検討されている。ESAC の評価と同様、ICCVAM のバリデーション試験の結果を受け入れることに問題はないと判断される。

特定の化学物質（アルコール類、固体、界面活性剤）ではその精度が十分ではない等の試験の限界を考慮に入れた上で、適切なプロトコールに基づき試験を実施すれば、ICE 法は、化学物質の眼刺激性の段階的評価の１つとして、腐食性・強刺激性物質をスクリーニングすることが可能であると判断される。

現在、EU では ICE 法の陽性結果をもって化学物質を R41 に区分することを既に受け入れている。また、米国では、FDA・EPA が化学物質の眼刺激性評価において、腐食性・強刺激性物質の判断に ICE 法の結果を受け入れることを公式に発表している。

わが国においても、GHS に準拠する化学物質に関わる法規制において、ICE 法による腐食性・強刺激性物質の眼刺激性を評価することが可能である。

１３　文献

Balls, M., Botham, P.A., Bruner, L.H. and Spielmann, H. (1995). The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicol In Vitro* 9:871-929.

Butscher, P. (1953). Beitrag zur therapie von augenschadigunen durch thioglykolsaur bei der herstellung der sogenannten kaltwelle. *Klin Monatsbl Augenheilkd* 122:349-350.

Chamberlain, M., Gad, S.C., Gautheron, P. and Prinsen, M.K. (1997). IRAG Working Group I: Organotypic models for the assessment/prediction of ocular irritation. *Food Chem Toxicol* 35:23-37.

ESAC (2007). ESAC statement on the conclusions of the ICCVAM retrospective study on Organotypic *in vitro* assays as screening tests to identify potential ocular corrosives and severe eye irritants as determined by US EPA, EU(R41) and UN GHS classifications in a tiered testing strategy, as part of a weight of evidence approach.

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

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.

Grant, W.M. (1974). Toxicology of the eye. 2nd ed. Springfield, IL: Charles C. Thomas.

ICCVAM. (2006). Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method. NIH Publication No.: 06-4513. Research Triangle Park: National Toxicology Program.

McDonald, T.O., Seabaugh, V., Shadduck, J.A. and Edelhauser, H.F. (1987). Eye

irritation. In Dermatotoxicology (Marzulli, F.N. and Maibach H.I., eds). Washington: Hemi sphere Publishing Corporation, 641-696.

Nakano, M. (1958). Effect of various antifungal preparations on the conjunctiva and cornea of rabbits. *Yakuzaigaku* 18:94-99.

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

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

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

Prinsen, M.K. and Koëter, B.W.M. (1993). Justification of 419 the enucleated eye test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. *Food Chem Toxicol* 31:69-76.

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

付表 1

定義 (Definitions)

Accuracy: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes (positive and negative) 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.

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

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

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

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

Corneal Swelling: An objective measurement in the ICE test of the extent of distention of the cornea following exposure to a test substance. It is expressed as a percentage and is calculated from corneal thickness measurements that are recorded at regular intervals during the ICE test. Increased corneal swelling is indicative of damage to the corneal epithelium.

False negative rate: 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 rate: The proportion of all negative substances that are falsely identified by a test method as positive. It is one indicator of test method accuracy.

Fluorescein retention: A subjective measurement in the ICE test of the extent of fluorescein sodium that is retained by epithelial cells in the cornea following exposure to a test substance. Increased fluorescein retention is indicative of damage to the corneal epithelium.

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

Good Laboratory Practices (GLP) : 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.

Interlaboratory reproducibility: 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: 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: 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.

Negative control : An untreated sample containing all components of a test system, except the test substance solvent, which is replaced with a known non-reactive 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: 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 methodaccuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

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 substancetreated 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: 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.

Sensitivity: 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).

Specificity: The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy

Tiered testing: A stepwise testing strategy where all existing information on a test substance is reviewed, in a specified order, using a weight of evidence process at each tier to determine if sufficient information is available for a hazard classification decision, prior to progression to the next tier. 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 sequential animal testing procedure is performed until an unequivocal classification can be made.

ICCVAM TEST METHOD EVALUATION REPORT: *IN VITRO* OCULAR TOXICITY TEST METHODS FOR IDENTIFYING SEVERE IRRITANTS AND CORROSIVES

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

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

[This Page Intentionally Left Blank]

ICCVAM TEST METHOD EVALUATION REPORT:

IN VITRO OCULAR TOXICITY TEST METHODS FOR IDENTIFYING SEVERE IRRITANTS AND CORROSIVES

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

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

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

September 2006
NIH Publication No.: 06-4511

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

[This Page Intentionally Left Blank]

Printed: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

°C	Degrees Centigrade
BCOP	Bovine Corneal Opacity and Permeability
BRD	Background Review Document
CAM	Chorioallantoic Membrane
CV	Coefficient of Variation
ECVAM	European Center for the Validation of Alternative Methods
EPA	U.S. Environmental Protection Agency
EU	European Union
FR	*Federal Register*
g	Gram
GHS	Globally Harmonized System
HET-CAM	Hen's Egg Test-Chorioallantoic Membrane
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye
IRE	Isolated Rabbit Eye
IS	Irritation Score
MeSH	Medical Subject Headings
mL	Milliliter
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NTP	U.S. National Toxicology Program
OTWG	Ocular Toxicity Working Group
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods
UN	United Nations
UV/VIS	Ultraviolet/Visible

INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS - AGENCY REPRESENTATIVES

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

Consumer Product Safety Commission
• Marilyn L. Wind, Ph.D. (Vice-Chair)
* Kailash C. Gupta, D.V.M., Ph.D.
* Patricia Bittner, M.S.
* Kristina Hatlelid, Ph.D.

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

Department of Defense
• Robert E. Foster, Ph.D.
◊ Patty Decot
* Harry Salem, Ph.D.

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

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

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

Environmental Protection Agency
Office of Science Coordination and Policy
• Karen Hamernik, Ph.D.
Office of Research and Development
◊ Julian Preston, Ph.D.
* Suzanne McMaster, Ph.D.
OECD Test Guidelines Program
* Jerry Smrchek, Ph.D.
Office of Pesticides Programs
* Amy Rispin, Ph.D.
* Deborah McCall

• Principal Agency Representative
◊ Alternate Principal Agency Representative
* Other Designated Agency Representatives

8-25-2006

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

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

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

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

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

National Library of Medicine
• Vera Hudson, M.S.
◊ Jeanne Goshorn, M.S.

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

ACKNOWLEDGMENTS

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

Department of Defense
Harry Salem, Ph.D.

Department of Transportation
Steve Hwang, Ph.D.

Environmental Protection Agency
Karen Hamernik, Ph.D. (OTWG, Co-Chair)
Meta Bonner, Ph.D.
Andrew Geller, Ph.D.
Karen Hicks
Marianne Lewis
Deborah McCall
Mark Perry, Ph.D.
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration
Jill Merrill, Ph.D. (OTWG, Co-Chair)
Robert Bronaugh, Ph.D.
Paul Brown, Ph.D.
Wiley Chambers, M.D.
Abigail Jacobs, Ph.D.
Donnie Lowther
Leonard Schechtman, Ph.D. (ICCVAM, Chair)

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

National Institute of Environmental Health Sciences
Raymond Grissom, Ph.D.
William Stokes, D.V.M., D.A.C.L.A.M. (Executive Director, ICCVAM)
Raymond Tice, Ph.D.

In Vitro Ocular Test Methods Expert Panel

Robert Scala, Ph.D. (Panel Chair), Consultant, Tucson, Arizona

Sally Atherton, Ph.D., Professor, Medical College of Georgia, Augusta, Georgia
Roger Beuerman, Ph.D., Professor, Louisiana State University, New Orleans, Louisiana
June Bradlaw, Ph.D., International Foundation for Ethical Research, Rockville, Maryland
Ih Chu, Ph.D., Health Canada, Ottawa, Canada
Henry Edelhauser, Ph.D., Professor, Emory University, Atlanta, Georgia
Nancy Flournoy, Ph.D., Professor, University of Missouri, Columbia, Missouri
Donald Fox, Ph.D., Professor, University of Houston, Houston, Texas
James Freeman, Ph.D., Section Head, ExxonMobil Biomedical Sciences, Inc., Annandale, New Jersey
Shayne Gad, Ph.D., D.A.B.T., A.T.S., Consultant, Gad Consulting Services, Cary, North Carolina
Sidney Green, Ph.D., A.T.S., Graduate Professor, Howard University, Washington, DC
Frederick Guerriero, M.S., Senior Occupational Toxicologist, GlaxoSmithKline, King of Prussia, Pennsylvania
A. Wallace Hayes, Ph.D., D.A.B.T., F.A.T.S., F.I.Biol., F.A.C.F.E., E.R.T., Scientist, Harvard School of Public Health, Andover, Massachusetts
Hiroshi Itagaki, Ph.D., Principal Scientist, Shiseido Co., Ltd., Japan
David Lovell, Ph.D., Reader in Medical Statistics, University of Surrey, United Kingdom
Yasuo Ohno, Ph.D., D.J.S.T.S., Director of Japanese Society of Alternatives to Animal Experiments and Director of Division of Pharmacology, National Institute of Health Sciences, Japan
Robert Peiffer, D.V.M., D.A.C.V.O., Senior Investigator, Merck Research Laboratories, West Point, Pennsylvania
Lionel Rubin, V.M.D., D.A.C.V.O., Emeritus Professor of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania
Horst Spielmann, Dr. Med., Director and Professor, ZEBET at the BfR, Berlin, Germany
Martin Stephens, Ph.D., Vice President for Animal Research, Humane Society of the United States, Washington DC
Katherine Stitzel, D.V.M., Consultant, West Chester, Ohio
Peter Theran, V.M.D., D.A.C.V.I.M., Vice President Animal Science, Massachusetts Society for the Prevention of Cruelty to Animals, Novato, California
Scheffer Tseng, M.D., Ph.D., Director, Ocular Surface Research and Education Foundation, Miami, Florida
Philippe Vanparys, Ph.D., Senior Research Fellow, Johnson and Johnson, Belgium

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

David Allen, Ph.D.
ILS, Inc.

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

Sue Brenzel
ILS, Inc.

Thomas Burns, M.S.
ILS, Inc.

Patricia Ceger, M.S.
ILS, Inc.

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

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

William Stokes, D.V.M., D.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 Contributors

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

PREFACE

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 (2000); available at http://iccvam.niehs.nih.gov/about/PL106545.pdf) with evaluating the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. Following such evaluations, ICCVAM is required to provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of such methods.

In October 2003, the U.S. Environmental Protection Agency (EPA) formally nominated several ocular toxicity test method activities to ICCVAM. ICCVAM determined that four *in vitro* test methods proposed for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy should have the highest priority for evaluation. This was based on the availability of existing validation data for all four methods and the fact that determining the adequacy of validation[1] is a prerequisite for test methods 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 (ICE) assay, and the Isolated Rabbit Eye (IRE) assay.

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 the test method evaluations. ICCVAM and NICEATM also collaborated closely with the European Centre for the Validation of Alternative Methods (ECVAM) in conducting the evaluations, with Drs. Chantra Eskes and Valérie Zuang serving as ECVAM liaisons to the OTWG.

NICEATM, in conjunction with the OTWG, prepared four comprehensive background review documents (BRDs) reviewing the available data and information for each of the four *in vitro* test methods. Each BRD described the current validation status of the *in vitro* test method, including its reliability and accuracy, the scope of the substances tested, and the availability of a standardized protocol. The BRDs were based on published studies using the respective test method, and other data and information submitted in response to a 2004 public call for information. The draft BRDs were made available to the public for comment on November 1, 2004, and a public independent expert panel meeting also was announced.

