BACKGROUND REVIEW DOCUMENT

Current Status of In Vitro Test Methods for Identifying Ocular Irritants :

Short Time Exposure (STE) test

Prepared by:

Kao Corporation
2606, Akabane Ichikai-Machi, Haga-gun
Tochigi 321-3497
Japan

Authers:

Yuko Nukada, Ph.D.
Kazuhiko Hayashi, M.S.
Takayuki Abo, M.S.
Taeko Mori, D.V.M.
Hitoshi Sakaguchi, Ph.D.

Final revision

May 2012

TABLE OF CONTENTS

Page Number

LIST OF TABLES

Page Number

LIST OF FIGURES

Page Number

LIST OF ACRONYMS AND ABBREVIATIONS

BCOP	Bovine Corneal Opacity and Permeability
BRD	Background Review Document
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service Registry Number
CV	Coefficient of variation
CM	Cytosensor Microphysiometer
EC	European Commission
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECVAM	European Centre for the Validation of Alternative Methods
EDTA	Ethylenediaminetetraacetic acid
EEC	European Economic Community
EPA	(U.S.) Environmental Protection Agency
EU	European Union
FBS	Fetal bovine serum
FDA	(U.S.) Food and Drug Administration
FL	Fluorescein Leakage
g	Gram
GHS	Globally Harmonized System (of Classification and Labeling of Chemicals)
GLP	Good Laboratory Practices
HET-CAM	Hen's Egg Test – Chorioallantoic Membrane
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye
µg	Microgram
µL	Microliter
MAS	Maximum average score
MEM	Minimum Essential Medium
MeSH	(National Library of Medicine) Medical subject heading
mg	Milligram
mL	Milliliter
MMAS	Modified maximum average score
NA	Not applicable
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

OD	Optical density
OECD	Organisation for Economic Co-operation and Development
SD	Standard deviation
SLS	sodium Lauryl Sulfate
TG	Test Guideline
UN	United Nations
v/v	Volume to volute ratio
w/w	Weight to weight ratio

PREFACE

Ocular irritation is a reaction caused by the direct contact of a chemical substance with the eye, inducing symptoms such as clouding of the cornea, inflammation of the iris, and redness/edema/secretion of the conjunctiva. It is important to assess ocular irritation, especially in products used on the face (such as cosmetics) or hair or household products, any of which can accidentally enter the eye.

The Draize test (Draize et al., 1944) using rabbits has been widely used to evaluate ocular irritation. In the Draize test, 0.1 ml or 0.1 g of a test substance is instilled into the palpebra of a rabbit; reactions in the cornea, iris, and conjunctiva are then macroscopically judged over time on the basis of a set of evaluation criteria. In evaluating the cornea, a maximum of 80 points are assigned on the basis of degree and area of opacity; for the iris, a maximum of 10 points are assigned on the basis of degree of congestion, swelling, and bleeding; and for the conjunctiva, a maximum of 20 points are assigned on the basis of redness, edema, and secretion. Thus, the total score is a maximum of 110 points. More weight is placed on changes in the cornea—as reflected in the higher number of points assigned there—given the significance of corneal injury. In this test, recovery from a reaction can be evaluated through successive judgments. Degree of irritation is evaluated on the basis of judgments made, and the four-step evaluation using the Maximum Average Score (MAS) obtained during the observation period (Kay and Calandra 1962) is used as the judgment standard. The ocular irritation tests described in the Organization for Economic Co-Operation and Development (OECD) test guidelines (OECD number 405, 1987) and the “Guidance for cosmetic safety evaluation, 2008” (edited by the Japan Cosmetic Industry Association, 2008) are basically those of the Draize eye irritation test.

The public interests in animal alternative tests have increased recently and the development of these tests has become a critical task for the cosmetic industry globally. In addition, the development of alternative methods is accelerating in the world due to new regulations like the banning of cosmetics in animal ocular irritation tests in the EU (Directive 2003/15/EC, 2003). A lot of ocular irritation alternative methods that use various cell lines and tissues are being developed around the world (Balls et al., 1999; Ohno et al., 1999; Eskes et al., 2005). The Bovine Corneal Opacity and Permeability (BCOP) test method and Isolated Chicken Eye (ICE) test method have only been accepted as OECD TG for predicting severe ocular irritation in last year. However, no other *in vitro* assay was accepted as a TG.

The Short Time Exposure (STE) test is an alternative ocular irritation test method developed by Takahashi et al. The STE test was well characterized in terms of employing cultured cells lines derived from cornea and possessing similar or shorter exposure times than many other cytotoxicity based methods, e.g. SIRC cells using crystal violet staining (SIRC-CVS) (Itagaki et al., 1991) and SIRC cells using neutral red uptake (SIRC-NRU) (Okamoto et al., 1990). The STE test has the advantage of being able to evaluate the ocular irritation potential of water insoluble chemicals (e.g. toluene and hexanol) by using mineral oil as test vehicle (Takahashi et al., 2008).

Category classification of ocular irritation by STE test is determined based on the relative viability assessed for 5% test concentrations. A concentration of test material that had a relative viability of 70% or less was categorized as an irritant (I) and a concentration of test material that had a relative viability greater than 70% was categorized as a non-irritant (NI). For STE test, as secondary approach in order to establish an ocular irritation potency ranking, a point system based on the test concentration and relative viability resulting from an exposure to 5% or 0.05% of test material was also proposed (Takahashi et al., 2008).

The objective of this background review document (BRD) is to describe the current validation status of the STE test, including what is known about its accuracy and reliability. However, a point system was not used for analysis of accuracy and reliability in this BRD.

EXECUTIVE SUMMARY

This Background Review Document (BRD) reviews available data and information regarding the validation status of the Short Time Exposure (STE) test for identifying ocular irritants. The test method was reviewed for its ability to predict ocular irritant as defined by the United Nations (UN) Globally Harmonized System (GHS) of classification and labeling of chemicals (UN 2003) and the U.S. Environmental Protection Agency (EPA) (EPA 1996) . The objective of this BRD is to describe the current validation status of the STE test, including its accuracy and reliability.

The information summarized in this BRD is based on publications obtained from the peer-reviewed literatures. A total of six publications that contained STE test results and protocol information were existed, of these publications, four publications that contained the STE test results obtained from two to five labs and GHS classifications (or Draize data for one publication) allowed for an evaluation of test method accuracy and reliability. However, in this BRD, all of the ocular irritancy classification (i.e, EPA [EPA 1996], and GHS [UN 2003]) of the test substances were all reclassified based mainly on the available *in vivo* data listed in the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Data Bank (ECETOC, 1998) and cited literature (e.g. Ohno Y. et al., 1999). Additionally, some in-house data and unpublished data provided by ECVAM were also used.

Other published STE test studies are reviewed in Section 9.0 (Other Scientific Reports). In this section the performance analysis (accuracy and intra-laboratory reproducibility) was not conducted again, the results in these publications were described without modifications. However, some of these data and some of unpublished data for STE test were used to analyze the overall accuracy of STE in Section 6 (STE test accuracy).

The STE test is an alternative method for identification of ocular irritant developed by Takahashi et al. (2008). The STE test was well characterized in terms of employing cultured cells lines derived from cornea and possessing similar or shorter exposure times (5min) than many other cytotoxicity based methods, e.g. SIRC cells using crystal violet staining (SIRC-CVS) (Itagaki et al., 1991) and SIRC cells using neutral red uptake (SIRC-NRU) (Okamoto et al., 1990).

A total of 119 substances were evaluated in the four validation/prevalidation studies and original study. A variety of chemical classes have been tested in the STE test. The chemical

classes with the greatest amount of STE test data was alcohols. Other chemical classes tested include, esters, surfactants (nonionic), ketones/lactones, surfactants (cationic), amines, organic salts, carboxylic acids and surfactants (anionic).

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

The accuracy evaluation of the STE test was limited to the substances evaluated in four *in vitro-in vivo* comparative studies. The ability of the STE test to correctly identify ocular irritants, as defined by the GHS (UN 2003) and the EPA (1996) was evaluated using two approaches. In the first approach, the accuracy of STE test was assessed separately for each *in vitro-in vivo* comparative study. In the second approach, the accuracy of STE test was assessed after pooling data across *in vitro-in vivo* comparative studies. The overall accuracy of the STE test ranged from 74% to 85%, depending on the classification system used. Sensitivity and specificity ranged from 68% to 84% and from 86% to 96%, respectively. The false positive rate ranged from 4% to 14%, while the false negative rate ranged from 14% to 32%.

For GHS classification, the accuracy analysis indicated that salts, ketones/lactones and esters are often overpredicted (100% [1/1], 25% [1/4] and 22% [2/9] false positive rate, respectively) in the STE test. In contrast, organic/Inorganic salts were most often underpredicted by the STE test (50% [2/4] false negative rate). Hydrocarbons (50% [2/4]), esters (25% [1/4]), and alcohols (25% [4/16]) also had high false negative rates. The numbers of substances among the remaining chemical classes were too few to resolve any definitive trends in false prediction by the STE test. For the purposes of these analyses, we considered three substances to be the threshold number per chemical class for consideration, and thus chemical classes represented by fewer than five substances were not considered.

For EPA classification, the accuracy analysis indicated that only one overpredicted (Polyethyleneglycol monolaurate (10E.O.)) substance was identified in the STE test. In

contrast, hydrocarbons (60% [7/9] false negative rate) were most often underpredicted by the STE test. Esters (50% [5/10]), salts (40% [2/5]), ketones/lactones (33% [2/6]) and alcohols (29% [5/17]) also had high false negative rates.

Exclusion of three discordant classes (i.e., organic/inorganic salts, ester and alcohols) from the data set resulted in an increased accuracy (from 85% to 90% [GHS], from 75% to 78% [EPA]), and a decreased false negative rate (from 16% to 7% [GHS], 32% to 29% [EPA]).

It's notable that 17 or 16 substances labeled as surfactants were not underpredicted by the STE test regardless the classification system used.

With regard to physical form of the substances overpredicted by the STE test, zero to five was liquids and zero or one was solids. Although the minor differences were existed depending on the classification system used, considering the proportion of the total available data, the rate of overprediction of liquids (74/97 and 73/96) and solids (23/97 and 23/96) were generally seems to be equivalent by the STE test.

With regard to physical form of the substances underpredicted by the STE test, five or six were solids and four to 18 were liquid. Although the minor differences were existed depending on the classification system used, despite the proportion of the total available database indicated above, the rate of underprediction of liquids and solids were generally seems to be equivalent by the STE test.

Among the underpredicted substances for which pH information was available, as one was acidic (pH ≤ 7.0) and zero was basic (pH > 7.0) regardless the classification system used. Therefore, the acidic substances (26/34; 76% or 25/33; 76%) may have a tendency to underestimate. However, it is noted that pH information was available for only 35 substances in all substances with *in vivo* data. The numbers of substance among the overpredicted substances were too few to resolve any definitive trends in overprediction by STE test.

With regard to volatility of the substances underpredicted by the STE test, the accuracy analysis indicated that the chemicals with the vapor pressure between 10kPa and 1kPa often underpredicted for GHS and EPA classification systems (50% [4/8] and 67% [8/12] false negative rate, respectively) in the STE test. When the substance with vapor pressure over 6kPa was excluded from the data set, changes performance statistics were noted regardless

the classification system used. When the substances with vapor pressure around over 6kPa were excluded from the data set; accuracy increased (from 86% to 88% [GHS], and from 71% to 73% [EPA]), and the false negative rate decreased (from 20% to 16% [GHS], and from 38% to 36% [EPA])

The substances with insoluble either in saline, saline with 5% DMSO and mineral oil are enable to assay. Colored test substances may be problematic as they could interfere with the optical density measured in MTT assay. In addition, the substances, which cause unexpected color change by direct MTT reduction, may be misjudged. When these substances were assigned as “non irritant” in STE test, it would be finally assigned as “inconclusive”. Moreover, as a result of the findings regarding the predictive capacity of the STE test, the false negative rates of high volatile substances, or inorganic/organic salts, alcohol, and hydrocarbons belonging to the solid substances were relatively high compared to that of other substances regardless the classification system used.