The ICCVAM organized an international independent Expert Panel meeting on January 11-12, 2005, to assess the validation status of these four *in vitro* test methods for identifying ocular corrosives or severe irritants. While a comprehensive review was conducted, public comments at the meeting revealed that additional relevant data were available that had not yet been provided in response to earlier requests for data. Accordingly, the Expert Panel recommended that if such data could be obtained, a reanalysis of each test method should be

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

performed. Availability of the Expert Panel's independent report was announced on March 21, 2005.

In response to the Expert Panel's recommendation, a second public request for *in vitro* data was published on February 28, 2005. In response to this request, 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. The additional data, together with clarified rules for hazard classification and reclassification of the chemical classes of the test substances necessitated a reanalysis of the accuracy and reliability of all four test methods. The accuracy and reliability reanalyses and a revised reference substances list for validation of *in vitro* tests to detect ocular corrosives and severe irritants were provided in a BRD Addendum released on July 26, 2005.

The Expert Panel was subsequently reconvened via teleconference on September 19, 2005 to discuss the BRD Addendum. The Expert Panel provided final conclusions regarding the effects of the information in the BRD Addendum on their original evaluation from the January 11-12, 2005 meeting. The report of this meeting also was published and public comments requested.

The draft BRDs, draft BRD Addendum, Expert Panel report and addendum, and all public comments were subsequently made available to the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) for comment at their meeting on December 12, 2005. The SACATM concurred with the consensus conclusions of the Expert Panel.

The ICCVAM and OTWG considered the Expert Panel report and addendum, the revised accuracy and reliability analyses, all public comments, and the comments of SACATM in preparing the final ICCVAM test method recommendations provided in this report. This report will be made available to the public and provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]) (Available at http://iccvam.niehs.nih.gov/about/PL106545.pdf). Agencies with applicable testing regulations and/or guidelines must respond to ICCVAM within 180 days after receiving the ICCVAM recommendations. These responses will be made available to the public on the ICCVAM website (http://iccvam.niehs.nih.gov) as they are received.

In this Test Method Evaluation Report, ICCVAM states that there are sufficient data to substantiate the use of BCOP and ICE test methods, with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants in a tiered-testing strategy, using a weight-of-evidence approach. When used in this manner, these methods should reduce the number of animals needed for ocular toxicity testing and refine animal use by avoiding the pain and distress associated with testing severely irritating and corrosive substances. Since ocular irritancy testing may involve more than slight or momentary pain or distress, alternative test methods must be considered prior to the use of animals, as required by U.S. Federal animal welfare regulations and policies. Accordingly, *in vitro* alternative test methods should be considered prior to *in vivo* ocular testing and used where determined appropriate for a specific testing situation. Consistent with the mission of ICCVAM,

appropriate use of these methods will support improved animal welfare while ensuring the continued protection of human health.

Acknowledgments
The efforts of many individuals who contributed to the preparation, review, and revision of this report are gratefully acknowledged. We especially recognize all of the Expert Panel members for their thoughtful evaluations and generous contributions of time and effort. Special thanks are extended to Dr. Robert Scala for serving as both the Panel Chair and a Group Chair and to Drs. Kathy Stitzel, James Freeman, and Shayne Gad for their service as Group Chairs. The efforts of the OTWG were invaluable for assuring a meaningful and comprehensive review. We especially would like to thank the co-chairs of the OTWG, Drs. Jill Merrill (U.S. Food and Drug Administration) and Karen Hamernik (U.S. Environmental Protection Agency) for their leadership. The efforts of the NICEATM staff in preparing the BRDs, organizing the Expert Panel meeting and teleconference, and preparing this final report are greatly appreciated. We also acknowledge Drs. David Allen, Jeff Charles, and Neepa Choksi and Messrs. Bradley Blackard, Thomas Burns, and James Truax of Integrated Laboratory Systems (ILS), Inc., the NICEATM Support Contractor, and Dr. Raymond Tice for his initial contributions as a member of the ILS, Inc. support contract and later as a member of NICEATM.

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

[This Page Intentionally Left Blank]

EXECUTIVE SUMMARY

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recently completed the technical evaluation of the validation status of four *in vitro* ocular irritation test methods proposed as screening tests[2] for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy[3], as part of a weight-of-evidence approach. The four test methods are the Bovine Corneal Opacity and Permeability (BCOP) assay, the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay, the Isolated Chicken Eye (ICE) assay, and the Isolated Rabbit Eye (IRE) assay. The U.S. Environmental Protection Agency (EPA) formally nominated these test methods for evaluation by ICCVAM in October 2003. In addition to evaluating their current usefulness and limitations as screening tests for identifying ocular corrosives and severe irritants, ICCVAM developed a recommended standardized protocol for each test method; made recommendations, where considered appropriate, for further research and development, optimization, and/or validation efforts; and developed a list of reference substances for such activities.

None of the four *in vitro* test methods evaluated can be considered to be replacements for the *in vivo* rabbit eye test. However, based on the available data, BCOP and ICE can be used, in appropriate circumstances and with certain limitations, as screening tests for the detection of ocular corrosives and severe irritants in a tiered-testing strategy, as part of a weight-of-evidence approach. At the present time, HET-CAM, using the decision criteria of Luepke (1985), and IRE are not recommended as screening tests for the identification of ocular corrosives and severe irritants for regulatory hazard classification purposes. Before HET-CAM and IRE can be recommended for this purpose, the protocol and the decision criteria for the identification of ocular corrosives and severe irritants need to be optimized and undergo further validation.

This evaluation provides validation information that should be helpful to various stakeholders (e.g., applicable U.S. Federal regulatory agencies, the international regulatory community, the pharmaceutical, pesticide, and commercial chemical industries) in determining when these test methods might be useful and which test method might be the most appropriate for a specific testing situation. These *in vitro* test methods, when used appropriately, will reduce and refine animal use for ocular safety testing.

---

[2]According to the *ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods*, a **screen** or **screening test** is "a rapid, simple test conducted for the purposes of a general classification of substances according to general categories of hazard. The results of a screen generally are used for preliminary decision making and to set priorities for more definitive tests. A screening test may have a truncated response range (e.g., be able to reliably identify active chemicals but not inactive chemicals)" (ICCVAM 2003).

[3]A tiered-testing strategy approach may not be applicable to purposes other than regulatory classification and labeling.

Specific Test Method Recommendations

BCOP Test Method

There are sufficient data to support the use of the BCOP test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, United Nations [UN] Globally Harmonized System of Classification and Labelling of Chemicals [GHS] Category 1, European Union [EU] R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations for this test method are based on the false negative and false positive rates observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols and solids range from 67% (2/3)[4] to 100% (2/2) and 42% (5/12) to 50% (5/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols, ketones, and solids range from 50% (7/14) to 56% (9/16), 40% (4/10), and 10% (2/20 to 2/21), respectively, depending on the hazard classification system. When substances within these chemical and physical classes are excluded from the database, the accuracy of BCOP across the EU, EPA, and GHS classification systems ranges from 87% (72/83) to 92% (78/85) and the false negative and false positive rates range from 0% (0/27) to 12% (3/26) and 12% (7/58) to 16% (9/56), respectively.

Coefficient of variation (CV) analysis of BCOP test method intralaboratory repeatability data (*In Vitro* Irritancy Scores) from two studies ranged from 11.8% to 14.2% for 16 substances of varying irritancy and from 1.1% to 13% for five substances predicted as severe irritants. Intralaboratory reproducibility evaluations indicated mean and median CV values for permeability values were 33.4% and 29.0%, respectively, for 25 surfactant-based personal care cleaning formulations in one study. Mean CV values of *In Vitro* Irritancy Scores for 16 substances tested two or more times in three laboratories ranged from 12.6% to 14.8%, while the median CV values ranged from 6.7% to 12.4%.

In a qualitative assessment of interlaboratory reproducibility of hazard classification category, 67% to 94% of the substances were classified the same by the participating laboratories. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies by performing a CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. In these studies, the mean and median CV values were (a) 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories, (b) 25% and 22%,

[4]The numbers in parentheses represent the numbers used to calculate the percentages. For the false negative or false positive rates, the numerators represent the total number of substances incorrectly identified as negatives or positives, respectively, by the *in vitro* test method, while the denominators represent the total number of substances identified as negatives or positives, respectively, by the *in vivo* rabbit eye test method.

respectively, for results obtained in five laboratories, and (c) 32.4% and 22.8%, respectively, for results obtained in three laboratories.

When studies are conducted using the BCOP test method, the study protocol should be based on the recommended standardized test method protocol provided in **Appendix D**. Exceptions and/or changes to the standardized test method protocol should be accompanied by a scientific rationale.

Users should be aware that BCOP's performance characteristics and the standardized test method protocol could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other chemical and product classes. Therefore, prior to initiation of BCOP studies, investigators are encouraged to consult the ICCVAM/National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To further characterize and potentially improve the usefulness of the BCOP test method for identifying ocular corrosives and severe irritants, and to evaluate its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations are recommended:

1. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be conducted. Such data will allow for the development of standardized decision criteria and a more comprehensive evaluation of the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.
2. Studies should be conducted to evaluate the impact of using a corneal holder that maintains normal corneal curvature (e.g., the corneal mounting system designed by Ubels et al. 2002) on accuracy and/or reliability of the BCOP test method.
3. The effect of modifying various test method protocol components (e.g., changing the duration of exposure) on the accuracy and/or reliability of the BCOP test method should be evaluated.

ICE Test Method

There are sufficient data to support the use of the ICE test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations

for this method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols, surfactants and solids range from 33% (1/3) to 50% (1/2), 44% (4/9) to 57% (4/7), and 46% (6/13) to 70% (7/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols range from 27% (3/11) to 50% (5/10), depending on the hazard classification system evaluated. When substances within these chemical and physical classes are excluded from the database, the accuracy of ICE across the EU, EPA, and GHS classification systems ranges from 91% (72/79 to 75/82) to 92% (69/75) and the false negative and false positive rates range from 29% (2/7) to 33% (3/9) and 5% (4/73) to 6% (4/68 to 4/70), respectively.

The range of CV values for the corneal thickness measurement, when results were compared within experiments, was from 0.9% to 6.1%. The other endpoints evaluated produced ranges of CV values that were larger, with variability most prominent with the nonirritating substance. The range of CV values for the corneal thickness measurement, when results were compared across experiments, was from 1.8% to 6.3%. The CV values for the remaining endpoints had a larger range (e.g., corneal swelling CV = 13.9% to 138.7%). However, if the nonirritating substance is removed, the range of CV values is reduced (e.g., corneal swelling CV = 13.9% to 22.4%).

One interlaboratory comparative study involving four laboratories contained test data on 59 substances for an assessment of interlaboratory reproducibility. Based on a qualitative analysis, 60% to 70% of the substances classified as ocular corrosives or severe irritants, depending on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001), were correctly identified by all four participating laboratories. A CV analysis of these same data indicated that the mean and median CV for severe substances tested was less than 35% for all test method endpoints, with the exception of corneal swelling.

When studies are conducted using this test method, the study protocol should be based on the recommended standardized ICE test method protocol provided in **Appendix E**. Exceptions and/or changes to the standardized test method protocol should be accompanied by a scientific rationale.

Users should be aware that ICE's performance characteristics and the standardized test method protocol could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of ICE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test

method for evaluating substances that are within a specific chemical, physical, or product classes.

To further characterize and potentially improve the usefulness of the ICE test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations are recommended:

1. Additional optimization studies/evaluations should be conducted in an attempt to decrease the 29% to 33% false negative rate of the ICE test method. After optimization, additional studies to further assess the reliability and accuracy of the test method are recommended.
2. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be included when the ICE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed on the optical pachymeter, which is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

IRE Test Method

Based on the accuracy (64% [68/107] to 69% [79/114]), false negative (24% [12/49] to 31% [14/45]), and false positive (35% [23/65] to 40% [25/62]) rates across the EU, EPA, and GHS classification systems, the use of the IRE test method for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended. There also are insufficient data using all four recommended IRE endpoints (corneal opacity, fluorescein penetration, corneal swelling, and observations of significant effect on corneal epithelium) to assess test method accuracy and reliability when all these endpoints are evaluated in a single study.

Based on a qualitative analysis of available data, 100% of the 12 to 18 substances were correctly identified as severe irritants or ocular corrosives in the IRE by four laboratories participating in a validation study, when compared to *in vivo* rabbit eye test data classification dependent on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001). Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds; and such product classes as organic solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for two studies by performing a CV analysis of corneal opacity, swelling, and, for the second study, fluorescein penetration measurements for substances tested in multiple laboratories. The CV analysis of

the first study indicated that the median CV for 59 substances tested was between 43.4% and 49.7% for the 4-hour corneal opacity and swelling endpoints, respectively. The CV was between 33.6% and 35.5% when only severe irritants are considered. In the second study using corneal opacity, swelling, and fluorescein penetration, the median CV for all substances ranged from 24.0% to 40.0% and from 15.4% to 35.5% when only severe irritants were considered.

When non-regulatory, validation, or optimization studies are conducted using the IRE test method, the protocol should be based on the standardized protocol provided in **Appendix F**. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that IRE's performance characteristics and the standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To potentially improve the usefulness of the IRE test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), the following evaluations should be conducted:

1. The IRE test method decision criteria should be optimized. Once optimized, additional validation studies should be conducted to further evaluate the relevance and reliability of the IRE test method.
2. A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, of the corneal tissue should be included when the IRE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed when an optical pachymeter is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

HET-CAM Test Method

ICCVAM evaluated several HET-CAM analysis methods proposed for identifying substances that are ocular corrosives or severe irritants. These included one analysis method termed Irritation Score (IS)(B)-10 and another analysis method termed IS(B)-100. The range of hazard classification accuracy rates across the EU, EPA, and GHS classification systems

for these two analysis methods ranged from 65% (64/98) to 68% (69/101) for IS(B)-10 and 52% (69/133) to 57% (94/164) for IS(B)-100, when the decision criteria of Luepke (1985) were used. The overall false negative and false positive rates of the IS(B)-10 analysis method range from 30% (10/33 to 12/40) to 32% (10/31) and 33% (20/61) to 36% (24/67), respectively, depending on the classification system. The overall false negative and false positive rates for the IS(B)-100 analysis method range from 6% (2/33) to 13% (5/39) and 52% (68/131) to 59% (58/99), respectively, depending on the classification system. Based on these rates, the use of these analyses methods and decision criteria for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended.

The analysis of intralaboratory repeatability was evaluated using data from two different publications for the IS(B) analysis method. In both studies, the hemorrhage endpoint had the highest CV value (109.10%-117.56%). Similar results were obtained from the analysis of intralaboratory reproducibility.

A qualitative analysis of interlaboratory reliability for the also was conducted for the IS(B) analysis method. 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, when compared to all three hazard classification systems. For the IS(B)-100 analysis method, the participating laboratories in a study were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated, when compared to all three hazard classification systems.

A quantitative evaluation of interlaboratory reproducibility for 14 substances, evaluated at 100% concentration (IS(B)-100), indicated that the mean and median CV values were 31.86% and 33.04%, respectively. For 12 substances evaluated at 10% concentration (IS(B)-10), the mean and median CV values were 66.29% and 60.75%, respectively. For the substances evaluated in another study which used the IS(B) analysis method, the mean and median CV values for substances tested at 10% concentration were 60.17% and 42.65%, respectively. For substances tested at 100% concentration in the same study, the mean and median CV values were lower: 35.21% and 26.22%, respectively.

When non-regulatory, validation, or optimization studies are conducted using the HET-CAM test method, the protocol should be based on the standardized protocol provided in **Appendix G**. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that HET-CAM's performance characteristics and the standardized test method protocol could be revised as additional data becomes available. Therefore, prior to initiation of HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, chemical and physical class performance characteristics, and the recommended standardized test method protocol. Evaluation of the most current information will allow users to determine

the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

To potentially improve the usefulness of the HET-CAM test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), additional studies should be conducted to further optimize the HET-CAM prediction models and the decision criteria (e.g., mtc10) that would be used to identify ocular corrosives and severe irritants for the EPA, GHS, or EU classification systems.

General Recommendations and Comparison of Performance Characteristics for Four *In Vitro* Test Methods

Results from appropriately validated *in vitro* ocular toxicity test methods are recommended for use in a weight-of-evidence decision making process in accordance with the EPA and EU ocular testing regulations (EPA 1998, EU 2004) and the GHS tiered-testing strategy (UN 2003). In these testing schemes, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are indicated based on a weight-of-evidence evaluation of supplemental information (e.g., structure-activity relationships, other testing data). Use of a weight-of-evidence decision making process and a tiered-testing strategy for classification of substances as ocular corrosives or severe irritants may eliminate the pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances.

The comparative accuracy and false positive/false negative rates of these four *in vitro* ocular toxicity test methods in identifying ocular corrosives and severe irritants using the EU, EPA, and GHS classification systems are summarized in **Table 6-1**. Exclusion of specific chemical and physical classes increases the accuracy and decreases the false positive and false negative rates for BCOP and ICE. ICCVAM recommends that users consider, to the extent possible, the chemical and physical structures of the substances to be tested to determine whether either of these test methods would be appropriate to use as a screening test for ocular corrosion or severe irritation. Additional studies with each test method are recommended to determine if modification of the test method standardized protocol and/or the decision criteria for classification of a test substance as a corrosive/severe irritant or as a nonsevere irritant/nonirritant can improve test method sensitivity and specificity.

Additional research and development, optimization, and/or validation efforts should use reference substances with existing rabbit data. Additional rabbit studies should be conducted only if important data gaps are identified. If such studies are conducted, they should be designed to minimize the number of rabbits tested, to minimize or avoid pain and distress, and to maximize the information collected. Design and conduct of such studies should be in

accordance with the recommendations from the Scientific Symposium on Mechanisms of Chemically-Induced Ocular Injury and the Scientific Symposium on Minimizing Pain and Distress in Ocular Safety Testing (see http://iccvam.niehs.nih.gov/methods/ocudocs/ocumeet/sympinfo.htm). These symposia were organized by ICCVAM, NICEATM, and the European Centre for the Validation of Alternative Methods.

All raw data generated using any of the recommended standardized *in vitro* ocular testing protocols and the *in vivo* rabbit eye test on the same substance should be submitted to NICEATM to expand the available validation database for these four test methods. The availability of such data will allow for additional retrospective evaluations of test method accuracy and/or reliability. Ideally, all substances should be completely identified (e.g., chemical name, chemical class, physicochemical properties). However, if this is not possible for proprietary reasons, data may be submitted using coded labels for each substance tested. If such coding is used, as much information as possible on physical and chemical properties should be provided to NICEATM.

Although the IRE and HET-CAM test methods cannot currently be recommended for meeting regulatory testing requirements, there may be non-regulatory uses for these two test methods. Accordingly, the four *in vitro* test methods should be considered prior to conducting *in vivo* ocular testing and an alternative test method should be used where determined appropriate for the specific testing situation. Since ocular irritancy testing frequently involves more than slight or momentary pain or distress, consideration of alternative test methods prior to the use of animals is necessary to comply with provisions of U.S. Animal Welfare Act regulations (9 CFR, Part 2, Section 2.31 and 9 CFR, Part 2, Section 2.32), the Public Health Service Policy on the Humane Care and Use of Laboratory Animals (PHS 2002), and the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (National Research Council 1996).

The potential usefulness of combining two or more *in vitro* test methods in a battery to identify ocular corrosives and severe irritants should be evaluated. Currently, there is insufficient guidance on the utility of a battery approach for such determinations.

Interested stakeholders are encouraged to support research and development of alternative test methods and technologies that may provide for a more accurate assessment of ocular toxicity and/or advantages in terms of time and cost.

ICCVAM Recommended Substances for Validation of *In Vitro* Ocular Toxicity Test Methods for the Evaluation of Ocular Corrosives and Severe Irritants

ICCVAM developed a list of reference substances recommended for the development of alternative ocular toxicity test methods and for evaluating the performance of any optimized test method protocol (**Appendix H**). Use of this standardized list of reference substances will aid in evaluating the comparative performance of different alternative test methods and, thus, in the selection of the most appropriate test method(s) to be used for a particular testing purpose. In accordance with ICCVAM procedures, once an adequate validation database is

available for any of these test methods, performance standards will be developed that can be used to evaluate the performance of other test methods that are structurally and functionally similar. These performance standards will include essential test method components, a minimum list of reference chemicals (i.e., a subset of the recommended list in this report), and comparable performance that should be achieved.

1.0 INTRODUCTION

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at http://iccvam.niehs.nih.gov/about/PL106545.pdf) to evaluate the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. Following such evaluations, ICCVAM is required to provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of such methods.

In August 2003, the ICCVAM Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) recommended that ICCVAM give high priority to reviewing the validation status of existing *in vitro* test methods proposed for identifying ocular corrosives and severe irritants. In October 2003, the U.S. Environmental Protection Agency (EPA) formally nominated four *in vitro* ocular irritation test methods and related activities for evaluation by ICCVAM. This included review of the current validation status of four *in vitro* test methods proposed for identifying potential ocular corrosives and severe irritants in a tiered-testing strategy, since validation[5] 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 (ICE) assay, and the Isolated Rabbit Eye (IRE) assay. Within Europe, the European Commission has concluded that positive results from these four methods can be used to classify and label substances as severe ocular irritants and corrosives (EU 2004). However, the policy specifically states:

> "These tests are not yet validated, and therefore not included in Annex V. Positive results can be used to consider a substance a severe irritant and R41 applied with no further testing. Where a negative result is obtained, an in vivo test should subsequently be required, as the in vitro tests have not been shown to adequately discriminate between eye irritants and non-irritants."

ICCVAM 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, ECVAM liaisons were designated for the ICCVAM OTWG to ensure input and contributions during the evaluation and review process.

---

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

NICEATM, in conjunction with the OTWG, subsequently prepared four comprehensive background review documents (BRDs) reviewing the available data and information for each of the four *in vitro* test methods. Each BRD described the current validation status of the *in vitro* test method, including what is known about its reliability and accuracy, the scope of the substances tested, and the availability of a standardized protocol.

The BRDs were based on published studies using the respective test method, and other data and information submitted in response to a 2004 public call for information, which was published in a *Federal Register* (*FR*) notice (*FR* Vol. 69, No. 57, pp. 13859-61; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). On November 3, 2004, the availability of the draft BRDs was announced in an *FR* notice (Vol. 69, No. 212, pp. 64081-2; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). The BRDs were made available in electronic format on the ICCVAM/NICEATM website (Available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) and from NICEATM on request.

The ICCVAM convened an international independent Expert Panel on January 11-12, 2005, to assess the validation status of these four *in vitro* test methods for identifying ocular corrosives or severe irritants. Comments from the public and scientific community on the BRDs were provided to the Expert Panel and made available on the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/methods/ocudocs/ocucomm.htm). Public comments at the meeting revealed that additional relevant data was available that had not yet 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. On March 21, 2005, the availability of *The ICCVAM Expert Panel Evaluation of the Current Validation Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants* was announced via an *FR* notice (Vol. 70, No. 53, pp. 13513-4; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm).