The BRD analysis indicated that solid salts, solid alcohols, solid hydrocarbons and high volatile substances with vapor pressure over around 6 kPa seems to be out of applicability domain of the STE test, regardless the classification system used. Moreover, the BRD analysis also indicated that the surfactant can be evaluated by STE test even if the substance was categorized into solid salts, solid alcohols, solid hydrocarbons, and high volatile substances.

Therefore, the possibility of a tiered approach combining the STE test, the EpiOcular assay, and the BCOP assay for predicting the eye irritation potential of substances not soluble in saline or mineral oil was assessed. The possibility of achieving accurate estimation of irritation potential for solid salts, solid alcohols, solid hydrocarbons and high volatile substances was also assessed. As a result, the tiered approach was allowed to estimate the eye irritation potential of not only insoluble substance but also the substances, which were categorized into solid salts, solid alcohols, solid hydrocarbons and high volatile substances accurately. From these results, this tiered approach might be a promising alternative eye irritation testing strategy capable of testing for wide range of test substances regardless of solubility and volatility with minimum under prediction (Hayashi et al., 2012a; Hayashi et al., 2012b).

A quantitative assessment of intra-laboratory data (viability values) from four studies (Takahashi et al. 2009, Takahashi et al. 2010, Sakaguchi et al. 2011, and Kojima et al. 2012) provides an indication of the extent of intra-laboratory repeatability of the STE test for

substances predicted as ocular irritants. For the 44 substances evaluated in the Takahashi (2009) study, the mean and median %CV values for viability values for replicate were in the range of 44.5 to 72.0 and 10.2 to 20.6 for the three laboratories, respectively. For the 70 substances evaluated in the Takahashi et al. 2010, the within experiment mean and median %CV values for viability values for replicate were in the range of 42.3 to 51.0 and 13.2 to 15.8 for the two laboratories, respectively. For the 25 substances evaluated in the Sakaguchi et al. 2011, the within experiment mean and median %CV values for viability values for replicate were in the range of 15.8 to 35.6 and 8.5 to 10.4 for the five laboratories, respectively. For the 40 substances evaluated in the Kojima et al. 2012, the within experiment mean and median %CV values for viability values for replicate were in the range of 30.2 to 51.0 and 14.5 to 35.5 for the two or three laboratories, respectively.

A qualitative assessment of the data provided for multiple laboratories in four studies (Takahashi et al. 2009, Takahashi et al. 2010, Sakaguchi et al. 2011, and Kojima et al. 2012) provides an indication of the extent of interlaboratory reproducibility. In an assessment of interlaboratory reproducibility of hazard classification (GHS or EPA), the three participating laboratories for the Takahashi et al. (2009) study, regardless of the classification system used, there was 100% agreement in regard to the ocular irritancy classification for all substances tested in the study. For the study by Takahashi et al. (2010), regardless of the classification system used, there was 100% agreement in regard to the ocular irritancy classification for all substances, which were tested in 2 laboratories. For the study by Sakaguchi et al. (2011), there was 100% agreement in regard to the ocular irritancy classification for 20 (83%) of the 24 substances against GHS classification or for 20 (87%) on the 23 substances against EPA classification, respectively. For the study by Kojima et al. (2012), there was 100% agreement in regard to the ocular irritancy classification for 33 (94%) of the 35 substances, regardless of the classification system used. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones/lactones, cationic surfactants and ester compounds.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies (Takahashi et al. 2009, Sakaguchi et al. 2011, and Kojima et al. 2012) by performing a %CV analysis of viability values obtained for substances tested in over than three laboratories. For the Takahashi et al. (2009) study, the 44 test substances had mean and median %CV values of 56.7 % and 11.2 %, respectively, for results obtained in three laboratories. For the Sakaguchi et al. (2011) study, the 25 test substances had mean and median %CV values of 32.3% and 8.6%, respectively, for results obtained in five laboratories. For the Kojima et al.

(2012) study, the mean and median %CV values for the viability values of the 10 substances that were evaluated for three laboratories were 58.8% and 51.4%, respectively, for three laboratories.

The STE test using cultured cells have the advantage of being simple, a quick procedure, and a low evaluation cost. Furthermore, poorly water-soluble chemicals like toluene, octanol, and hexanol could be evaluated in the STE test by using mineral oil as the vehicle. Therefore, the STE test could be considered a building block assay in the tiered, especially bottom up, approach to establishing an ocular irritation animal alternative testing paradigm.

As stated above, this BRD provides a comprehensive summary of the current validation status of the STE test, including what is known about its reliability and accuracy, and the scopes of the substances tested.

1.0 Introduction and Rationale for the Proposed use of In Vitro Test Methods to Identify Ocular Irritants

1.1 Introduction

1.1.1 Historical Background of *In Vitro* Ocular Irritation Tests and Rationale for Their Development

For many years, the ocular irritation potential of chemicals mostly has been evaluated by the Draize test. The public interests in animal alternative tests have increased recently and the development of these tests has become a critical task for the cosmetic industry globally. In addition, the development of alternative methods is accelerating in the world due to new regulations like the banning of cosmetics in animal ocular irritation tests in the EU (Directive 2003/15/EC, 2003). A lot of ocular irritation alternative methods that use various cell lines and tissues are being developed around the world (Balls et al., 1999; Ohno et al., 1999; Eskes et al., 2005). The Bovine Corneal Opacity and Permeability (BCOP) test method and Isolated Chicken Eye (ICE) test method have only been accepted as OECD TG for predicting severe ocular irritation in last year. However, no other *in vitro* assay was accepted as a TG.

The Short Time Exposure (STE) test is an alternative ocular irritation method developed by Takahashi et al. (2008). The STE test was well characterized in terms of employing cultured cells lines derived from cornea and possessing similar or shorter exposure times than many other cytotoxicity based methods, e.g. SIRC cells using crystal violet staining (SIRC-CVS) (Itagaki et al., 1991) and SIRC cells using neutral red uptake (SIRC-NRU) (Okamoto et al., 1990).

Generally, cytotoxicity tests using cultured cells have the advantage of being simple, a quick procedure, and a low evaluation cost. In addition, since the scattering of results from multiple replicas for one sample would be small, intra-laboratory and inter-laboratory reproducibility would be good for a cytotoxicity test. Such a method could be easily standardized as an alternative test method. However, this theoretical method will have some issues such as not being able to evaluate water insoluble materials, acids, alkalis, and alcohols as well as having the test sample be neutralized by the buffering capacity of medium (Bagley et al., 1994; Harbell et al., 1997; Ohno, 1999).

The STE test has the advantage of being able to evaluate the ocular irritation potential of water insoluble chemicals (e.g. toluene and hexanol) by using mineral oil as test vehicle

(Takahashi et al., 2008).

This test method was based on the cytotoxixity using cultured cell line derived from rabbit cornea because the cornea cells is one of the main targets during accidental eye exposures, and damage to the cornea can result in visual impairment or loss. In addition, corneal effects are weighted heavily in the original *in vivo* ocular irritancy scoring systems (e.g., 80 out of a possible 110 points in the Draize eye test scoring system).

In the STE test, cytotoxicity is determined by the viability of SIRC cells. The viability of cells is conventionally measured by MTT assay method. While these *in vitro* toxicity measurements using the corneal cell are correlated with *in vivo* ocular irritation corneal effects, they represent only one aspect of the overall complex response of the eye to irritants, which involves other tissues such as the iris and conjunctiva.

For the ocular irritation animal alternative test, it may be unlikely to completely replace the Draize test by a single *in vitro* test because the Draize test evaluates a range of criteria for injury and inflammation to the eye. The tiered approach of several *in vitro* assays combined was proposed in order to estimate the irritation potential for a wide range of chemical classes (Hagino et al., 2008; McNamee et al., 2009; Scott et al., 2010).

The STE test, an alternative ocular irritation test, involves exposing SIRC (rabbit corneal cell line) cells for 5 min to a 5% concentration of test material. Furthermore, poorly water-soluble chemicals like toluene, octanol, and hexanol could be evaluated in the STE test by using mineral oil as the vehicle (Takahashi et al., 2008). For these reasons, the STE test could be considered a building block assay in the tiered approach to establishing an ocular irritation animal alternative testing paradigm.

The STE test is currently used to by Kao Corporation as in-house method to assess the ocular irritation potential of industry chemicals, cosmetics and personal care product etc. For non-registered household products, the STE test is used to predict the relative ocular irritation potential of newly developed products compared to products on the market or substances for which the ocular irritation potential has already been determined.

1.1.2 Overview of prior development and validation activities

The Ministry Health and Welfare (MHW) Scientific Study Group conducted validation studies of several some *in vitro* ocular irritation tests (e.g., RBC assay, SIRC-CVS assay, SIRC-NRU

assay, HeLa-MTT assay, CHL-CVS assay). Accuracy between the results of RBC assay and *in vivo* Draize data was 70% (21/30) when a Draize irritation score of 15 points was used as a cut-off value. Meanwhile, the accuracy of the SIRC-CVS assay, SIRC-NRU assay, HeLa-MTT assay, and CHL-CVS assay was around 71% (24/34). (Ohno et al., 1999).

Recently, a retrospective validation activities are ongoing of *in vitro* assays (Neutral Red Release: NRR, Red Blood Cell: RBC, Fluorescein Leakage: FL, Cytocensor Microphysiometer: CM) is the European Center for the Validation of Alternative Methods (ECVAM) Scientific Advisory Committee (ESAC) conducted (McNamee et al., 2009). Draft guidelines of FL and CM were proposed on July 2010.

1.2 Scientific basis for the proposed test

1.2.1 Purpose and Mechanistic Basis of STE test

The STE test is a cytotoxicity-based method that used SIRC cells. As mentioned above, the STE test was developed as an alternative ocular irritation test method in order to obviate the need for laboratory animals as the source for test eyes. It was reported that the 90% of solution dropped into the eye will excrete in 1-2 minutes in human, moreover, 80% of that will excrete through the conjunctival sac in 3-4 minutes in rabbit (Mikkelson et al., 1973; Motose, 1984). Therefore, none of the test material seems to retention in the eye over 5 minutes. As mentioned in section 1.1.1, the cornea cells are one of the main targets during accidental eye exposures. Substances such as surfactants and organic compounds that can lyse cell membranes and aggregated proteins are cytotoxicity immediately. In order to reflect an actual exposure situation mentioned above, the endpoint evaluated in the STE test to measure the extent of damage to the SIRC cells following exposure to a chemical substance is cytotoxicity. Cytotoxicity is quantitatively measured by the relative viability of SIRC cells. Cell viability is measured by MTT assay method. Decrease of cell viability is significant adverse of some irritants that can lead to corneal damage.

1.2.2 Similarities and Differences of Modes of Action Between the STE test and Ocular Irritancy in Humans and/or Rabbits

1.2.2.1 *The In Vivo Rabbit Eye Test Method*

For many years, the ocular irritation potential of chemicals mostly has been evaluated by the Draize test. This test method involves instillation of the test substance into the lower conjunctival sac of the rabbit eye, and evaluates the cornea, the iris, and the conjunctiva for adverse effects after exposure to the potential irritant. The cornea is evaluated both for the

degree of corneal opacity and the area of the cornea in which opacity is involved. The iris is assessed for inflammation, iridal folds, congestion, swelling, circumcorneal injection, reaction to light, hemorrhage, and gross destruction. The conjunctiva is evaluated for the degree of redness, chemosis (swelling), and discharge (Draize et al. 1944).