In response to the Expert Panel's recommendation, an *FR* notice was published on February 28, 2005 (Vol. 70, No. 38, pp. 9661-2; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). The notice 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 re-sent 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, which were used for reanalysis of test method performance.

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, and necessitated a reanalysis of the accuracy and reliability of all four test methods.

The original draft BRDs also provided an evaluation of the accuracy of each test method by chemical class. 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 at 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.

Finally, an additional accuracy analysis was conducted. In this analysis, the accuracy of each *in vitro* ocular irritancy test method for detecting ocular corrosives or severe irritants, depending on whether the *in vivo* rabbit classification was based on the severity of the response and/or its persistence to day 21 post-treatment, was determined.

A list of proposed reference substances for validation of *in vitro* tests to detect ocular corrosives and severe irritants was included in the draft BRDs released on November 3, 2004. A revised list of proposed reference substances was prepared after consideration of the following:

- Recommendations of the Expert Panel that resulted from their deliberations on January 11-12, 2005
- Submission of additional Draize rabbit eye test results for approximately 300 substances
- Clarification regarding the United Nations (UN) Globally Harmonized System (GHS) rules for classification of severe irritants (UN 2003) that resulted in the reclassification of two proposed reference substances from nonsevere to severe irritants
- Reassignment of the candidate reference substances to chemical classes using MeSH (NLM 2005)

The accuracy and reliability reanalyses and the revised reference substances list for validation of *in vitro* tests to detect ocular corrosives and severe irritants were presented in a BRD Addendum that was released on July 26, 2005, with notification of its release through the ICCVAM electronic mailing list and via an *FR* notice (Vol. 70, No. 142, p. 43149; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). The BRD Addendum was made available in electronic format on the ICCVAM/NICEATM website and from NICEATM on request.

The Expert Panel was subsequently reconvened via 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 (http://iccvam.niehs.nih.gov/methods/ocudocs/addendcomm.htm). The Expert Panel provided formal comment on each of the four *in vitro* test methods, as well as the proposed list of reference substances. In addition, the public were provided time at the public meeting to comment (although no public comments were provided). The Expert Panel then provided final endorsement regarding the impact, if any, of the information in the BRD Addendum on their original evaluation from the January 11-12, 2005 meeting. The availability of *The ICCVAM Expert Panel Evaluation of the Draft Background Review Document for In Vitro*

Test Methods For Identifying Ocular Corrosives and Severe Irritants - Addendum was announced via an *FR* notice (Vol. 70, No. 211, p. 66451; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) on November 2, 2005.

Subsequently, the draft BRDs and the draft BRD Addendum, the Expert Panel report and its addendum, and all public comments were made available to the SACATM for their consideration at their meeting on December 12, 2005. The SACATM agreed with the conclusions of the Expert Panel.

The ICCVAM and OTWG considered the Expert Panel report and its addendum (**Appendix A**), the revised accuracy and reliability analyses (see **Appendix B** for accuracy analyses results), all public comments, and the comments of SACATM in preparing the final test method recommendations that are provided in this report. This report will be made available to the public and provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at http://iccvam.niehs.nih.gov/about/PL106545.pdf). Agencies with applicable testing regulations and guidelines (**Appendix C**) must respond to ICCVAM within 180 days of receiving the ICCVAM recommendations. These responses will be made available to the public on the ICCVAM website (http://iccvam.niehs.nih.gov) as they are received.

2.0 THE BCOP TEST METHOD

2.1 BCOP Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the BCOP BRD, which reviewed the available data and information for the test method.[6] The BRD describes the current validation status of the BCOP test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/).

2.1.1 Test Method Description

The BCOP test method is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the bovine cornea in an isolated system. In this test method, damage by the test substance is assessed by quantitative measurements of changes in corneal opacity and permeability with an opacitometer and an ultraviolet/visible (UV/VIS) spectrophotometer, respectively. Both measurements are used to calculate an *In Vitro* Irritancy Score, which is used to assign an *in vitro* irritancy classification for prediction of the *in vivo* ocular irritation potential of a test substance. Although histopathological data could not be formally evaluated by ICCVAM, a histopathological assessment can be included on a case-by-case basis to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category) or to identify ocular damage that does not produce opacity or permeability changes in the isolated cornea.[7] Histopathology also is used for chemical classes or formulations that are not well characterized in the BCOP assay, where the mode of action cannot be easily predicted, when delayed effects might be anticipated, or when a more complete characterization of damage is needed.

The BCOP test method protocols used in the various studies are similar, but not identical.[8] Variations in the publicly available BCOP protocols include different instrumentation to evaluate opacity, different decision criteria (i.e., prediction models) or *in vitro* classification systems, and differences in the use of positive controls, among other methodological variations. The essential principles of the test method protocol include isolating and culturing the bovine cornea, treating the isolated cornea with a test substance, collecting opacity and permeability data, and evaluating the data in relation to a prediction model. However, given the various uses and applications of the BCOP test method by different investigators and laboratories, and the evolution of the test method over time, a number of laboratory-specific differences have been noted regarding the conduct of the test method.

---

[6]Comparison of the performance analysis for BCOP to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

[7]For the studies discussed here, histopathological endpoints were not evaluated or incorporated into the accuracy assessment.

[8]For additional information on this evaluation, please see the BCOP BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#bcop).

2.1.2 Validation Database

A total of 158 substances in eight studies were used to evaluate BCOP test method accuracy. These substances represented a variety of chemical and product classes (ICCVAM 2006a). The chemical classes tested included alcohols, heterocyclic compounds, carboxylic acids, ketones, esters, inorganic salts, ethers, hydrocarbons, amines, and onium compounds. The product classes tested included solvents, surfactants, chemical/synthetic intermediates, drugs/pharmaceuticals/therapeutic agents, petroleum products, cleaners, personal care cleansers, hair shampoos, pesticides, plasticizers, reagents, bactericides, and insect repellents.

2.1.3 Test Method Accuracy

Based on all available data, the BCOP test method has an overall accuracy of 79% (113/143)[9] to 81% (119/147), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), European Union (EU; 2001), or GHS (UN 2003) classification systems. Furthermore, the BCOP test method has an overall false positive rate of 19% (20/103) to 21% (22/103) and an overall false negative rate of 16% (7/43) to 25% (10/40), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), EU (2001), or GHS (UN 2003) classification systems.

There were some notable trends in the performance of the BCOP test method among substances grouped according to chemical class and/or physicochemical properties (**Table 2-1**). The chemical classes of substances that were most consistently overpredicted (i.e., were false positives) by the BCOP test method, according to the GHS classification system are alcohols (53%, 8/15) and ketones (40%, 4/10). With regard to physical form, liquids (26%, 18/68) appear more likely than solids (10%, 2/20) to be overpredicted by the BCOP test method.

Alcohols (67%, 2/3) also were most often underpredicted (i.e., were false negatives) by the BCOP test method, according to the GHS classification system. With regard to physical form, solids (42%, 5/12) appear more likely than liquids (4%, 1/24) to be underpredicted by the BCOP test method. There was no definitive difference among the underpredicted substances for which pH information was available.

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

[9]The numbers in parentheses represent the data used to calculate the percentages noted.

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

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

Abbreviations: BCOP = Bovine Corneal Opacity and Permeability; 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*.

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

[4]Data used to calculate the percentage.

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

[6]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.

[7]Total number of GHS Category 1 substances for which pH information was obtained.

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

[9]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.

[10]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.

Finally, the underpredicted substances were more likely to be classified *in vivo* (according to the GHS classification system) based on persistent lesions, rather than on severe lesions. However, three substances that caused severe lesions *in vivo* (corneal opacity=4) were false negatives in BCOP.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the BCOP test method for the EPA and EU classification systems can be obtained from **Section 6.0**, **Appendix B**, and the BCOP BRD.

2.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Quantitative BCOP test method data were available for replicate corneas within individual experiments or for replicate experiments within an individual laboratory for three studies. Therefore, an evaluation of the intralaboratory repeatability and reproducibility of the BCOP test method could be conducted. Intralaboratory repeatability of *In Vitro* Irritancy Scores was assessed by analyzing two studies for substances predicted as severe eye irritants (*In Vitro* Scores $\geq$55.1). For 16 substances of varying irritancy evaluated in one study, the median coefficient of variation (CV) for *In Vitro* Irritancy Scores for replicate corneas (n=3) ranged from 11.8% to 14.2%. In a second study, the range of mean and median CV values for *In Vitro* Irritancy Scores for replicate corneas (n=4) was 1.1% to 13% for five substances predicted as severe irritants.

A CV analysis of intralaboratory data (*In Vitro* Irritancy Scores) from two studies indicated the following intralaboratory reproducibility of the BCOP test method for substances predicted as severe eye irritants. In one study, the between experiment (n=3) mean and median CV values for permeability values were 33.4% and 29.0%, respectively, for 25 surfactant-based personal care cleaning formulations. In the second study, the between experiment mean CV values of *In Vitro* Irritancy Scores for 16 substances tested two or more times in three laboratories ranged from 12.6% to 14.8%, while the median CV values ranged from 6.7% to 12.4%.

Additionally, comparable BCOP data were available for multiple laboratories within each of three comparative validation studies, which allowed for an evaluation of the interlaboratory reproducibility of the BCOP test method. For these studies, interlaboratory reproducibility was evaluated qualitatively based on the ocular irritancy classification assigned to each substance by each laboratory, and quantitatively using *In Vitro* Irritancy Scores. In the qualitative assessment of interlaboratory reproducibility of hazard classification category, 67% to 94% of the substances were classified the same by the participating laboratories. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as solvents, surfactants, chemical intermediates, and pesticides. A quantitative evaluation of interlaboratory reproducibility was conducted for these three studies by performing a CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. In one study, the 17 substances predicted as severe in the BCOP assay had mean and median CV values of 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories. In a second study, the 32 substances predicted as

severe in the BCOP assay had mean and median CV values of 25% and 22%, respectively, for results obtained in five laboratories. In a third study, the mean and median CV values for the *In Vitro* Irritancy Scores of the 16 substances were 32.4% and 22.8%, respectively, for results obtained in three laboratories.

Finally, the interlaboratory correlation between BCOP test method endpoint data generated by each laboratory was determined for 60 substances, as well as for various subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances. Interlaboratory correlation coefficients for the *In Vitro* Irritancy Score generally spanned a range of 0.867 to 0.958 depending on the specific subsets of substances being evaluated.

2.2 ICCVAM Recommendations for the BCOP Test Method

2.2.1 Use of the BCOP Test Method

ICCVAM recognizes that the BCOP test method is not proposed as a stand alone replacement for the *in vivo* rabbit eye test method currently used for regulatory classification and labeling. ICCVAM concludes that there are sufficient data to support the use of the BCOP test method, in appropriate circumstances and with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach.[10]

The identified limitations for this test method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols and solids range from 67% (2/3) to 100% (2/2) and 42% (5/12) to 50% (5/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols, ketones, and solids range from 50% (7/14) to 56% (9/16), 40% (4/10), and 10% (2/20 to 2/21), respectively, depending on the hazard classification system. When substances within these chemical and physical classes are excluded from the database, the accuracy of BCOP across the EU, EPA, and GHS classification systems ranges from 87% (72/83) to 92% (78/85) and the false negative and false positive rates range from 0% (0/27) to 12% (3/26) and 12% (7/58) to 16% (9/56), respectively.

A tiered-testing strategy for ocular irritation/corrosion (e.g., as described in the Globally Harmonized System of Classification and Labelling of Chemicals; UN 2003) allows for the use of validated and accepted *in vitro* methods prior to the use of animals for ocular safety testing. In a tiered testing strategy, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the

[10]The recommendations are based on the performance results for BCOP without the use of histopathology for decision making purposes.

opportunity for confirmatory testing if false positive results are suggested based on a weight-of-evidence evaluation of supplemental information (e.g., pH, structure-activity relationships, other testing data). Using *in vitro* data in a tiered-testing strategy with a weight-of-evidence decision process to classify substances as ocular corrosives or severe irritants will avoid the potential pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances. A tiered-testing strategy may not be applicable to purposes other than regulatory classification and labeling.

Users should be aware that BCOP's performance characteristics could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of BCOP studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

2.2.2 BCOP Test Method Protocol

ICCVAM recommends that when testing is conducted, the BCOP test method protocol should be based on the BCOP standardized test method protocol provided in **Appendix D**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the proposed standardized test method protocol should be accompanied by a scientific rationale. Users should be aware that the test method protocol could be revised based on future optimization and/or validation studies. ICCVAM, therefore, recommends that test method users consult the ICCVAM/NICEATM website to ensure use of the most current recommended test method protocol http://iccvam.niehs.nih.gov/methods/eyeirrit.htm).

2.2.3 Optimization of the Current BCOP Test Method Protocol

The current ICCVAM recommendations are focused on the use of the BCOP test method as a screening test for ocular corrosives and severe irritants (see **Section 2.2.1**). For that use, the current test method protocol should be sufficient. To further the use of this test method and to evaluate the use of the BCOP test method as a potential replacement for the *in vivo* rabbit eye test method or for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), ICCVAM recommends additional studies be considered and undertaken to decrease the false positive rate of this test method.

A histopathological evaluation of the corneal tissue, using a standardized scoring scheme, should be conducted. Such data will allow for the development of standardized decision criteria and a more comprehensive evaluation of the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results

Studies should be conducted to evaluate the impact of using a corneal holder that maintains normal corneal curvature (e.g., the corneal mounting system designed by Ubels et al. 2002) on accuracy and/or reliability of the BCOP test method.

ICCVAM also recommends that an evaluation be conducted on the effect of modifying various test method protocol components (e.g., duration of test substance exposure) on the accuracy and/or reliability of the BCOP test method.

[This Page Intentionally Left Blank]

3.0 THE ICE TEST METHOD

3.1 ICE Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the ICE BRD, which reviewed the available data and information for the test method.[11] The BRD describes the current validation status of the ICE test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/).

3.1.1 Test Method Description

The ICE test method is an organotypic model that provides short-term maintenance of the chicken eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, opacity, and fluorescein retention. While the latter two parameters involve a subjective assessment, analysis of corneal swelling provides an objective measurement. This objective measure potentially provides improved precision and reduced interlaboratory variability compared to the traditional *in vivo* rabbit eye test, which relies only on subjective measurements. Each measurement is either converted into a quantitative score used to calculate an overall Irritation Index, or assigned a qualitative categorization that is used to assign an *in vitro* irritancy classification. Either of these outcomes can then be used to predict the *in vivo* ocular irritation potential of a test substance. A histopathological assessment also can be included on a case-by-case basis to discriminate borderline cases (i.e., substances that produce results that preclude assignment to a single category).

The ICE test method protocols used in the various studies are similar, but not identical.[12] The primary difference among these protocols was the number of treated eyes per test substance. Acceptable ranges for negative control responses, historical data used to establish these ranges, and procedures to determine the optimum quantity of test substance to be applied have not been published.

3.1.2 Validation Database

A total of 154 substances in five studies were used to evaluate ICE test method accuracy. These substances represent a variety of chemical and product classes (ICCVAM 2006b). The chemical classes tested included, but were not limited to, acyl halides, alcohols, alkalis, amines/amidines, carboxylic acids, esters, heterocyclic, hydrocarbons, inorganic salts, ketones, onium compounds, and organophosphates. Commercial products or formulations tested included, but were not limited to, detergents, pesticides, silicone powder, ink, solvents, surfactants, toilet cleaners, and thermal paper coatings.

---

[11]Comparison of the performance analysis for ICE to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

[12]For additional information on this evaluation, please see the ICE BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice).

3.1.3 Test Method Accuracy

Based on all available data, the ICE test method has an overall accuracy of 83% (120/144) to 87% (134/154), an overall false positive rate of 6% (7/122) to 8% (9/114 to 9/116), and an overall false negative rate of 41% (13/32) to 50% (15/30), when compared to *in vivo* rabbit eye test method data classified according to the EPA (1996), EU (2001), or GHS (UN 2003) classification systems.

There were some notable trends in the performance of the ICE test method among substances grouped according to chemical class and/or physicochemical properties (**Table 3-1**). The chemical class of substances that was most consistently overpredicted (i.e., were false positives) by the ICE test method according to the GHS classification system is alcohols (50%, 5/10). With regard to physical form, liquids (10%, 9/90) appear more likely than solids (0%, 0/24) to be overpredicted by the ICE test method.

No single chemical class was prominently represented among 15 substances that were underpredicted. Five of the 15 underpredicted substances were unclassified coded substances and three were carboxylic acids. No other chemical class was represented more than twice. However, these studies do suggest that surfactants or formulations containing surfactants (e.g., detergents) (56%, 5/9) may be underpredicted by the ICE test method. They also suggest that pesticides (60%, 3/5) may be underpredicted.

With regard to physical form, eight of the 15 underpredicted substances were liquids while seven were solids. However, considering that the total number of solids (36) in the database is much smaller than the number of liquids (108), solids, with a false negative rate of 58% (7/12), appear more likely to be underpredicted than liquids, with a false negative rate of 44% (8/18).

ICE test method performance statistics also were evaluated when substances from the classes that gave the most discordant results were excluded (i.e., alcohols, surfactants, solids). When using the GHS classification system, exclusion of surfactants and solids individually resulted in small changes in the performance statistics. However, exclusion of alcohols from the data set caused a two-fold decrease in the false positive rate from 8% (9/114) to 4% (4/104). When both alcohols and surfactants were excluded, the false positive rate decreased from 8% (9/114) to 4% (4/92). The largest changes were observed when all three discordant classes were excluded from the data set; accuracy increased from 83% (120/144) to 92% (69/75), the false negative rate decreased from 50% (15/30) to 29% (2/7), and the false positive rate decreased from 8% (9/114) to 6% (4/68).

Among the eight underpredicted substances for which pH information was available, four were acidic (pH <7.0) and four were basic (pH >7.0). Basic substances (8) occupy a smaller proportion of the total database than acidic substances (12), and were more often underpredicted (50% vs. 33%). However, pH information was obtained for only 20 of the 30 total Category 1 substances.

Finally, the underpredicted substances were more likely to be classified *in vivo* based on persistent lesions (according to the GHS classification system) than on severe lesions.

Table 3-1 False Positive and False Negative Rates of the ICE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall	144	8	9/114	50	15/30
Chemical Class[5]					
Alcohol	12	50	5/10	50	1/2
Amine/Amidine	5	0	0/2	33	1/3
Carboxylic acid	10	0	0/3	43	3/7
Ester	9	13	1/8	0	0/1
Heterocyclic	9	0	0/3	33	2/6
Onium compound	8	0	0/2	33	2/6
Properties of Interest					
Liquids	108	10	9/90	44	8/18
Solids	36	0	0/24	58	7/12
Pesticide	11	0	0/6	60	3/5
Surfactant – Total	21	0	0/12	56	5/9
-nonionic	4	0	0/3	100	1/1
-anionic	2	0	0/1	100	1/1
-cationic	7	0	0/1	33	2/6
pH – Total[6]	20	-	-	40	8/20
- acidic (pH < 7.0)	12	-	-	33	4/12
- basic (pH > 7.0)	8	-	-	50	4/8
Category 1 Subgroup[7]					
- Total	23[9]	-	-	35	8/23
- 4 (CO=4 at any time)	12	-	-	33	4/12
- 3 (severity/persistence)	2	-	-	50	1/2
- 2 (severity)	4	-	-	0	0/4
- 2-4 combined[8]	18	-	-	28	5/18
- 1 (persistence)	5	-	-	60	3/5

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); ICE = Isolated Chicken Eye.

[1]N = number of substances.

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

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

[4]Data used to calculate the percentage.

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

[6]Total number of GHS Category 1 substances for which pH information was obtained.

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

[8]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.

[9]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.

However, four substances that caused severe lesions *in vivo* (corneal opacity=4) were false negatives in ICE.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the ICE test method for the EPA and EU classification systems can be obtained from **Section 6.0**, **Appendix B**, and the ICE BRD.

3.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Data were received that allowed for a quantitative analysis of intralaboratory repeatability and reproducibility of ICE test method endpoints. The range of CV values for the corneal thickness measurement, when results were compared within experiments, was from 0.9% to 6.1%. The other endpoints evaluated produced ranges of CV values that were larger, with variability most prominent with the nonirritating substance. However, this could be an exaggeration of variability given the relatively small values that were produced from the nonirritating substance relative to the irritating and corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher CV than values of 11, 14, and 18). A similar discussion also can be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively). The range of CV values for the corneal thickness measurement, when results were compared across experiments, was from 1.8% to 6.3%. The CV values for the remaining endpoints had a larger range (e.g., corneal swelling CV = 13.9% to 138.7%). However, if the nonirritating substance is removed, the range of CV values is reduced (e.g., corneal swelling CV = 13.9% to 22.4%).

One interlaboratory comparative study involving four laboratories contained sufficient ICE test data on 59 substances for an assessment of interlaboratory reproducibility. Based on a qualitative analysis, 60% to 70% of the substances classified as ocular corrosives or severe irritants, depending on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001), were correctly identified by all four participating laboratories. A CV analysis of these same data indicated that the mean and median CV for severe substances tested was less than 35% for all test method endpoints, with the exception of corneal swelling.

3.2 ICCVAM Recommendations for the ICE Test Method

3.2.1 Use of the ICE Test Method

ICCVAM recognizes that the ICE test method is not proposed as a stand alone replacement for the *in vivo* rabbit eye test method currently used for regulatory classification and labeling. ICCVAM concludes that there are sufficient data to support the use of the ICE test method, in appropriate circumstances with certain limitations, as a screening test to identify substances as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach.

The identified limitations for this method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available

database, the false negative rates for alcohols, surfactants and solids range from 33% (1/3) to 50% (1/2), 44% (4/9) to 57% (4/7), and 46% (6/13) to 70% (7/10), respectively, depending on the hazard classification system. Additionally, the false positive rates for alcohols range from 27% (3/11) to 50% (5/10), depending on the hazard classification system evaluated. When substances within these chemical and physical classes are excluded from the database, the accuracy of ICE across the EU, EPA, and GHS classification systems ranges from 91% (72/79 to 75/82) to 92% (69/75) and the false negative and false positive rates range from 29% (2/7) to 33% (3/9) and 5% (4/73) to 6% (4/68 to 4/70), respectively.

A tiered-testing strategy for ocular irritation/corrosion (e.g., as described in the Globally Harmonized System of Classification and Labelling of Chemicals; UN 2003) allows for the use of validated and accepted *in vitro* methods prior to the use of animals for ocular safety testing. In a tiered testing strategy, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are suggested based on a weight-of-evidence evaluation of supplemental information (e.g., pH, structure-activity relationships, other testing data). Using *in vitro* data in a tiered-testing strategy with a weight-of-evidence decision process to classify substances as ocular corrosives or severe irritants will avoid the potential pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances. A tiered-testing strategy may not be applicable to purposes other than regulatory classification and labeling.