1.2.2.2 *Comparison of STE test with the In Vivo Rabbit Eye Test*

The STE method is an alternative ocular irritation method developed by Takahashi et al. (2008). The STE test was well characterized in terms of employing cultured cells lines derived from cornea and possessing similar or shorter exposure times (5min) than many other cytotoxicity based methods, e.g. SIRC cells using crystal violet staining (SIRC-CVS) (Itagaki et al., 1991) and SIRC cells using neutral red uptake (SIRC-NRU) (Okamoto et al., 1990). In the STE test, cytotoxicity is determined by the viability of the SIRC cells. The viability is measured by MTT assay method. While these *in vitro* toxicity measurements using the cultured cell line are correlated with *in vivo* ocular irritation corneal effects, they represent only one aspect of the overall complex response of the eye to irritants, which involves other tissues such as the iris and conjunctiva.

In contrast, the *in vivo* rabbit eye test involves a qualitative visual evaluation of the severity of adverse effects on the cornea, the iris, and the conjunctiva, as well as the reversibility of any ocular effects detected at selected intervals up to 21 days after exposure. In STE test, a test substance is exposed as solution (5%) to cells for just 5 min, and then rinsed off. In the *in vivo* rabbit eye test, test substances are applied to the conjunctival sac. Because the rabbit eye can blink and/or tear, exposure of the cornea to the test substance will be affected by these factors in terms of coverage or duration. The neurogenic components that drive tear film production are also not present in the STE test. When compared with an *in vivo* rabbit eye study, application of a test substance in the absence of this protective barrier might be expected to cause an increase in false positive outcomes. On the other hands, in some test substances (e.g., solids), blinking can also induce mechanical damage *in vivo*, contributing to a higher degree of irritation. However, this protective mechanism for the eye are absent in the STE test. Moreover, the STE test does not account for systemic effects following ocular instillation that may be noted with the *in vivo* rabbit eye test (e.g., toxicity or lethality as in the case of certain pesticides).

1.2.3 Intended Range of Substances Amenable to the STE test and/or Limits of STE test

While a wide range of substances with various physicochemical characteristics can be tested in the STE test, substances with insoluble either in saline, saline with 5% DMSO and mineral oil are enable to assay. Colored test substances may be problematic as they could interfere with the optical density measured in MTT assay. In cases where the substances are shown to reduce MTT directly, only substances that remain bound to the cells after washing, resulting in a false MTT reduction signal. This will cause the under-prediction.

Sakaguchi et al. (2011) noted the results found that the STE test provided an excellent predictive ability. However, there were a few chemicals that exhibited cell viability around 70% in the STE test. Those that did have viability around 70% seemed also to have higher variability in classification between laboratories. In addition, since scattering in the intermediate range (around 20–85%) of mean cell viability was relatively high, the interpretation of classification results must be performed carefully. Hence, many more chemicals with cell viability near the 70% cut-off point need to be evaluated and added to the databank for future analysis. On the other hands, among the 25 chemicals evaluated in the present validation study, ethanol and cyclohexanol had predictive rankings in the STE test in all laboratories that differed from the rankings of the GHS classification. In contrast, n-hexanol, 1-octanol, and 2-ethyl-1-hexanol were evaluated correctly by the STE test. Although alcohols can be evaluated in the STE test, alcohols should be considered chemicals that could generate false negative results or be predicted to have a weaker toxicity potential.

1.3 Regulatory rationale and applicability

1.3.1 Current Regulatory Testing Requirements

In recent years, several regulations and regulatory agencies have contributed to a greater emphasis on alternative animal testing for ocular irritation (7th amendment to the Cosmetic Directive [Directive 2003/15/EC, 2003], Registration Evaluation Authorization and Restriction of Chemicals [REACH]).

As described below in Section 1.1.1, for the assessment of ocular irritation, one *in vitro* alternative test may not completely replace the Draize test. Therefore, a tiered approach combining several *in vitro* assays, including cytotoxicity assays, is proposed in order to estimate the irritation potential for a wide range of chemical classes. The STE test is a cytotoxicity test involves exposing to SIRC cells. Furthermore, poorly water-soluble chemicals like toluene, octanol, and hexanol could be evaluated in the STE test by using mineral oil as

the vehicle (Takahashi et al., 2008). For these reasons, the STE test could be considered a building block assay in the tiered approach to establishing an ocular irritation animal alternative testing paradigm.

1.3.2 Intended regulatory use(s)

Cell-based cytotoxicity test (e.g. NRR, RBC) such as the STE test have been proposed for identifying of ocular irritancy (e.g., Category 1 or Category 2 per the GHS classification system [UN 2003], and Category I to III per the EPA classification system [EPA 1996]).

1.3.3 The similarities and differences in the endpoint measured in the proposed test and currently used *in vivo* reference test

As mentioned in Section 1.1.1, this test method was based on the cytotoxicity of cornea cells because the cornea cells is one of the main targets during accidental eye exposures, and damage to the cornea can result in visual impairment or loss.

In the STE test, cytotoxicity is determined by the viability of the SIRC cells. The viability is measured by MTT assay method. While these *in vitro* toxicity measurements using the cultured cell line are correlated with *in vivo* ocular irritation corneal effects, they represent only one aspect of the overall complex response of the eye to irritants, which involves other tissues such as the iris and conjunctiva.

1.3.4 How the proposed test fits into the overall strategy of hazard or safety

As mentioned in Section 1.2.1, for the ocular irritation animal alternative test, it may be unlikely to completely replace the Draize test by a single *in vitro* test because the Draize test evaluates a range of criteria for injury and inflammation to the eye. The tiered approach combined several *in vitro* assays was proposed in order to estimate the irritation potential for a wide range of chemical classes (Hagino et al., 2008; McNamee et al., 2009; Scott et al., 2010).

The STE test is being considered for use in identification ocular irritancy (e.g., GHS category 1 or category 2). For these reasons, the STE test could be considered a building block assay in the tiered approach to establishing an ocular irritation animal alternative testing paradigm.

2.0 STE test Protocol Components

2.1 Overview of How the Test is Conducted

A protocol of the present test is attached as **Appendix A**, and procedures are described in greater detail below.

The procedure of Takahashi et al. (2008) is used. Briefly, physiological saline (Otsuka Pharmaceutical Co., Ltd. Tokyo, Japan) is used as first vehicle for test chemicals. If the chemicals revealed low aqueous solubility, 5% dimethyl sulfoxide (DMSO, Sigma-Aldrich Co.) in saline is used as 2nd vehicle. In case of water insoluble test chemicals, mineral oil (Sigma-Aldrich Co.) is used as 3rd vehicle. When saline, DMSO or mineral oil is used with a test sample, similar vehicle conditions are used as the corresponding control samples. The cells cultured in 96-well plates are exposed to 200μL of 5% test chemical solutions for 5 minutes. After exposure, the cells are washed with phosphate buffered saline (-) [PBS (-); Takara Bio Inc., Siga, Japan] twice and 200μl of methylthiazolydiphenyl-tetrazolium bromide (MTT, Sigma Aldrich) solution (0.5mg MTT/ml of medium) is added. After a 2h reaction time, MTT formazan is extracted with 0.04 N HCl-isopropanol (Kanto Chemical Co., Inc., Tokyo, Japan) for 30min, and the absorbance of the extract is measured at 570nm with a plate reader (Lab A : Corona Electric Co., Ltd., Ibaraki, Japan, Lab B : DS Phama Biomedical Co., Ltd. Osaka, Japan, Lab C : Thermo Fisher Scientific Co., Ltd. Kanagawa, Japan). The ratio of absorbance (%) on each test sample to that of control is represented as relative viability (triplicate determinations). The control group cells are exposed to physiological saline, saline with 5% DMSO, or mineral oil. The mean of three wells for each test concentration is calculated. This is the mean relative viability for one independent test. A total of three independent tests are conducted for each concentration of a test material, and the calculated overall mean of three independent tests is used for estimation of ocular irritation.

$$\text{Cell Viability} = \frac{\text{Absorbance of Test Sample (Absorbance of Test Sample - Blank)}}{\text{Absorbance of Solvent Control (Absorbance of Solvent Control - Blank)}} \times 100$$

Category classification of ocular irritation by STE test is determined based on the relative viability assessed for 5% test concentrations. A concentration of test material that have a relative viability of 70% or less is categorized as an irritant (I) and a concentration of test material that had a relative viability greater than 70% is categorized as a non-irritant (NI). The GHS classifications of the chemicals are estimated as NI (not classified or not an eye

irritant), and I (an eye irritant of category 2 or category 1) based on the Draize data listed in the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Data Bank (ECETOC, 1998), the publicly-available documents (such as Appendix H in ICCVAM, 2006) or unpublished data (in-house data or data from ICCVAM referred from NIHS Japan, TSCA, ZEBET, and NLDLS in **Appendix B2**). By comparing between STE category classification at 5% test concentration and GHS classification, the predictive capacity is confirmed. The following parameters are determined by the method of Cooper et al. (1982): sensitivity (percent of "I" chemicals classified by STE test), specificity (percent of "NI" chemicals classified by STE test), positive and negative predictivity (percent of chemicals classified by STE that are true NI or I) and accuracy (total percent of exact classification). The predictive capacity based on ocular irritation category classification is evaluated in terms of these parameters.

Classification	Viability
Non-irritant	> 70%
Irritant	≤ 70%

2.2 Description and Rationale for the Test Components

2.2.1 Materials, equipment, and supplies needed

2.2.1.1 *Cells*

As mentioned in Section 2.1, SIRC (rabbit corneal cell line, ATCC CCL-60) cells are obtained from American Type Culture Collection (Manassas, VA, USA). The cells should be used between 3 weeks and 3 months after the start of cultivation or within 25 passages.

2.2.1.2 *Instrument to Measure Viability*

The viability of cells is conventionally measured by MTT assay method. Resulting MTT formazan is extracted with 0.04 N HCl-isopropanol, and the absorbance of the extract is measured at 570 nm with a micro plate reader made by various manufacturer (e.g., Corona Electric Co., Ltd., Ibaraki, Japan, DS Phama Biomedical Co., Ltd. Osaka, Japan, Thermo Fisher Scientific Co., Ltd. Kanagawa, Japan).

2.2.2 Dose-Selection Procedures, Including the Need for Any Dose Range-Finding Studies

As described below in Section 2.1, test substances are applied at the concentration of 5%. Therefore, the dose-selection procedure is not conduct for STE test.

2.2.3 Endpoint(s) measured

As described below in Section 1.1.1, the Test was based on the cytotoxicity of cornea cells because the cornea cells is one of the main targets during accidental eye exposures, and damage to the cornea can result in visual impairment or loss. In the STE test, cytotoxicity is determined by the viability of the rabbit corneal cell line (SIRC cells). The viability of cells is conventionally measured by MTT assay method.

2.2.4 Duration of exposure

2.2.4.1 *Pre-Exposure Preparations*

As mentioned in Section 2.1, SIRC cells are cultured in Eagle's MEM (EMEM, Sigma-Aldrich) containing 10% (v/v) fetal bovine serum , 2mM L-glutamine , 50units/mL penicillin , and 50µg/mL streptomycin (Invitrogen Co., Carlsbad, CA, USA). When the cells proliferated in the culture flask to confluence, the cells are dispersed with trypsin-EDTA solution (Sigma-Aldrich). The dispersed cells are spread into 96-well flat-bottomed plates (Corning Coster Co., Cambridge, MA) at 3.0×10^3 cells/well. After incubation (37°C, 5% CO_2) for 5 days (or 6.0×10^3 cells/well for 4 days), the cells reach confluence.

2.2.4.2 *Selection of vehicles for test substance preparation*

As mentioned in Section 2.1, physiological saline (Otsuka Pharmaceutical Co., Ltd. Tokyo, Japan) is used as first vehicle for test chemicals. If the chemicals revealed low aqueous solubility, 5% dimethyl sulfoxide (DMSO, Sigma-Aldrich Co.) in saline is used as 2nd vehicle. In case of water insoluble test chemicals, mineral oil (Sigma-Aldrich Co.) is used as 3rd vehicle.

2.2.4.3 *Test substance exposure volume*

A 200µL of test sample is applied to cells in STE test. This exposure volume was decided from the view point both of ease of sample solution application and maximum volume per well of 96-well microplate.

2.2.4.4 *Concentration tested*

In STE, a monolayer of SIRC cells is used for assay. In contrast, the cornea tissue consists of multilayer of cornea epithelium. Due to the differences, monolayer cells are susceptible for the substances compared to the tissue. It is difficult to rinse completely for viscous substance after neat exposure. And the neat exposure may obtain unnecessarily severe toxicity by osmotic pressure of the substance. Therefore, as described below in Section 2.1, test substance is applied at the concentration of 5% in STE test. To address the correspondence

between the category classification by the STE test and GHS classification, 24 chemicals were tested by STE test at six concentrations from 10% to 0.05% (**Table 2-1**). It was found that the test substance concentration of over 5%, with cut-off value of 70%, produced better correlation to *in vivo* results for 24 chemicals (**Table 2-2**). However, when the 10% was applied for the STE test as a test substance concentration, the number of insoluble test substances in the vehicle was markedly increased (in house data was not shown). Based on these results, the test concentration of 5% was decided to apply for STE test.

Table 2-1 Viability values of test substances for several concentrations (test substances were exposed to cells for 5 min)

Chemical	GHS	Concentration tested					
		10%	5%	1%	0.5%	0.1%	0.05%
Benzalkonium chloride	1	3.1	2.1	3.5	5.3	6.8	3.1
Cetylpyridinium bromide	1	-0.5	0.6	1.8	2.0	8.8	4.2
Cyclohexanol	1	3.3	1.4	6.1	91.5	97.9	104.5
Sodium hydroxide	1	0.7	1.4	-0.3	0.6	-0.4	-1.4
TritonX-100	1	0.2	-0.1	3.9	0.7	-0.2	0.7
1-Octanol	2A	1.6	-0.5	84.2	89.1	102.8	96.8
2-Ethyl-1-hexanol	2A	6.6	44.0	95.6	95.4	100.6	93.4
Acetone	2A	55.6	9.6	98.2	98.5	99.3	101.4
Ethanol	2A	40.1	98.2	98.9	108.2	97.7	97.1
Methyl ethyl ketone (2-butanone)	2A	20.6	44.7	100.6	93.9	100.7	100.7
n-Hexanol	2A	5.5	-0.3	67.4	84.2	101.8	98.3
2-Methyl-1-pentanol	2B	5.0	1.8	91.7	93.4	103.7	101.6
2-Ethylhexyl p-dimethyl-amino benzoate	NI	97.0	106.4	99.3	101.0	98.4	98.3
3.3-Dimethylpentane	NI	95.4	92.6	95.3	98.8	100.6	102.4
3-Methoxy-1.2-propanediol	NI	94.5	93.6	92.0	94.5	100.5	98.1
Gluconolactone	NI[1]	60.4	88.2	89.7	83.2	93.4	91.0
Glycerol	NI	106.5	95.7	100.0	99.2	100.9	100.2
Methyl amylketone	NI	45.7	91.7	96.0	101.6	102.9	101.7
Methyl cyclopentane	NI	97.6	102.2	100.1	103.3	99.4	104.9
Methyl isobutyl ketone	NI	94.1	88.5	101.8	104.9	101.4	107.3
n.n-Dimethylguanidine sulfate	NI	96.6	78.6	94.2	95.6	95.9	101.0
Pplyethyleneglycol 400	NI	93.9	92.1	94.6	97.4	100.7	85.9
Propylene glycol	NI	96.1	96.4	97.6	98.2	101.9	100.5
Toluene	NI	94.0	101.3	91.6	97.5	96.3	99.5
Tween20	NI	89.8	21.1	106.9	98.5	97.6	99.5

The viabilities should be classified into "irritant"were indicated in colored cells.
(70 % of cut-off value was used in STE test)
[1]A GHS category was classified based on the in vivo data reported by Gautheron et al., 1994

Table 2-2 Accuracy of STE test using several test substance concentration

	Concentration tested					
	10%	**5%**	**1%**	**0.5%**	**0.1%**	**0.05%**
Accuracy	91.7% (22/24)	91.7% (22/24)	75.0% (18/24)	66.7% (16/24)	66.7% (16/24)	66.7% (16/24)
False positive rate	16.7% (2/12)	8.3% (1/12)	0% (0/12)	0% (0/12)	0% (0/12)	0% (0/12)
False negative rate	0% (0/12)	8.3% (1/12)	50.0% (6/12)	66.7% (8/12)	66.7% (8/12)	66.7% (8/12)

70 % of cut-off value was used in STE test.

2.2.4.5 *Test substance exposure duration*

As described below in Section 1.2.1, the SIRC cells are exposed for 5 min in order to reflect an actual exposure situation in the STE test. On the other hands, to address the correspondence between the category classification by the STE test and GHS classification, 24 chemicals were tested by STE test at three different exposure periods from 1 min to 10 min (**Table 2-3**). It was found that the exposure period of 5 min, with cut-off value of 70%, produced best correlation to *in vivo* results for 24 chemicals (**Table 2-4**).

Table 2-3 Viability values of test substances for several exposure time (5 % test substance concentration)

Chemical	GHS	Exposure time		
		10min	5min	1min
Benzalkonium chloride	1	5.2	2.1	6.1
Cetylpyridinium bromide	1	1.5	0.6	-1.0
Cyclohexanol	1	3.2	1.4	1.4
Sodium hydroxide	1	1.5	1.4	0.0
TritonX-100	1	2.4	-0.1	2.0
1-Octanol	2A	5.0	-0.5	1.6
2-Ethyl-1-hexanol	2A	18.2	44.0	67.2
Acetone	2A	89.4	91.3	106.0
Ethanol	2A	102.2	98.2	96.4
Methyl ethyl ketone (2-butanone)	2A	18.9	44.7	93.3
n-Hexanol	2A	7.5	-0.3	3.6
2-Methyl-1-pentanol	2B	4.7	1.8	3.3
2-Ethylhexyl p-dimethyl-amino benzoate	NI	97.0	106.4	90.3
3.3-Dimethylpentane	NI	99.1	92.6	98.1
3-Methoxy-1.2-propanediol	NI	105.7	93.6	98.6
Gluconolactone	NI[1]	21.4	88.2	94.3
Glycerol	NI	102.3	95.7	99.9
Methyl amylketone	NI	97.0	91.7	94.8
Methyl cyclopentane	NI	102.1	102.2	87.6
Methyl isobutyl ketone	NI	92.3	88.5	92.4
n.n-Dimethylguanidine sulfate	NI	77.2	78.6	98.5
Pplyethyleneglycol 400	NI	101.3	92.1	110.8
Propylene glycol	NI	99.9	96.4	100.2
Toluene	NI	99.0	101.3	90.1
Tween20	NI	5.5	21.1	106.1

The viabilities should be classified into "irritant"were indicated in colored cells.
(70 % of cut-off value was used in STE test)

[1]A GHS category was classified based on the in vivo data reported by Gautheron et al., 1994

Table 2-4 Accuracy of STE test using several exposure time

	Exposure time		
	10min	**5min**	**1min**
Accuracy	83.3% (20/24)	87.5% (21/24)	87.5% (21/24)
False positive rate	16.7% (2/12)	8.3% (1/12)	0.0% (0/12)
False negative rate	16.7% (2/12)	16.7% (2/12)	25.0% (3/12)

70 % of cut-off value was used in STE test.

2.2.5 Known limits of use

As described in Section 1.2.3, while a wide range of substances with various physicochemical characteristics can be tested in the STE test, substances with insoluble either in saline, saline with 5% DMSO and mineral oil are enable to assay. Colored test substances may be problematic as they could interfere with the optical density measured in MTT assay.

Takahashi et al. (2009) noted ethanol, isopropylalcohol and sodium salicylate, were underestimated by the STE method. The false prediction of alcohols may be attributed to vaporization of the chemicals from saline (Tani et al., 1999). However the mechanism(s) for the under-prediction still remains to be established. Regarding sodium salicylate, the irritation score of the neat substance in the Draize test was high at 83.7 while that of a 10% diluted solution was 0 (Ohno et al., 1999). Based on the Draize scores for each test concentration, a category 1 classification was given to sodium salicylate by GHS. Ohno et al. (1999) indicated that the ocular irritation caused by the neat substance was largely driven by physical stimulation while chemical ocular irritation was likely insignificant since the Draize test of neat substance was conducted with particles of sodium salicylate.

Additionally, the false-positive result has been noted for Tween 20. The MMAS (Modified Maximum Average Score) value of Tween 20 (neat sample) in the Draize test was reported as 4 (ECETOC, 1998). The high molecular weight of Tween 20 (within 1128) may be the major factor for the discrepancy observed between the Draize test score and the STE test results. Due to its size, Tween 20 was not able to reach to the deep part of the cornea (wing cell and basal cell layer) and basically remain on the cornea surface of the rabbit (Wilhelm et al., 2001). On the other hand, a monolayer of SIRC cells is used in the STE test, which results in a direct contact between the chemical the cell membrane (Takahashi et al. 2009).

Moreover, in cases where the substance is shown to reduce MTT, the under-prediction might

be induced. This is caused by the direct reduction of MTT by substances, resulting that the higher cell viability was obtained than the actual. Therefore, pre-test confirmation of MTT direct reduction by substances should be done. To demonstrate whether the substance can directly reduce the MTT, 200 μL of the MTT medium and 20 μL of 5% diluent of substance in appropriate solvents are mixed and incubated for 2 hours. After incubation at 37°C, the color change to purple will be checked. If the substance, which has changed the color of MTT medium, is observed, it is able to reduce the MTT directly. In cases of these chemicals, non irritant in STE should become inconclusive because the substances might induce cytotoxicity actually.

2.2.6 Nature of the response assessed

2.2.6.1 *Cell viability*

As described below in Section 2.2.1.2, the viability of cells is conventionally measured by MTT assay method. Resulting MTT formazan is extracted with 0.04 N HCl-isopropanol, and the absorbance of the extract is measured at 570 nm with a micro plate reader (e.g., BMG Labtech Ltd., Bio-Rad, Fisher Scientific).

2.2.7 Appropriate vehicle, positive, and negative controls and the basis for their selection

2.2.7.1 *Negative Controls*

Culture medium (EMEM supplemented 10% FBS) is used as negative control in STE test. The medium control is commonly-used as negative control for the many cytotoxicity assays.

2.2.7.2 *Positive Controls*

Sodium lauryl sulfate (SLS) is used as positive control at the concentration of 0.01% in STE test. SLS is commonly-used as positive controls for cytotoxicity assays. Results from the positive control are compared to the historical control range and used to evaluate whether a particular study is acceptable.

2.2.7.3 *Vehicle Controls*

The protocol for testing requires that the test substance be dissolved or suspended in either saline, saline with 5% DMSO and mineral oil. Therefore, saline, saline with 5% DMSO and mineral oil (vehicles for STE test) are used for the vehicle control. No adverse effects of all vehicles mentioned above on the viability of SIRC cells have been confirmed.

2.2.8 Acceptable range of vehicle, positive and negative control responses and the basis for the acceptable ranges

2.2.8.1 *Negative/vehicle Controls*

Negative and vehicle controls must produce the anticipated response to ensure the test system is functioning properly and that the specific test is valid.