Users should be aware that ICE's performance characteristics could be revised as additional data become available. For example, the current validation database did not allow for adequate evaluation of all chemical or product classes (e.g., formulations). Additional data may allow for further evaluation of this, as well as other, chemical and product classes. Therefore, prior to initiation of ICE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (seehttp://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

3.2.2 ICE Test Method Protocol

ICCVAM recommends that when testing is conducted, the ICE test method protocol should be based on the ICE standardized test method protocol provided in **Appendix E**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the proposed standardized test method protocol should be accompanied by a scientific rationale. Users should be aware that the test method protocol could be revised based on future optimization and/or validation studies. ICCVAM, therefore, recommends that test method users consult the ICCVAM/NICEATM website to ensure use of the most current recommended test method protocolhttp://iccvam.niehs.nih.gov/methods/eyeirrit.htm).

3.2.3 Optimization of the Current ICE Test Method Protocol

The current ICCVAM recommendations are focused on the use of the ICE test method as a screening test for ocular corrosives and severe irritants (see **Section 3.2.1**). For that use, the current test method protocol should be sufficient. To further the use of this test method and to evaluate the use of the ICE test method as a potential replacement for the *in vivo* rabbit eye test method or for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36), ICCVAM recommends additional studies be considered and undertaken.

Additional optimization studies/evaluations should be conducted in an attempt to decrease the 29% to 33% false negative rate of the ICE test method. After optimization, additional studies to further assess the reliability and accuracy of the test method are recommended.

ICCVAM recommends that a histopathological evaluation of the corneal tissue, using a standardized scoring scheme, be included when the ICE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed on the optical pachymeter, which is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

4.0 THE IRE TEST METHOD

4.1 IRE Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the IRE BRD, which reviewed the available data and information for the test method.[13] The BRD describes the current validation status of the IRE test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/).

4.1.1 Test Method Description

The IRE test is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the entire rabbit eye in an isolated system. In this test method, damage by the test substance is assessed by determination of corneal swelling, corneal opacity, fluorescein retention, and effects on the corneal epithelium. Identification of severe ocular irritants and corrosives is based on reaching or exceeding predetermined cut-off values in any one of the four endpoints (e.g., product of the corneal opacity and area scores ≥3; product of area and intensity scores for fluorescein penetration ≥4; corneal swelling ≥25%; or any significant effect on corneal epithelium (pitting, mottling, stippling, ulceration) (See **Appendix F** for details).

The IRE test method protocols used in the various studies are similar, but not identical.[14] Examples of some of the test method components that differed among the IRE protocols used to generate data include:

- temperature of solution used to rinse solids from the eyes ranged from room temperature to 32°C,
- amount of substance applied as a solid ranged from 25 mg to 100 mg, and
- decision criteria used for classification of substances was based on scores from two to four endpoints.

4.1.2 Validation Database

A total of 149 substances were evaluated in three studies, of which 25 were commercial products or formulations (ICCVAM 2006c). The chemical classes tested included, but were not limited to, alcohols, amides, amines, carboxylic acids, esters, ethers, formulations, heterocyclic, ketones, onium compounds, and sulfur compounds. The commercial products or formulations tested were skin cleansers, soaps, shampoos, conditioners, surfactants, and solvents.

---

[13]Comparison of the performance analysis for IRE to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

[14]For additional information on this evaluation, please see the IRE BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ire).

4.1.3 Test Method Accuracy

The overall accuracy (based on the pooled data set[15]) for the IRE test method ranged from 64% (68/107) to 69% (79/114) when compared to the *in vivo* test method data classified according to the GHS (UN 2003), EPA (1996), and EU (2001) regulatory classification systems. The overall false positive rates, when compared to these regulatory classification systems, ranged from 35% (23/65) to 40% (25/62). The overall false negative rates, when compared to the three regulatory classification systems, ranged from 24% (12/49) to 31% (14/45).

There were some trends in the performance of the IRE test method among substances grouped according to chemical class and/or physicochemical properties (**Table 4-1**). The chemical classes that were consistently overpredicted (i.e., false positives), when compared to classifications based on the GHS classification system, were alcohols (55%, 6/11), amines (50%, 3/6), and ketones (67%, 4/6). The chemical classes that were underpredicted (i.e., false negatives), when compared to classifications based on the GHS classification system, were carboxylic acids (67%, 4/6) and organic compounds (50%, 3/6).

With regard to physical form, liquids have a higher false positive rate (49%, 18/37) when compared to solids (22%, 5/23) for the IRE test method. The false negative rates for liquids and solids were relatively similar (29%, 8/28 vs. 32%, 6/19; respectively).

A subset of the substances evaluated had pH information available. For these substances, the overall false positive rate was 24% (4/17) and the overall false negative rate was 0% (0/10).

Of the surfactant-based formulations evaluated by this test method, the false positive rate was 25% (2/8) and the false negative rate was 38% (6/16). Comparatively, for substances identified as surfactants in the database, the false positive rate was 40% (2/5) and the false negative rate was 12% (1/8).

Finally, the underpredicted substances were more likely to be classified *in vivo* (according to the GHS classification) system based on persistent lesions, rather than severe lesions. However, three substances that caused severe lesion *in vivo* (corneal opacity=4) were false negatives.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the IRE test method for the EPA and EU classification systems can be obtained from **Section 6.0**, **Appendix B**, and the IRE BRD.

4.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

Due to the lack of available quantitative IRE test method data for replicate eyes within individual experiments or for replicate experiments within an individual laboratory, an

[15]The pooled dataset represents the results from all the available studies combined, regardless of the number of endpoints evaluated by each of the individual studies. Additional information about this dataset can be obtained from the IRE BRD.

Table 4-1 False Positive and False Negative Rates of the IRE Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System (Analysis Based on the Pooled Data Set)

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall	107	38	23/60	30	14/47
Chemical Class[5]					
Alcohol	13	55	6/11	50	1/2
Amide	5	0	0/3	0	0/2
Amine	11	50	3/6	20	1/5
Carboxylic acid	12	33	2/6	67	4/6
Ester	10	30	3/10	-	0/0
Ether	9	33	2/6	0	0/3
Formulation	24	25	2/8	38	6/16
Heterocycle	18	44	4/9	11	1/9
Ketone	6	67	4/6	-	0/0
Onium compound	10	33	1/3	0	0/7
Organic	12	17	1/6	50	3/6
Sulfur compound	8	20	1/5	33	1/3
Properties of Interest					
Liquid/Solution	65	49	18/37	29	8/28
Solids	42	22	5/23	32	6/19
Surfactant-based formulation	24	25	2/8	38	6/16
Surfactants	13	40	2/5	12	1/8
-nonionic	4	33	1/3	0	0/1
-anionic	2	0	0/1	100	1/1
-cationic	7	100	1/1	0	0/6
pH – Total[6]	27	24	4/17	0	0/10
-acidic	18	20	2/10	0	0/8
-basic	7	33	2/6	0	0/1
-equals 7	2	0	0/1	0	0/1
Category 1 Subgroup[7] **- Total**	37[9]	-	-	32	12/37
- 4 (CO=4 at any time)	11	-	-	27	3/11
- 3 (severity/persistence)	4	-	-	25	1/4
- 2 (severity)	3	-	-	33	1/3
- 2-4 combined[8]	18	-	-	28	5/18
- 1 (persistence)	19	-	-	37	7/19

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); IRE = Isolated Rabbit Eye.
[1]N = number of substances.
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[3]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[4]Data used to calculate the percentage.
[5]Chemical classes included in this table are represented by at least five substances tested in the IRE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[6]Total number of GHS Category 1 substances for which pH information was obtained.
[7]NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO = 4 at any time.
[8]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.
[9]The number of substances evaluated in the Category 1 subgroup analysis may be less than the number of *in vivo* Category 1 substances evaluated, since some substances could not be classified into the subgroups used in the evaluation.

evaluation of the intralaboratory repeatability and reproducibility of the IRE test method could not be conducted. However, two studies contained sufficient IRE test data (n=59 and 21 substances, respectively) for an assessment of interlaboratory reproducibility based on data reported for three or four different laboratories. For these studies, interlaboratory reproducibility was evaluated qualitatively based on the ocular irritancy classification assigned to each substance by each laboratory and quantitatively using corneal opacity, swelling in one study, and corneal opacity, corneal swelling and evaluation of fluorescein penetration in the second study.

Based on a qualitative analysis, 100% of the 12 to 18 substances were correctly identified as severe irritants or ocular corrosives in the IRE by all four participating laboratories, when compared to *in vivo* rabbit eye test data classification dependent on the regulatory classification system employed (i.e., EPA 1996, GHS [UN 2003], EU 2001). Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds; and such product classes as organic solvents, surfactants, chemical intermediates, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for these two studies by performing a CV analysis of corneal opacity, swelling, and, for the second study, fluorescein penetration measurements for substances tested in multiple laboratories. The CV analysis of the first study indicated that the median CV for all 59 substances tested was between 43.4% and 49.7% for the 4-hour corneal opacity and the 4-hour swelling endpoints, respectively. The CV was between 33.6% and 35.5% when only severe irritants are considered. In the second study using corneal opacity, swelling, and fluorescein penetration, the median CV for all substances ranged from 24.0% to 40.0% (the largest variability was for corneal swelling) and from 15.4% to 35.5% when only severe irritants were considered.

4.2 ICCVAM Recommendations for the IRE Test Method

4.2.1 Use of the IRE Test Method

Based on the accuracy (64% [68/107] to 69% [79/114]), false negative (24% [12/49] to 31% [14/45]), and false positive (35% [23/65] to 40% [25/62]) rates across the EU, EPA, and GHS classification systems, the use of the IRE test method for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended. There also are insufficient data using all four recommended IRE endpoints (corneal opacity, fluorescein penetration, corneal swelling, and observations of significant effect on corneal epithelium) to assess test method accuracy and reliability when all these endpoints are evaluated in a single study.

Users should be aware that IRE's performance characteristics could be revised as additional data become available. Therefore, prior to initiation of non-regulatory, validation, or optimization IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (seehttp://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to

determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

4.2.2 IRE Test Method Protocol

When non-regulatory, validation, or optimization studies are conducted using the IRE test method, the protocol should be based on the standardized protocol provided in **Appendix F**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that IRE's standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of IRE studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current recommended standardized test method protocol.

ICCVAM recommends that, for all studies, raw data be collected and maintained. The availability of such data will allow for further retrospective evaluation of test method accuracy and/or reliability.

4.2.3 Optimization of the Current IRE Test Method Protocol

ICCVAM recommends that additional evaluation studies be conducted to increase the current IRE database and optimize the IRE test method decision criteria. Once these studies are conducted, ICCVAM recommends that additional validation studies be conducted to further evaluate the relevance and reliability of the IRE test method.

ICCVAM recommends that a histopathological evaluation of the corneal tissue, using a standardized scoring scheme, be included when the IRE test method is conducted. Such data will allow for development of decision criteria and future assessments on the usefulness of this endpoint for classifying and labeling substances, especially those that may otherwise produce borderline or false negative results.

ICCVAM also recommends that centering lights be installed when an optical pachymeter is used to measure corneal thickness, to ensure consistent central corneal thickness measurements across laboratories.

[This Page Intentionally Left Blank]

5.0 THE HET-CAM TEST METHOD

5.1 HET-CAM Technical Summary

The following technical summary provides a synopsis of the performance analysis described in the HET-CAM BRD, which reviewed the available data and information for the test method.[16] The BRD describes the current validation status of the HET-CAM test method, including what is known about its reliability and accuracy, the scope of the substances tested, and a standardized protocol. The BRD may be obtained from the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/).