The Test uses the optical density (OD_{570}) of the culture medium procedural control as negative controls. An acceptable optical density values for the negative control was at least 0.3 after subtraction the mean optical density of blank wells (without cells). If the acceptable range should be not satisfied, additional assay needs to be performed.

Any of saline, saline with 5% DMSO or mineral oil is used as vehicle controls in the Test. An acceptable viability value of vehicle control is 80% or higher when the cell viability in the medium control is considered as 100%. If the acceptable range should be not satisfied, additional assay needs to be performed.

2.2.8.2 *Positive Controls*

The acceptable range is a viability that fell within two SDs of the historical mean value (21.1 – 62.3%). If the acceptable rang e should be not satisfied, additional assay needs to be performed. The historical data for negative controls is shown in the table below.

Positive Control	**Viability**
Sodium lauryl sulfate (0.01%)	
Mean (n=71)	41.7
SD	10.3
CV	24.7%
Upper and lower limits	21.1 - 62.3

2.2.8.3 *Cell Variability*

Cell variability in the STE test has been evaluated by calculating the mean ± SD for the viability values. The acceptable range of SD of the viability values with triplicate experiments performed independently for each test sample concentration is less than 15. If the SD of cell viabilities is more than 15%, another three independent assays are performed for that concentration and the SD is re-calculated from the final viabilities derived from newly performed three assays.

2.2.9 Nature of the data to be collected and the methods used for data collection

As described below in Section 2.2.1.2, the viabilities of cells are conventionally measured by MTT assay method. Resulting MTT formazan is dissolved with 0.04 N HCl-isopropanol, and the absorbance of the formazan solution is measured at 570 nm with a micro plate reader. Raw data are typically recorded electronically.

2.2.10 Type of media in which data are stored

The data from STE test be stored and archived in an appropriate manner as an electronic file and printed matter.

2.2.11 Measures of cell variability

As described below in Section 2.2.8.3, the cell variability in the STE test is evaluated by calculating the mean ± SD for the viability values. This value allow for an assessment of the performance of the test conducted and whether the observed variability between replicates is greater than would be considered acceptable.

2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data

The STE test uses the mean OD_{570} values for each test sample to calculate a viability of SIRC cells. The ratio of OD_{570} values on each test sample to that of control is represented as relative viability (triplicate determinations). The control group cells are exposed to vehicle (physiological saline, saline with 5% DMSO, or mineral oil). The mean of three wells for each test concentration is calculated. This is the mean relative viability for one independent test. A total of three independent tests are conducted for each concentration of a test material, and the calculated overall mean of three independent tests is used for estimation of ocular irritation.

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

As described below in Section 2.1, 70% is applied as cut-off value that be used to classify the ocular irritation potential of irritant and non irritant in STE test. To address the correspondence between the category classification by the STE test and GHS classification, 24 chemicals were tested by STE test at different cut-off values from 10% to 90%. It was found that the cut-off value of 70%, with the test concentration of 5%, produced better correlation to *in vivo* results for 24 chemicals (**Table 2-5**). Based on these results, the cut-off value of 70% was applied for STE test.

Table 2-5 Accuracy of STE test using different test substance concentration

	Cut-off value					
	80%	**70%**	**60%**	**50%**	**40%**	**30%**
Accuracy	87.5% (21/24)	91.7% (22/24)	91.7% (22/24)	91.7% (22/24)	83.3% (20/24)	83.3% (20/24)
False positive rate	16.7% (2/12)	8.3% (1/12)	8.3% (1/12)	8.3% (1/12)	8.3% (1/12)	8.3% (1/12)
False negative rate	8.3% (1/12)	8.3% (1/12)	8.3% (1/12)	8.3% (1/12)	25.0% (3/12)	25.0% (3/12)

5% of test substance concentration was used in STE test.

Category classification of ocular irritation by STE test was determined based on the relative viability assessed for 5% test concentrations. A concentration of test material that had a relative viability of 70% or less was categorized as an irritant (I) and a concentration of test material that had a relative viability greater than 70% was categorized as a non-irritant (NI).

2.3 Basis for Selection of the Test System

As discussed in Section 1.1.1, this test method is based on the cytotoxicity of cornea cells because the cornea cells is one of the main targets during accidental eye exposures, and damage to the cornea can result in visual impairment or loss. In addition, corneal effects are weighted heavily in the original *in vivo* ocular irritancy scoring systems (e.g., 80 out of a possible 110 points in the Draize eye test scoring system).

In the STE test, cytotoxicity is determined by the viability of the rabbit corneal cell line (SIRC cells). The viability of cells is conventionally measured by MTT assay method. While these *in vitro* toxicity measurements using the isolated cornea are correlated with *in vivo* ocular irritation corneal effects, they represent only one aspect of the overall complex response of the eye to irritants, which involves other tissues such as the iris and conjunctiva.

For the ocular irritation animal alternative test, it may be unlikely to completely replace the Draize test by a single *in vitro* test because the Draize test evaluates a range of criteria for injury and inflammation to the eye. The tiered approach of several *in vitro* assays combined was proposed in order to estimate the irritation potential for a wide range of chemical classes (Hagino et al., 2008; McNamee et al., 2009; Scott et al., 2010).

The STE test showed a highly predictive capacity and applicable domain were also observed for the STE test because of the correlation both estimated ocular irritation category classification by the STE test have with those classifications of the GHS for the different classes of chemicals involved like the poorly water-soluble chemicals (Takahashi et al., 2009, Sakaguchi et al., 2011, Kojima et al., 2012). For these reasons, the STE test could be considered a building block assay in the tiered approach to establishing an ocular irritation animal alternative testing paradigm.

2.4 Basis for Number of Replicate and Repeat Experiments

2.4.1 Sample Replicates

For each sample concentration, three wells per experiment are used and the mean value is calculated to obtain cell viability. Based on scientific judgment, it would seem reasonable to average the viabilities in triplicate determinations for reducing the variation of viability caused by the seeding uniformity of cells.

2.4.2 Experimental Replicates

The mean value of three independent experiments is used as the final cell viability for each sample concentration. This mean value of cell viability is defined as the mean cell viability. Although the basis of them is not defined, however, based on scientific judgment, it would seem reasonable to predict that equivocal or unexpected results obtained among tests would mandate repeating the experiment.

3.0 Substance Used for Validation of STE test

3.1 Rationale for the Chemicals Selected for Use

A total 96 substances were evaluated in the four validation studies (containing two pre-validation studies). In addition, another 23 substances were evaluated in one laboratory. These 23 substances have only one STE data (16 substances were referred from Takahashi et al., 2011, and 7 substances were in-house data). These data of 23 substances were added to the dataset to enhance the reliability of predictive performance. Section 3.1.1 through 3.1.5 address the rationale for the chemicals tested in each of these studies.

3.1.1 Takahashi et al. (2009) - *pre-validation study 1* -

The rationale for the chemicals selected for use in the study is not known.

3.1.2 Takahashi et al. (2010) - *pre-validation study 2* -

The rationale for the chemicals selected for use in the study is not known.

3.1.3 Sakaguchi et al. (2011) – *Phase I validation study* -

Most substances had already being used in for the prevalidating of other alternative eye irritation test methods, for which results have been reported (Van Goethem et al., 2006). In addition, the other chemicals were selected for which data have been published by the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC, 1998), Gautheron et al. (1992), or Ohno et al. (1999). In terms of the 25 chemicals selected, the corresponding categories were as follows: 5 chemicals were Category 1; 7 chemicals were Category 2; and 13 chemicals were non-irritants (NI). The test chemicals covered the whole range of eye irritation potencies and represented different chemical classes.

3.1.4 Kojima et al. (2012) – *Phase II validation study* -

It is preferable to examine more chemicals to estimate the predictivity of STE test. Therefore, the additional 40 chemicals (for the estimation of reproducibility, containing 2 chemicals that were commonly-selected in former validation study [Sakaguchi et al. 2011]) were selected for this validation study to make up for the lack of chemicals selected for former validation study (Sakaguchi et al., 2011). Consequently, total 63 chemicals with the balanced GHS classification, chemical class, physical state and vehicle used were selected for the both validation studies.

3.1.5 In-house data

Twenty-three additional substances from Takahashi et al., 2011 and unpublished data were added to dataset to enhance the reliability of investigation of predictive performance. Twenty-two out of 23 substances were not judged as irritant in GHS.

3.2 Rationale for the Number of Substance Tested

The rationale for the number of chemicals selected in four validation/pre-validation studies is not known. The number of the Additional in-house data was determined based on the balance between the irritant and non irritant for GHS and EPA classification.

3.3 Chemicals Evaluated

Descriptive information for each of the substances tested in the STE test was obtained, to the extent possible, from the information provided in the study reports. No attempt was made to identify the purity of a substance if the information was not included in the study report. However, if chemical classes were not assigned in the study reports, the information sought from other sources, including the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at http//www.nlm.nih.gov/mesh). **Appendix B** provides the available information on the name, CASRN, chemical/product class, pH, vapor pressure of each substance evaluated in the STE test. In addition, Appendix B also provides its solubility, the skin irritating potency, and the potency of direct reduction of MTT. **Tables 3-1** provide the chemical classes of the test substances evaluated with the STE test. The detailed information is provided **Appendix B** with the STE test data.

Table 3-1 Chemical Classes Tested in the STE test

Chemical Class	No. of Substances
Alcohols	22
Aldehydes	3
Alkalis	1
Amidines	1
Amines	6
Carboxylic acids	5
Color additives	1
Disulphides	1
Esters	14
Ethers	1
Fatty acids	1
Hydrocarbons	19
Inorganic chemicals	1

Chemical Class	No. of Substances
Inorganic salts	1
Ketones/Lactones	8
Organic salts	5
PABA derivatives	1
Paraffins	1
Propane derivatives	1
Silicon compounds	1
Sulphoxides	1
Surfactants (anionic)	5
Surfactants (cationic)	7
Surfactants (nonionic)	11
Not categorized	1

As shown in **Table 3-1**, the chemical classes with the greatest amount of the STE test data are alcohols, hydrocarbons, esters and surfactants (nonionic). Other chemical classes tested include ketones/lactones, surfactants (cationic), amines, organic salts, carboxylic acids and surfactants (anionic).

3.3.1 Takahashi et al. (2009) *- pre-validation study -*

Regarding descriptive information about test substances, the forty-four chemicals consisting of 12 surfactants (2 anionic, 6 cationic, 4 nonionic), 12 alcohols, 4 amines, 3 ketones, 3 carboxylic acid, 3 hydrocarbons, 2 esters, 1 alkali, 1 organic salt, 1 inorganic chemical, 1 co lour additive, 1 PABA derivative were selected. Sodium lauryl sulfate (SLS), calcium thioglycolate and Tween 80 were selected as standard chemicals to confirm transferability.

3.3.2 Takahashi et al. (2010) *- pre-validation study -*

Seventy raw materials included 22 surfactants (10 nonionic activators, 5 anionic activators, 7 cationic activators), 16 alcohols, 5 esters, 4 amines, 4 ketones, 3 carboxylic acids, 3 hydrocarbons, 3 organic salts, 1 alkali, 1 color additive, 1 fatty acid, 1 inorganic chemical, 1 PABA derivative, 1 paraffin,1 propane derivative, 1 silicone compound, 1 sulphoxide, and 1 not categorized chemical.

3.3.3 Sakaguchi et al. (2011) *– Phase I validation study -*

The 25 test substances included 9 alcohols, 3 hydrocarbons, 5 ketones/lactones, 4 surfactants, 1 alkali, 1 amidine, 1 ester, and 1 PABA derivative were selected. The test

substances used were encoded and participating laboratories were blinded to the identification of the test substances.

3.3.4 Kojima et al. (2012) – *Phase II validation study -*

The selected 38 test substances included 8 alcohols, 6 esters, 4 hydrocarbons, 4 organic salts, 4 surfactants, 3 aldehydes, 3 amines, 2 carboxylic acids, 2 ketones/lactones, 1 inorganic salt, and 1 ether were selected. The chemicals were selected to make up for the lack of chemicals selected for former validation (Sakaguchi et al. 2011). The test substances used were encoded and participating laboratories were blinded to the identification of the test substances.

3.3.5 In-house data

The selected 23 test substances included 12 hydrocarbons, 2 alcohols, 6 esters, 1 amine, 1 ketone/lactone and 1 disulphide were selected.

3.4 Coding Procedures Used in the Validation Studies

The coded chemicals were only used for the two validation study (Sakaguchi et al., 2011, Kojima et al., 2012). The test substances used for these two validation studies were encoded and participating laboratories were blinded to the identification of the test substances. Coding procedures used in these studies is not known.

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

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

4.1.1 Draize Rabbit Eye Test

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

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

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

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

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

4.2 Detailed Reference Data Used for BRD Analysis

The STE test results evaluated in this document include *in vivo* reference data generated using the basic procedures described above for the *in vivo* rabbit eye test method.

Most *in vivo* reference data were referred from ECETOC reference chemicals data bank (ECETOC 1998). Three to six rabbits were used per test substance and MAS (Draize et al. 1944) were calculated. Sufficient *in vivo* data were provided for 64 of these substances to be classified according to the GHS (UN 2003) and the EPA (EPA 1996) ocular irritancy classification systems (**Appendix B2**). Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal were available from ECETOC reference chemicals data bank (ECETOC 1998).

Regarding another 45 substances, sufficient *in vivo* data referred from the publicly-available documents (such as ICCVAM, 2006) or unpublished data (in-house data or data from ICCVAM referred from NIHS Japan, TSCA, ZEBET, and NLDLS in Appendix B2) were provided to be classified according to the EPA, and the GHS ocular irritancy classification. (**Appendix B2**). 10 substances have not sufficient *in vivo* data.

4.3 In Vivo Classification Criteria Used for BRD Analysis

4.3.1 GHS classification Rules Used for BRD Analysis

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

Table 4-2 Criteria for Classification of Rabbits According to the GHS Classification System

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

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

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

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

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

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

4.3.2 EPA classification Rules Used for BRD Analysis

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

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

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

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

4.4 Availability of Original Records for the In Vivo Reference Data

Although the original study records were not obtained from cited literatures, the detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal were available from ECETOC reference chemicals data bank (ECETOC 1998). For other references including the publicly-available documents (such as ICCVAM, 2006) or unpublished data (in-house data or data from ICCVAM referred from NIHS Japan, TSCA, ZEBET, and NLDLS in **Appendix B2**), similar detailed *in vivo* data were available.

4.5 In Vivo Data Quality

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

Based on the available information, the Draize data listed in ECETOC reference chemicals data bank (ECETOC 1998) were obtained according to GLP guidelines. For the remaining reports (ICCVAM, 2006, in-house data or data provided by ICCVAM referring to NIHS Japan, TSCA, ZEBET, and NLDLS in **Appendix B2**), the extent of GLP compliance was not provided, so the extent of GLP compliance is not known.

4.6 Information About Accuracy and Reliability of the In Vivo Test Method

4.6.1 Information About the Accuracy of the *In Vivo* Test

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

In comparison, there have been studies where the responses to ocular irritants differ between humans and rabbits. In some cases, test substances produced more severe responses in humans than in rabbits (Lewin and Guillery 1913; Gartner 1944; Estable 1948; Marsh and Maurice 1971; Grant 1974). For example, Marsh and Maurice (1971) evaluated

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

4.6.2 Information About the Reliability of the *In Vivo* Test

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

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

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

5.0 STE Test Data and Results

5.1 Description of the STE Test Protocols Used to Generate Data

STE test was conducted based on the method of Takahashi et al. (2008). As mentioned in preface, rabbit cornea-derived SIRC cells are exposed to a substance evaluated at a constant concentration for 5 min. Resulting viability is determined by the incorporation of methylthiazolydiphenyl-tetrazolium bromide (MTT: tetrazolium salt substance). In the STE test, a 5% concentration of test sample is used for evaluating the irritation potential. The irritation category and score can differ depending on whether cell viability is greater than 70%. Physiological saline is used as a test vehicle to evaluate water-soluble substances, while physiological saline containing 5% DMSO or mineral oil is used for water-insoluble substance. Details of protocol were described in the **appendix A**. In all existing publications, any modification and difference not exist.

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

All data including copies of original data to evaluate the accuracy and reliability are available upon request.

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

As described below in Section 2.2.12, the STE test used the mean OD_{570} values for each test sample to calculate a viability of SIRC cells. The ratio of OD_{570} values on each test sample to that of control was represented as relative viability (triplicate determinations). The control group cells were exposed to vehicle (physiological saline, saline with 5% DMSO, or mineral oil). The mean viability of three wells for each test concentration was calculated. This was the mean relative viability for one independent test. A total of three independent tests were conducted for each concentration of a test material, and the calculated overall mean of three independent tests was used for estimation of ocular irritation.

As described below in Section 2.2.13, Category classification of ocular irritation by STE test was determined based on the relative viability assessed for 5% test concentrations. A concentration of test material that had a relative viability of 70% or less was categorized as

an irritant (I) and a concentration of test material that had a relative viability greater than 70% was categorized as a non-irritant (NI).

5.4 Summary of Results

STE data were collected for total 119 test substances including 96 test substances among the four validation/pre-validation studies evaluated. A summary of results used to evaluate test accuracy based on the ocular irritation category (Non Irritant or Irritant) is shown in **Appendix D**. **Appendix D1** provides a table, sorted by alphabetically by substance with the name of the substance tasted, the CASRN, the concentration tested, the STE data (mean viability value, standard deviation, number of replicates), category classification of the test substance, and the reference. **Appendix D2** provides the same information, but is sorted alphabetically by test substance to indicate which substances were tested in multiple studies. Other supporting information, such as the source, purity and physicochemical characteristics of the test substances, was included in the tables to the extent this information was available. No attempt was made to identify the source, purity, and physicochemical characteristics of a test substance, if the authors did not provide such information. Chemical and product classes were assigned based on the MeSH classification system (available at http//www.nlm.nih.gov/mesh). Each of the eight studies evaluated varied with respect to the level of detail of data that was provided, as described below.

5.4.1 Takahashi et al. (2009) - *pre-validation study 1* -

In this inter-laboratory study with 3 laboratories, 44 chemicals with a wide range of classes were evaluated for the transferability, between-lab reproducibility and predictive capacity of the STE test as an alternative ocular irritation test. GHS classification based on Draize eye irritation test data was used as the comparative *in vivo* data. Transferability was assessed using standard chemicals (sodium lauryl sulfate, calcium thioglycolate, and Tween 80) and the coefficient variations (CVs) of relative viabilities between 3 labs were less than 0.13. The irritation category (Irritant or Non irritant) at a 5% test concentration in STE test was the same in 3 laboratories for all 44 tested chemicals. The predictive capacity irritation category classification between STE test results and GHS were compared, and a good correlation was confirmed (accuracy was 90.9% at all laboratories). In addition, the STE rankings of 1, 2, and 3 classified by the prediction model (PM) based on the relative viability at two concentrations (5% and 0.05%) were highly correlated with the GHS ranks of non-irritant, category 1, and category 2, respectively (accuracy was 75.0% at all laboratories).

5.4.2 Takahashi et al. (2010) *- pre-validation study 2 -*

In this inter-laboratory study, 2 laboratories conducted the test using 70 raw materials in order to evaluate transferability, between-lab reproducibility, and predictive capacity of STE test as an alternative ocular irritation test. Transferability was assessed using saline as a negative control and 0.01% sodium lauryl sulfate as a positive control. Relative viabilities obtained for both laboratories were almost the same. Therefore, transferability was considered to be excellent. Both laboratories showed similar relative viabilities for all 70 raw materials at each test concentration. Correspondence rates of ocular irritation category (Irritants or Non irritants) were over 97% for each concentration tested, exhibiting high between-laboratory reproducibility. The correspondence rate for predicting ocular irritation potential of neat raw materials and a diluted solution (10%) were over 85% at each laboratory for the 5% and 0.05% test concentration in STE. Lastly, the correspondence rate for the rank classification by the prediction model at each laboratory was over 72%, and the correspondence became almost 90% when acids, amines, and alcohols were excluded from the analysis.

5.4.3 Sakaguchi et al. (2011) *- Phase I validation study -*

The Japanese Society for Alternative to Animal Experiments organized an Executive Committee and conducted the validation study at four laboratories to evaluate the transferability, inter-laboratory reproducibility, and predictive ability of the STE method. The mean cell viability determined for three standard substances (sodium lauryl sulfate [SLS], calcium thioglycolate [CT], and Tween 80 [TW80]) were equivalent. Furthermore, the rank classification of the three standard substances derived at each of the four laboratories was 3, 2, and 1 for SLS, CT, and TW80, respectively. In the evaluation of 25 blinded test substances, the correspondence between the STE test result (5% data) and GHS category (non-irritant vs. irritant) was good, and nearly the same results were obtained by each of the laboratories. Further, in terms of correspondence between the ranks obtained from the mean cell viability at 5% and 0.05% in the STE test and the GHS irritation classification (three criteria), good results were confirmed and almost the same results were obtained by all the laboratories. The standard deviations of mean cell viability for the tested chemicals in their respective vehicles (physiological saline for water soluble chemicals and physiological saline with 5% (w/w) DMSO or mineral oil for water insoluble chemicals) were small and inter-laboratory reproducibility was also good.

5.4.4 Kojima et al. (2012) *- Phase II validation study -*

In this second-phase validation study, a new VMT was organized by JaCVAM (Japanese Center for the Validation of Alternative Methods); we re-evaluated the predictive capacity of the STE test using an additional 40 blinded substances in three laboratories. After that, we evaluated the predictive ability of GHS category in the STE test, using 63 blinded substances along with the results from the first-phase validation study. The results showed that the STE test was not only easy to acquire and implement among three laboratories, it also had a high intra- and inter-laboratory reproducibility. Furthermore, the STE classification was highly effective in predicting the GHS classification of various substances. However, a predictive ability for predicting the STE rank was not good compared with that of GHS categories. Therefore, the STE test can assess not only severe/corrosive ocular irritants (corresponding to UN GHS Category 1) but also mild or moderate ocular irritants (corresponding to UN GHS Category 2). The predictive ability for predicting the STE rank was insufficient for identification of UN GHS categories (Category 1, Category 2, and No Category). From these results, we recommend the STE test as an initial step within a Bottom-Up approach to identifying substances that do not require classification as eye irritants (UN GHS No Category), as well as a step within a Top-Down approach to identify severe, moderate or mild irritants, and substances that do not require classification as eye irritants (UN GHS No Category) from other toxicity classes, specifically for limited types of substances. On the other hand, we do not consider the STE test adequate or valid for the identification of mild or moderate irritants (i.e., UN GHS Categories 2A and 2B) or severe irritants (UN GHS Category 1).

5.4.5 In-house data

Additional data of 23 substances are not assessed independently because these data are only used for the analysis of overall predictive performance by using pooled data.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

The coded chemicals were only used for the phase I and phase II validation study (Sakaguchi et al., 2011, Kojima et al., 2012). Ideally, all data supporting the validity of a test should be obtained and reported in accordance with GLP guidelines and with the use of coded chemicals (OECD 1998; EPA 2003a, 2003b; FDA 2003). No report that identified following GLP guidelines or used data obtained according to GLP guidelines existed.