5.1.1 Test Method Description

The HET-CAM test method uses the chorioallantoic membrane (CAM), which is a vascular fetal membrane, composed of the fused chorion and allantois. 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. 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.

The HET-CAM test method protocols used in the various studies evaluated 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 mL or 0.3 mL for liquids and 0.3 g for solids,
- number of replicate eggs per test substance ranged from three to six, and
- some studies included concurrent positive control substances, while others did not.

5.1.2 Validation Database

There were several HET-CAM analysis methods used by the various studies.[17] For the Irritation Score (IS)(A)[18] and IS(B)[19] analysis methods, data were available to conduct additional sub-analyses (ICCVAM 2006d). For these sub-analyses, substances tested at a 10% concentration or 100% concentration *in vitro* were compared to responses observed at a 100% concentration tested *in vivo* (e.g., IS(A)-10, IS(B)-10, IS(B)-100).

---

[16]Comparison of the performance analysis for HET-CAM to the other three *in vitro* test methods evaluated can be reviewed in **Section 6.0** and **Appendix B**.

[17]For additional information on this evaluation, please see the HET-CAM BRD (http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#hetcam).

[18]Analysis method described in Luepke (1985).

[19]Analysis method described in Kalweit et al. (1987).

A total of 24 and 20 substances were evaluated for the IS(A)-10 and IS(A)-100 analysis methods, respectively, using the decision criteria of Luepke (1985). For the IS(B)-10 and IS(B)-100 analysis methods, using the decision criteria of Luepke (1985), 101 and 138 substances were evaluated, respectively. The chemical classes tested included, but were not limited to, alcohols, amines, esters, ethers, formulations, heterocyclic compounds, inorganic salts, ketones, and organic salts. The product classes tested included, but were not limited to, cosmetics, solvents, shampoos, flavor ingredients, and pharmaceutical synthetics.

5.1.3 Test Method Accuracy

For the IS(A) analysis method, accuracy increased when substances were evaluated at *in vitro* were tested at 100% concentration compared to the 10% concentration and where *in vivo* data were classified according to the EPA (1996), EU (2001), and GHS (UN 2003) classification systems. The opposite pattern was observed for the IS(B) analysis method; test method accuracy increased when substances were evaluated *in vitro* at 10% concentration (IS(B)-10) compared to the 100% concentration (IS(B)-100) and where *in vivo* data were classified according to the EPA (1996), EU (2001), and GHS (UN 2003) classification systems.

Chemical classes that were overpredicted by the HET-CAM IS(B) analysis methods, when testing substances at either a 10% or at 100% concentration, include alcohols (IS(B)-10: 89% [8/9]; IS(B)-100: 88% [14/16]), ethers (IS(B)-10: 50% [5/10]; IS(B)-100: 50% [6/12]), amines (IS(B)-10: 60% [3/5]; IS(B)-100: 83% [5/6]), organic salts (IS(B)-10: 57% [4/7]; IS(B)-100: 86% [6/7]), and heterocyclic compounds (IS(B)-10: 86% [6/7]; IS(B)-100: 78% [7/9]). Formulations appeared to have the lowest false positive rates for both IS(B)-10 and IS(B)-100 (**Table 5-1**). Chemical classes that were underpredicted by both analysis methods were amines and ethers.

An evaluation based on the physical form of the test substance *in vivo* depended on the analysis method being evaluated. For the IS(B)-100 analysis method, substances tested as solids *in vivo* had a false positive rate of 67% (16/24) and substances tested as liquids *in vivo* had a false positive rate of 65% (33/51) (**Table 5-1**). For the IS(B)-100 analysis method, substances tested as liquids *in vivo* had a false negative rate of 0% (0/9) and substances tested as solids *in vivo* had a false negative rate of 24% (4/17). For the IS(B)-10 analysis method, liquids had a false positive rate of 19% (3/16) and false negative rate of 37% (7/19) while solids had false positive and false negative rates of 58% (11/19) and 13% (1/8), respectively.

An analysis of the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular lesions (i.e., severity and/or persistence) responsible for classification of a substance as an ocular corrosive/severe irritant, indicated that, for IS(B)-10, the underpredicted substances were more likely to be substances classified as corrosive or severely irritating *in vivo* based on persistent lesions, with a false negative rate of 37% (10/27) compared to 15% (2/13) for substances classified as corrosive or severely irritating *in vivo* based on severity. For the IS(B)-100 analysis method, the underpredicted substances were more likely to be substances classified as corrosive or severely irritating *in vivo* based on severe lesions, with a false negative rate of 11% (2/19)

Table 5-1 False Positive and False Negative Rates of the HET-CAM Test Method, by Chemical Class and Properties of Interest, for the GHS Classification System

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall IS(B)-10 (Entire database)	101	33	20/61	30	12/40
Overall IS(B)-100 (Entire database)	138	59	58/99	13	5/39
Chemical Class-IS(B)-10[5]					
Alcohol	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-IS(B)-100[5]					
Alcohol	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 Formulation – IS(B)-10	24	0	0/8	44	7/16
pH – IS(B)-10[6]	35	58	11/19	13	2/16
- acidic (pH < 7.0)					

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
- basic (pH > 7.0)	24	50	7/14	20	2/10
	11	80	4/5	0	0/6
pH – IS(B)-100[6]	35	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[7]					
- 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[8]	13	-	-	15	2/13
- 1 (persistence)	27	-	-	37	10/27
Category 1 Subgroup-IS(B)-100[7]					
- Total	38[9]	-	-	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[8]	22	-	-	14	3/22
- 1 (persistence)	16	-	-	6	1/16

Abbreviations: CO = corneal opacity; GHS = Globally Harmonized System (UN 2003); HET-CAM = Hen's Egg Test – Chorioallantoic Membrane.

[1]N=number of substances.

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

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

[4]Data used to calculate percentages.

[5]Chemical classes included in this table are represented by at least five substances tested in the HET-CAM test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

[6]Total number of GHS Category 1 substances for which pH information was obtained.

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

[8]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.

[9]The number of substances evaluated in the Category 1 subgroup analysis may be less than the number of *in vivo* Category 1 substances evaluated, since some substances could not be classified into the subgroups used in the evaluation.

compared to 6% (1/16) for substances classified as corrosive or severely irritating *in vivo* based on persistence. However, two substances that were classified based on severe lesions (i.e., CO=4) were underpredicted by the HET-CAM IS(B)-10 and IS(B)-100 analysis methods.

The performance statistics for the EPA and EU classification systems are similar to those discussed for the GHS classification system. Additional information on the performance characteristics of the HET-CAM test method for the EPA and EU classification systems can be obtained from **Section 6.0**, **Appendix B**, and the HET-CAM BRD.

5.1.4 Test Method Reliability (Inter- and Intra-Laboratory Reproducibility)

The analysis of intralaboratory repeatability was evaluated using data from two different publications for the IS(B) analysis method. In both studies, the hemorrhage endpoint had the highest CV value (109.10%-117.56%). The CV values for the coagulation endpoint ranged from 41.78% to 95.69%. 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 dynamic ranges for each of the endpoints (0.02 to 5 for hemorrhage, 0.02 to 7 for lysis, and 0.03 to 9 for coagulation). Similar results were obtained from the analysis of intralaboratory reproducibility.

A qualitative analysis of interlaboratory reliability also was conducted. 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, when compared to all three hazard classification systems. For the IS(B)-100 analysis method, the participating laboratories in a study were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated, when compared to all three hazard classification systems. There was 100% agreement in regard to the ocular irritancy classification for 11 (64% to 69%) of the 16 to 17 substances evaluated in five laboratories using the IS(A) analysis method, when compared to all three hazard classification systems.

The overall reliability statistics, arranged by HET-CAM data analysis method, were consistent with what was observed for the individual studies evaluated. For the IS(B)-10, the statistics were 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 that described above.

A quantitative evaluation of interlaboratory reproducibility was conducted for the same analysis methods. For one study, two different evaluations were conducted based on the concentration tested *in vitro* using the IS(B) analysis method. For 14 substances evaluated at 100% concentration, the mean and median CV values were 31.86% and 33.04%, respectively. In the same study, for 12 substances evaluated at 10% concentration, the mean and median CV values were 66.29% and 60.75%, respectively. For the substances evaluated in another study which used the IS(B) analysis method, the mean and median CV values for substances tested at 10% concentration were 60.17% and 42.65%, respectively. For substances tested at 100% concentration in the same study, the mean and median CV values were lower: 35.21% and 26.22%, respectively. When substances that were tested in three different testing laboratories (instead of two) were removed from the assessment, little change was seen in the mean and median CV values for both concentrations tested. For a study using the IS(A) analysis method, the mean and median CV for substances classified as GHS Category 1 (UN 2003) were 26.09% and 27.08%, respectively. The mean and median CV for substances classified as EPA Category I (EPA 1996) were 25.86% and 26.43%, respectively.

5.2 ICCVAM Recommendations for the HET-CAM Test Method

5.2.1 Use of the HET-CAM Test Method

ICCVAM evaluated several HET-CAM analysis methods proposed for identifying substances that are ocular corrosives or severe irritants. These included one analysis method termed the IS(B)-10 and another analysis method termed IS(B)-100. The range of hazard classification accuracy rates across the EU, EPA, and GHS classification systems for these two analysis methods ranged from 65% (64/98) to 68% (69/101) for IS(B)-10 and 52% (69/133) to 57% (94/164) for IS(B)-100, when the decision criteria of Luepke (1985) were used. The overall false negative and false positive rates of the IS(B)-10 analysis method range from 30% (10/33 to 12/40) to 32% (10/31) and 33% (20/61) to 36% (24/67), respectively, depending on the classification system. The overall false negative and false positive rates for the IS(B)-100 analysis method range from 6% (2/33) to 13% (5/39) and 52% (68/131) to 59% (58/99), respectively, depending on the classification system. Based on these rates, the use of these analyses methods and decision criteria for screening and identifying ocular corrosives and severe irritants (i.e., EPA Category I, GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach, is not recommended.

Users should be aware that HET-CAM's performance characteristics could be revised as additional data become available. Therefore, prior to initiation of non-regulatory, validation, or optimization HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current validation database, overall performance characteristics, and chemical and physical class performance characteristics. Evaluation of the most current information will allow users to determine the appropriateness of this test method for evaluating substances that are within a specific chemical, physical, or product classes.

5.2.2 HET-CAM Test Method Protocol

When non-regulatory, validation, or optimization studies are conducted using the HET-CAM test method, the protocol should be based on the standardized protocol provided in **Appendix G**. This will facilitate collection of consistent data and expand the current validation database. Exceptions and/or changes to the test method protocol should be accompanied by a scientific rationale.

Users should be aware that HET-CAM's standardized test method protocol could be revised as additional data become available. Therefore, prior to initiation of HET-CAM studies, investigators are encouraged to consult the ICCVAM/NICEATM website (see http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) to review the most current recommended standardized test method protocol.

ICCVAM recommends that, for all studies, raw data be collected and maintained. The availability of such data will allow for further retrospective evaluation of test method accuracy and/or reliability.

5.2.3 Optimization of the Current HET-CAM Test Method Protocol

ICCVAM recommends that additional studies should be conducted to further optimize the HET-CAM prediction models and the decision criteria (e.g., mtc10) that would be used to identify ocular corrosives and severe irritants for the EPA, GHS, or EU classification systems. Such studies could potentially improve the usefulness of the HET-CAM test method for identifying severe ocular irritants and corrosives and its possible future use for the identification of mild and moderate ocular irritants (e.g., EPA Category II, III, and IV; GHS Category 2; EU R36).