However, the two validation studies mentioned above (Sakaguchi et al. 2011 and Kojima et al.

2012) were performed in the spirit of GLP compliance. To guarantee data quality, the following considerations were applied. To start the experiments, several recording sheets were prepared for recording the necessary items. Information regarding adjustment/operation confirmation of instruments, culture media (EMEM supplemented with 10% FBS) preparation, vehicle selection for the test substances, use of reagent/test substance, preparation of test sample, preparation of 0.04 N HCl-isopropanol, passage of SIRC cells, and plate seeding was recorded on the recording sheets under the authority of the experiment personnel and the person responsible for the experiment. These records were stored at each laboratory. To ensure measured values were appropriately recorded in the data sheet prepared for this study, a data audit was performed, and consistency between the values recorded in the printout from the plate reader and values inputted into the data sheet was confirmed.

5.6 Lot-to-lot Consistency of Test Substances

There was no information about the lot-to-lot consistency in any reports.

5.7 Availability of Data for External Audit

All study notebooks and other supporting records are available, upon request, for an external audit, for the following studies: Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011) and Kojima et al. (2012).

6.0 STE test Accuracy

6.1 Accuracy of the STE test

The ability of the STE test to correctly identify ocular irritant, as defined by the GHS (UN 2003) and the EPA (1996), was evaluated using two approaches. In the first approach, the performance of the STE test was assessed separately for each in vitro-*in vivo* comparative study (i.e., publication or data submission) reviewed in Sections 4.0 and 5.0. In the second approach, the performance of the STE test was assessed after pooling data across comparative studies that used the same method of data collection and analysis. When there was any test substances that had no *in vivo* data (e.g., individual irritation score of cornea, iris, and conjunctiva at observation period up to 21 days) and could not determinate the classifications of GHS or EPA, these substances were excluded out of the accuracy analysis.

Accuracy of STE for Individual Studies:

The ocular irritation category of each test substance in each study was summarized in **Appendix C**. All studies reviewed in this BRD where the same test substance was evaluated in multiple laboratories within the same study, an overall ocular irritation category classification was assigned for each chemical in the study based on the majority of classification calls (e.g., if two tests classified a substance as a non irritant and three tests classified a substance as a irritant, the overall ocular irritation category classification for the substance would be irritant). When there was an even number of different ocular irritation category for test substances (e.g., two tests classified a substance as a non irritant and two tests classified a substance as an irritant), these substances were assigned as irritant. Once the ocular irritation category classification was determined for each substance in each of the studies, the ability of the STE test to identify non irritant and irritant, based on the two different classification systems, was determined for each study (**Appendix D**).

Accuracy of STE for Pooled Studies:

For overall analysis of accuracy for the STE test, results from the four different comparative studies and additional 23 data obtained in one laboratory were combined and an overall classification was determined for each substances. The classification was compared to the regulatory classification of ocular irritation (**Appendix D**).

Among 119 substances, 2-ethylhexyl thioglycolate had an ability to induce direct reduction of MTT. Because it was assigned as “non irritant” in STE test, it was finally assigned as “inconclusive”. Therefore 2-ethylhexyl thioglycolate was excluded out from the overall

analysis of accuracy for STE test. The overall analysis was conducted against 99 or 98 substances for GHS or EPA classification, respectively.

6.1.1 GHS Classification System: STE test Accuracy

Accuracy analyses for irritant (category 1, category 2A or category 2B) and non irritant (not classified), as defined by the GHS classification system (UN 2003), were performed for the following four studies: Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), and Kojima et al. (2012). The GHS classification assigned to each test substance is presented in **Appendix D**. The performance characteristics (i.e., accuracy, sensitivity, specificity, positive predictivity, negative predictivity, false positive rate, and false negative rate) were determined for each of the four studies based on the available *in vivo* reference data for the substances tested in these studies (**Table 6-1**). All studies, Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), and Kojima et al. (2012) provided STE data for substances tested in multiple laboratories; the accuracy calculations for these studies in **Table 6-1** represents the results obtained using the consensus call for each test substance was considered.

Based on the data provided in the four studies, the STE test has an accuracy of 74% to 88%, a sensitivity of 77% to 86%, a specificity of 60% to 100%, a false positive rate of 0% to 40%, and a false negative rate of 14% to 23% (**Table 6-1**).

As described below in Section 3.1.4, in a study reported by Kojima et al. (2012) the 40 chemicals (for the estimation of reproducibility, containing 2 chemicals that were commonly-selected in former validation study [Sakaguchi et al. 2011]) were selected for this validation study to make up for the lack of chemicals selected for former validation study (Sakaguchi et al. 2011). As a result, larger number of irritant substance (GHS category 1 and category 2) and only four non irritant substances were selected. For these reasons, especially low specificity and high false positive rate was confirmed in this study. However, the reasonable performance characteristics for the combined results of two studies (Sakaguchi et al., 2011 and Kojima et al., 2012) were confirmed.

In terms of an overall accuracy analysis, using all data, the STE test has an accuracy of 85%, a sensitivity of 84%, a specificity of 86%, a false positive rate of 14%, and a false negative rate of 16%. The performance characteristics for the pooled studies are provided in **Table 6-1**.

As described in Sections 3.0 and 4.0, appropriate *in vivo* data were not available for all of the substances evaluated in some of the studies. For example, in the Takahashi et al. (2010) study, only 55 of the 70 substances had appropriate *in vivo* data to assign a GHS classification.

Table 6-1 Evaluation of the Performance of the STE test In Predicting Ocular Irritant Compared to In Vivo findings, as Defined by the GHS Classification System, by Study and Overall

Data Source	N[1]	Accuracy		Sensitivity		Specificity		Positive predictivity		Negative predictivity		False Positive Rate		False Negative Rate	
		%	No.[2]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Takahashi et al. 2009	40/44	87.5[3]	35/40	85.7	24/28	91.7	11/12	96.0	24/25	73.3	11/15	8.3	1/12	14.3	4/28
Takahashi et al. 2010	55/70	83.6	46/55	85.7	30/35	80.0	16/20	88.2	30/34	76.2	16/21	20.0	4/20	14.3	5/35
Sakaguchi et al. 2011 (A)	24/25	87.5	21/24	76.9	10/13	100.0	11/11	100.0	10/10	78.6	11/14	0.0	0/11	23.1	3/13
Kojima et al. 2012 (B)	35/40	74.3	26/35	76.7	23/30	60.0	3/5	92.0	23/25	30.0	3/10	40.0	2/5	23.3	7/30
(A) + (B)	57/63	80.7	46/57	78.0	21/41	87.5	14/16	94.1	32/34	60.9	14/23	12.5	2/16	22.0	9/41
Pooled Studies[4]	99/119	84.8	84/99	83.6	46/55	86.4	38/44	88.5	46/52	80.9	38/47	13.6	6/44	16.4	9/55

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

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

[3]Performance calculated using the overall classification based on the majority classification among the multiple testing laboratories.

[4]Data from Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), Kojima et al. (2012), and 23 additional data were pooled together and an overall classification was assigned for each test substance based on the majority classification obtained in each testing laboratory.

2-Ethylhexyl thioglycolate was excluded out from the overoall analysis due to its ability of direct reduction of MTT.

6.1.1.1 *Discordant Results According to the GHS Classification System*

In order to evaluate discordant responses of the STE test relative to the regulatory classification of ocular irritation, several accuracy sub-analyses were performed. These included specific classes of chemicals with robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, physical form).

As indicated in **Table 6-2**, there were some notable trends in the performance of the STE test among these subgroups of substances. Although the number of substance was too few to resolve any definitive trends in overprediction by STE test, organic/inorganic salts was most often overpredicted. Although there were a relatively small number (13 or 7) of substances represented, esters and ketones/lactones were also often overpredicted (i.e., were false positives) by the STE test according to the GHS classification system (see **Appendix D**).

With regard to physical form of the substances overpredicted by the STE test, one was solids and another 5 were liquids. Despite the proportion of the total available data, solids (23/97; 24%) appear more likely than liquids (74/97; 76%) to be overpredicted by the STE test.

Although there were a relatively small number (5) of substances represented, organic/inorganic salts were most often underpredicted (i.e., were false negatives) by the STE test according to the GHS classification system. Additional chemical classes represented among the underpredicted substances were hydrocarbons, alcohols, and esters. Among the 17 substances labeled as surfactants, no was underpredicted by the STE test.

With regard to physical form of the substances underpredicted by the STE test, four were liquids and five were solids. Despite the proportion of the total available data, solids (23/97; 24%) appear more likely than liquids (74/97; 76%) to be underpredicted by the STE test.

Among the underpredicted substances for which pH information was available, as three was acidic ($pH \leq 7.0$) and zero was basic ($pH > 7.0$), therefore, the acidic substances (26/35; 74%) may have a tendency to underestimate. Although the numbers of substance among the overpredicted substances were few to resolve any definitive trends in overprediction by STE test, there was two false positive substance (Polyethyleneglycol monolaurate (10E.O.) and tween 20) of acidic (see **Appendix D**).

With regard to volatility of the substances underpredicted by the STE test, substance with the vapor pressure between 10 kPa and 1 kPa were most often underpredicted by the STE test according to the GHS classification system. Additional vapor pressure ranges represented among the underpredicted substances were between 1 kPa and 0.1 kPa, and under 0.01 kPa. The numbers of substance among the overpredicted substances were few to resolve any definitive trends in overprediction by STE test. However, the substance with the vapor pressure between over 10 kPa, between 1 kPa and 0.1 kPa, and between 0.1 kPa and 0.01 kPa may have a tendency to overestimate.

Table 6-3 shows the effects on the STE test performance statistics of excluding from the data set problematic classes (i.e., that gave the most discordant results, according to the GHS classification system).

In general, exclusion of organic/inorganic salts, esters or alcohols individually resulted in small changes in the performance statistics. When both salts and alcohols were excluded from the data set, changes performance statistics were noted, with accuracy increasing from 85 (84/99) to 89% (64/72), and the false negative rate decreasing from 16% (9/55) to 9% (3/35). The largest changes were observed when all three discordant classes were excluded from the data set; accuracy increased from 85 (84/99) to 90% (54/60), and the false negative rate decreased from 16% (9/55) to 7% (2/31).

With regard to false positive rate, the exclusion of the combination of organic/inorganic salts and alcohols from the data set caused no substantial change with the exception that the exclusion of the combination with organic/inorganic salts and esters of resulted in slightly improve in the performance statistics.

Table 6-4 shows the effects on the STE test performance statistics of excluding from the data set problematic substances with volatility.

When the substance with vapor pressure over 6kPa was excluded from the data set, changes performance statistics were noted. The largest changes were observed when the substances with vapor pressure over 3kPa were excluded from the data set; accuracy increased from 86% (71/83) to 89% (59/66), and the false negative rate decreased from 20% (9/45) to 14% (5/37). The similar larger changes were observed when the substances with vapor pressure over 6 kPa were excluded; accuracy increased from 86% to 88% (61/69), and the

false negative rate decreased from 20% to 16% (6/39).