[This Page Intentionally Left Blank]

6.0 GENERAL RECOMMENDATIONS AND COMPARISON OF PERFORMANCE CHARACTERISTICS FOR FOUR EVALUATED *IN VITRO* TEST METHODS

In addition to the test method specific recommendations discussed in **Sections 2.0** through **5.0**, ICCVAM also makes some general recommendations that relate to all the *in vitro* test methods discussed.

Table 6-1 provides a comparison of the accuracy, false positive, and false negative rates for all four *in vitro* ocular toxicity test methods evaluated for each of the regulatory hazard classification systems evaluated (EPA, EU, and GHS). As noted in the sections discussing each of the test methods individually (**Sections 2.0** through **5.0**), these performance characteristics are similar among the three hazard classification systems.

Although both BCOP and ICE can be used as screens for the detection of ocular corrosives and severe irritants in a tiered testing strategy, as part of a weight-of-evidence approach, both test methods as well as HET-CAM and IRE have limitations. As shown in **Table 6-1**, exclusion of specific chemical and physical classes increases the accuracy and decreases the false positive and false negative rates for BCOP and ICE. ICCVAM recommends that users consider, to the extent possible, the chemical and physical structures of the substances to be tested to determine whether either of these test methods would be appropriate to use as a screening test for ocular corrosion or severe irritation. Also, additional studies with each test method are recommended to determine if modification of the test method standardized protocol and/or the decision criteria for classification of a test substance as a corrosive/severe irritant or as a nonsevere irritant/nonirritant can improve test method sensitivity and specificity.

Results from appropriately validated *in vitro* ocular toxicity test methods are recommended for use in a weight-of-evidence decision making process in accordance with the EPA and EU ocular testing regulations (EPA 1996, EU 2004) and the GHS tiered-testing strategy (UN 2003)[20]. In these testing schemes, when a positive result is obtained in an appropriately validated *in vitro* test, a test substance may be classified as an ocular hazard without testing in rabbits. A substance that tests negative in the *in vitro* ocular toxicity test would need to be tested in the *in vivo* ocular test to identify possible *in vitro* false negatives and to identify moderate and mild ocular irritants. As is appropriate for any test system, there is the opportunity for confirmatory testing if false positive results are indicated based on a weight-of-evidence evaluation of supplemental information (e.g., structure-activity relationships, other testing data). Use of a weight-of-evidence decision making process and a tiered-testing strategy for classification of substances as ocular corrosives or severe irritants will eliminate the pain and distress that might be experienced by rabbits who otherwise would have been administered these test substances.

[20]A tiered-testing strategy approach may not be applicable to purposes other than regulatory classification and labeling.

Table 6-1 Comparison of Performance Characteristics of Four *In Vitro* Ocular Test Methods for the Identification of Severe Ocular Irritants or Corrosives, for Three Hazard Classification Systems

Test Method	Database	EPA Classification System				EU Classification System				GHS Classification System			
		N[1]	Accuracy (%)[2]	False Positive Rate[3] (%)	False Negative Rate[4] (%)	N[1]	Accuracy (%)[2]	False Positive Rate[3] (%)	False Negative Rate[4] (%)	N[1]	Accuracy (%)[2]	False Positive Rate[3] (%)	False Negative Rate[4] (%)
BCOP	All	143	79 (113/143)	19 (20/103)	25 (10/40)	143	80 (114/143)	21 (22/103)	18 (7/40)	147	81 (119/147)	20 (21/104)	16 (7/43)
	Excluding alcohols, ketones, and solids	83	87 (72/83)	14 (8/57)	12 (3/26)	82	88 (72/82)	16 (9/56)	4 (1/26)	85	92 (78/85)	12 (7/58)	0 (0/27)
ICE	All	145	84 (122/145)	8 (9/116)	48 (14/29)	154	87 (134/154)	6 (7/122)	41 (13/32)	144	83 (120/144)	8 (9/114)	50 (15/30)
	Excluding alcohols, surfactants, and solids	79	91 (72/79)	6 (4/70)	33 (3/9)	82	91 (75/82)	5 (4/73)	33 (3/9)	75	92 (69/75)	6 (4/68)	29 (2/7)
IRE	Pooled Data Set	107	64 (68/107)	40 (25/62)	31 (14/45)	114	69 (79/114)	35 (23/65)	24 (12/49)	107	65 (70/107)	38 (23/60)	30 (14/47)
HET-CAM	IS(B)-10	98	65 (64/98)	36 (24/67)	32 (10/31)	95	67 (64/95)	34 (21/62)	30 (10/33)	101	68 (69/101)	33 (20/61)	30 (12/40)
	IS(B)-100	133	52 (69/133)	58 (61/105)	11 (3/28)	164	57 (94/164)	52 (68/131)	6 (2/33)	138	54 (75/138)	59 (58/99)	13 (5/39)

Abbreviations: BCOP = Bovine Corneal Opacity and Permeability; EPA = U.S. Environmental Protection Agency; EU = European Union; GHS = Globally Harmonized System (UN 2003); HET-CAM = Hen's Egg Test – Chorioallantoic Membrane; ICE = Isolated Chicken Eye; IRE = Isolated Rabbit Eye.

[1]N=number of substances.

[2]Numbers in parentheses represent data used to calculate percentages.

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

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

Additional research and development, optimization, and/or validation efforts should use reference substances with existing rabbit data. Additional rabbit studies should be conducted only if important data gaps are identified. If such studies are conducted, they should be designed to minimize the number of rabbits tested, to minimize or avoid pain and distress, and to maximize the information collected. Designing and conducting such studies should be in accordance with the recommendations from the Scientific Symposium on Mechanisms of Chemically-Induced Ocular Injury and the Scientific Symposium on Minimizing Pain and Distress in Ocular Safety Testing (see http://iccvam.niehs.nih.gov/methods/ocudocs/ocumeet/sympinfo.htm). These symposia were organized by ICCVAM, NICEATM, and ECVAM.

All raw data generated using any of the recommended standardized *in vitro* ocular testing protocols and the *in vivo* rabbit eye test on the same substance should be submitted to NICEATM to expand the available validation database for these four test methods. The availability of such data will allow for additional retrospective evaluations of test method accuracy and/or reliability. Ideally, all substances should be completely identified (e.g., chemical name, chemical class, physicochemical properties). However, if this is not possible for proprietary reasons, data may be submitted using coded labels for each substance tested. If such coding is used, as much information as possible on physical and chemical properties should be provided to NICEATM.

Although the IRE and HET-CAM test methods cannot currently be recommended for meeting regulatory testing requirements, there may be non-regulatory uses for these two test methods. Accordingly, the four *in vitro* test methods should be considered prior to conducting *in vivo* ocular testing and an alternative test method should be used where determined appropriate for the specific testing situation. Since ocular irritancy testing frequently involves more than slight or momentary pain or distress, consideration of alternative test methods prior to the use of animals is necessary to comply with provisions of U.S. Animal Welfare Act regulations (9 CFR, Part 2, Section 2.31 and 9 CFR, Part 2, Section 2.32), the Public Health Service Policy on the Humane Care and Use of Laboratory Animals (PHS 2002), and the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (National Research Council 1996).

The potential usefulness of combining two or more *in vitro* test methods in a battery to identify ocular corrosives and severe irritants should be evaluated. Currently, there is insufficient guidance on the utility of a battery approach for such determinations.

Interested stakeholders are encouraged to support research and development of alternative test methods and technologies that may provide for a more accurate assessment of ocular toxicity and/or advantages in terms of time and cost.

[This Page Intentionally Left Blank]

7.0 ICCVAM RECOMMENDATIONS ON SUBSTANCES FOR VALIDATION OF *IN VITRO* OCULAR TOXICITY TEST METHODS FOR THE EVALUATION OF OCULAR CORROSIVES AND SEVERE IRRITANTS

In addition to evaluating the validation status of four *in vitro* ocular toxicity test methods for their ability to identify ocular corrosives and severe irritants, ICCVAM developed a list of reference substances for the optimization and/or validation of *in vitro* tests to identify ocular corrosives and severe irritants. This section provides ICCVAM's recommendations on these reference substances.

ICCVAM reviewed the Expert Panel's report and addendum (provided in **Appendix A**), the results of the analysis in the BRDs, and the public comments received to both. Based on these sources, ICCVAM makes the following recommendations with relation to the list of reference substances for the optimization and/or validation of *in vitro* ocular toxicity test methods for identification of ocular corrosives and severe irritants.[21]

ICCVAM endorses the reference substances list of 122 substances. The list of substances (see **Appendix H**) includes:

- 79 GHS Category 1 substances (UN 2003); 10 of which the Category 1 classification is based solely on human data
- 28 GHS Category 2 substances (UN 2003)
  - 15 GHS Category 2A substances (moderate irritants)
  - 13 GHS Category 2B substances (mild irritants)
- 15 GHS nonirritant substances (UN 2003)
- 34 chemical classes
- 24 product classes
- 79 liquids
- 43 solids

ICCVAM further endorses the use of the reference substance list as a source for generating a subset of substances to be used for evaluating *in vitro* ocular toxicity test methods on a scientifically sound case-by-case basis. It is recommended that the subset of substances that are developed from the reference substance list comprise a scientifically sound distribution of substances among various properties including, but not limited to, chemical class, product class, physical form, irritancy severity classification, mechanism of action, physical and chemical characteristics, and molecular weight. In situations where a listed substance is not available, other substances of the same class for which there is high quality *in vivo* reference data may be used. Following completion of optimization and/or validation studies, substances from this list can be selected for inclusion in performance standards and proficiency testing (ICCVAM 2003).

[21]The recommendations discussed here are based on the ability of the *in vitro* test method to identify *in vivo* classifications based on the GHS classification system.

[This Page Intentionally Left Blank]

8.0 REFERENCES

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

EU. 2004. Manual of Decisions for Implementation of the Sixth and Seventh Amendments to Directive 67/548/EEC on Dangerous Substances (Directives 79/831/EEC and 92/32/EEC). European Chemicals Bureau, IHCP, European Commission DG-JRC, report in. EUR 20519 EN, updated version of July 2004.

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L255:1-333.

ICCVAM. 2006a. ICCVAM Evaluation of the Bovine Corneal Opacity and Permeability Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#bcop [accessed 10 Jan 2006].

ICCVAM. 2006b. ICCVAM Evaluation of the Isolated Chicken Eye Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ice [accessed 10 Jan 2006].

ICCVAM. 2006c. ICCVAM Evaluation of the Isolated Rabbit Eye Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#ire [accessed 10 Jan 2006].

ICCVAM. 2006d. ICCVAM Evaluation of the Hen's Egg Test – Chorioallantoic Membrane Test for the Identification of Severe Ocular Irritants and Corrosives. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm#hetcam [accessed 10 Jan 2006].

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. 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.

Kalweit S, Gerner I, Spielmann H. 1987. Validation project of alternatives for the Draize eye test. Mol Toxic 1:597-603.

Luepke N. 1985. Hen's egg chorioallantoic membrane test for irritation potential. Food Chem Toxic 23:287-291.

National Research Council. 1996. Guide for the Care and Use of Laboratory Animals. Washington, DC: National Academy Press.

NLM. 2005. Medical Subject Heading. Available: http://www.nlm.nih.gov/mesh [accessed 09 January 2006].

PHS. 2002. Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals.

Ubels JL, Paauw JD, Casterton PL, Kool DJ. 2002. A redesigned corneal holder for the bovine cornea opacity and permeability assay that maintains normal corneal morphology. Toxicol In Vitro 16:621-628.

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html. [accessed 16 October 2006].