Table 6-2 False Positive and False Negative Rates of the STE test, by Chemical Class and Properties of Interest, for the GHS classification

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall	99	13.6	6/44	16.4	9/55
Chemical Class[5]					
Alcohols	22	0.0	0/6	25.0	4/16
Amines	4	0.0	0/1	0.0	0/3
Carboxylic acids	3	-	0/0	0.0	0/3
Esters	13	22.2	2/9	25.0	1/4
Hydrocarbons	19	0.0	0/15	50.0	2/4
Ketones/Lactones	7	25.0	1/4	0.0	0/2
Organic/Inorganic salts	5	100.0	1/1	50.0	2/4
Properties Interest					
Liquid[6]	74	12.2	5/41	12.1	4/33
Solid[6]	23	33.3	1/3	25.0	5/20
Surfactants - Total	17	50.0	2/4	0.0	0/13
- nonionic	7	50.0	2/4	0.0	0/3
- anionic	3	-	0/0	0.0	0/3
- cationic	7	-	0/0	0.0	0/7
pH - Total[7]	35	22.2	2/9	11.5	3/26
- acidic (pH ≤ 7.0)	26	28.6	2/7	15.8	3/19
- basic (pH > 7.0)	8	0.0	0/1	0.0	0/7
Volatility - Total[8]	83	7.9	3/38	20.0	9/45
- VP[9] **≧ 10 kPa**	10	16.7	1/6	0.0	0/4
- 10 kPa > VP ≧ 1 kPa	14	0.0	0/6	50.0	4/8
- 1 kPa > VP ≧ 0.1 kPa	18	9.1	1/11	14.3	1/7
- 0.1 kPa > VP ≧ 0.01 kPa	16	12.5	1/8	0.0	0/8
- VP < 0.01 kPa	25	0.0	0/7	22.2	4/18

[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 STE test and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

[6]Physical form is unable to be classified for some substances that have proparties of both solid and liquid, and therefore the overall number does not equal the sum of the solid and liquid substances.

[7]Total number of substances for which pH information was obtained. One substance was not categorized due to limited information.

[8]Total number of substances for which volatility information was obtained.

[9]VP = Vapor pressure

Table 6-3 Effect of Exclusion of Discordant Classes on False Negative and False Positive Rates of the STE test, for the GHS Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	84.8	84/99	13.6	6/44	16.4	9/55
w/o Salts (Organic/Inorganic)	87.2	82/94	11.6	5/43	13.7	7/51
w/o Esters	86.0	74/86	11.4	4/35	15.7	8/51
w/o Alcohols	85.7	66/77	15.8	6/38	12.8	5/39
w/o Salts & Esters	87.8	72/82	11.4	4/35	12.8	6/47
w/o Salts & Alcohols	88.9	64/72	13.5	5/37	8.6	3/35
w/o Esters & Alcohols	87.5	56/64	13.8	4/29	11.4	4/35
w/o Salts & Esters & Alcohols	90.0	54/60	13.8	4/29	6.5	2/31

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.

Table 6-4 Effect of Exclusion of Discordant Volatile Chemicals on False Negative and False Positive Rates of the STE test, for the GHS Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall with vapor pressure (VP) data	85.5	71/83	7.9	3/38	20.0	9/45
w/o VP > 15 kPa	84.4	65/77	8.8	3/34	20.9	9/43
w/o VP > 10 kPa	84.9	62/73	6.3	2/32	22.0	9/41
w/o VP > 8 kPa	85.9	61/71	8.6	2/31	20.0	8/40
w/o VP> 7kPa	87.1	61/70	8.6	2/31	17.9	7/39
w/o VP > 6kPa	88.4	61/69	8.6	2/31	15.8	6/39
w/o VP > 4kPa	88.0	59/67	6.9	2/29	15.8	6/38
w/o VP > 3 kPa	89.4	59/66	6.9	2/29	13.5	5/37
w/o VP > 2kPa	88.7	55/62	7.1	2/27	14.3	5/35

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.
VP, vapor pressure

6.1.2 EPA Classification System: STE test Accuracy

Accuracy analyses for irritant (category I to III) and non irritant (category IV), as defined by the EPA classification system (EPA 1996), were performed for the following four studies: Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), and Kojima et al.

(2012). The EPA classification assigned to each test substance is presented in **Appendix D**. The performance characteristics (i.e., accuracy, sensitivity, specificity, positive predictivity, negative predictivity, false positive rate, and false negative rate) were determined for each of the four studies based on the available *in vivo* reference data for the substances tested in these studies (**Table 6-5**). All studies, Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), and Kojima et al. (2012) provided STE data for substances tested in multiple laboratories; the accuracy calculations for these studies in **Table 6-5** represents the results obtained using the consensus call for each test substance was considered.

Based on the data provided in the four studies, the STE test has an accuracy of 65% to 80%, a sensitivity of 53% to 78%, a specificity of 90% to 100%, a false positive rate of 0% to 10%, and a false negative rate of 22% to 47% (**Table 6-5**).

In terms of an overall accuracy analysis, using all data, the STE test has an accuracy of 74%, a sensitivity of 68%, a specificity of 96%, a false positive rate of 4%, and a false negative rate of 32%. The performance characteristics for the pooled studies are provided in **Table 6-5**.

As described in Sections 3.0 and 4.0, appropriate *in vivo* data were not available for all of the substances evaluated in some of the studies. For example, in the Takahashi et al. (2010) study, only 54 of the 70 substances had appropriate *in vivo* data to assign an EPA classification.

Table 6-5 Evaluation of the Performance of the STE test In Predicting Ocular Irritant Compared to In Vivo findings, as Defined by the EPA Classification System, by Study and Overall

Data Source	N[1]	Accuracy		Sensitivity		Specificity		Positive predictivity		Negative predictivity		False Positive Rate		False Negative Rate	
		%	No.[2]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Takahashi et al. 2009	39/44	79.5[3]	31/39	77.4	24/31	100.0	7/7	100.0	24/24	53.3	7/15	0.0	0/7	22.6	8/31
Takahashi et al. 2010	54/70	75.9	41/54	72.7	32/44	90.0	9/10	97.0	32/33	42.9	9/21	10.0	1/10	27.3	12/44
Sakaguchi et al. 2011 (A)	23/25	65.2	15/23	52.9	9/17	100.0	6/6	100.0	9/9	42.9	6/14	0.0	0/6	47.1	8/17
Kojima et al. 2012 (B)	35/40	80.0	28/35	78.1	25/32	100.0	3/3	100.0	25/25	30.0	3/10	0.0	0/3	21.9	7/32
(A) + (B)	56/63	73.2	41/56	68.8	33/48	100.0	8/8	100.0	33/33	34.8	8/23	0.0	0/8	31.3	15/48
Pooled Studies[4]	98/119	73.7	73/99	67.6	50/74	95.8	23/24	98.0	50/51	48.9	23/47	4.2	1/24	32.4	24/74

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

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

[3]Performance calculated using the overall classification based on the majority classification among the multiple testing laboratories.

[4]Data from Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), and Kojima et al. (2012) and 23 additional data were pooled together and an overall classification was assigned for each test substance based on the majority classification obtained in each testing laboratory.
2-Ethylhexyl thioglycolate was excluded out from the overoall analysis due to its ability of direct reduction of MTT.

6.1.2.1 *Discordant Results According to the EPA Classification System*

In order to evaluate discordant responses of the STE test relative to the regulatory classification of ocular irritation, several accuracy sub-analyses were performed. These included specific classes of chemicals and certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants, physical form) same as indicated in **Table 6-2** in Section 6.1.1.1

As indicated in **Table 6-6**, there were some notable trends in the performance of the STE test among these subgroups of substances.

As can be seen in **Table 6-6**, although there was insufficiently numbers of irritants (i.e., positive substances), only one false positive result of polyethyleneglycol monolaurate (10E.O.) was obtained in all chemical classes and properties for each of the four studies (see **Appendix D**).

Although there were a small number (13 or 19) of substances represented, esters and hydrocarbons were underpredicted (i.e., were false negatives) at higher rate by the STE test according to the EPA classification system (see **Appendix D**). Additional chemical classes represented among the underpredicted substances were organic/inorganic salts, ketones/lactones, alcohols, and amines. Among the 16 substances labeled as surfactants, no was underpredicted by the STE test.

With regard to physical form of the substances underpredicted by the STE test, 18 were liquids and 6 were solids. Considering the proportion of the total available data, the rate of underprediction of liquids (73/96; 76%) and solids (23/96; 24%) were seems to be equivalent by the STE test.

There was no difference among the underpredicted substances for which pH information was available, as three was acidic (pH ≤ 7.0) and one was basic (pH > 7.0), therefore, the acidic substances (25/34; 74%) may have a tendency to underestimate. Although the numbers of substance among the overpredicted substances were few to resolve any definitive trends in overprediction by STE test, there was only one false positive substance (Polyethyleneglycol monolaurate (10E.O.)) of acidic (see **Appendix D**).

With regard to volatility of the substances underpredicted by the STE test, substance with the

vapor pressure between 10 kPa and 1 kPa were most often underpredicted by the STE test according to the EPA classification system. Additional often underpredicted vapor pressure ranges represented among the underpredicted substances were between 1 kPa and 0.1 kPa, under 0.01 kPa and over 10 kPa. The numbers of substance among the overpredicted substances were too few to resolve any definitive trends in overprediction by STE test.

Table 6-7 shows the effects on the STE test performance statistics of excluding from the data set problematic classes (i.e., that gave the most discordant results, according to the EPA classification system).

When the combination of organic/inorganic salts and esters were excluded from the data set, the accuracy was increased from 75% (73/98) to 78% (62/80), and the false negative rate was slightly decreased from 32% (24/74) to 29% (17/59). The largest changes were observed when all of organic/inorganic salts, esters, and alcohols were excluded from the data set; accuracy increased from 75% to 78% (45/58), and the false negative rate decreased from 32% to 29% (12/42).

Table 6-8 shows the effects on the STE test performance statistics of excluding from the data set problematic substances with volatility.

When the substance with vapor pressure over 6kPa was excluded from the data set, changes performance statistics were noted, with accuracy increasing from 71% (59/83) to 73% (50/69), and the false negative rate decreasing from 38% (24/63) to 36% (19/53). The largest changes were observed when the substances with vapor pressure over 2kPa were excluded from the data set; accuracy increased from 71% to 76% (47/62), and the false negative rate decreased from 38% to 32% (15/47).

Table 6-6 False Positive and False Negative Rates of the STE test, by Chemical Class and Properties of Interest, for the EPA classification

Category	N[1]	False Positive Rate[2]		False Negative Rate[3]	
		%	No.[4]	%	No.
Overall	98	4.2	1/24	32.4	24/74
Chemical Class[5]					
Alcohols	22	0.0	0/5	29.4	5/17
Amines	4	-	0/0	25.0	1/4
Carboxylic acids	3	-	0/0	0.0	0/3
Esters	13	0.0	0/3	50.0	5/10
Hydrocarbons	19	0.0	0/10	77.8	7/9
Ketones/Lactones	7	0.0	0/1	33.3	2/6
Organic/Inorganic salts	5	-	0/0	40.0	2/5
Properties Interest					
Liquid[6]	73	4.3	1/23	36.0	18/50
Solid[6]	23	0.0	0/1	27.3	6/22
Surfactants - Total	16	33.3	1/3	0.0	0/13
-nonionic	6	33.3	1/3	0.0	0/3
-anionic	3	-	0/0	0.0	0/3
-cationic	7	-	0/0	0.0	0/7
pH - Total[7]	34	14.3	1/7	14.8	4/27
- acidic (pH ≤ 7.0)	25	16.7	1/6	15.8	3/19
- basic (pH > 7.0)	8	-	-	12.5	1/8
Volatility - Total[8]	83	0.0	0/20	38.1	23/63
- VP[9] **≧ 10 kPa**	10	0.0	0/3	28.6	2/7
- 10 kPa > VP ≧ 1 kPa	14	0.0	0/2	66.7	8/12
- 1 kPa > VP ≧ 0.1 kPa	18	0.0	0/5	46.2	6/13
- 0.1 kPa > VP ≧ 0.01 kPa	16	0.0	0/5	18.2	2/11
- VP < 0.01 kPa	25	0.0	0/5	30.0	6/20

[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 STE test and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

[6]Physical form is unable to be classified for some substances that have proparties of both solid and liquid, and therefore the overall number does not equal the sum of the solid and liquid substances.

[7]Total number of substances for which pH information was obtained.
One substance was not categorized due to limited information.

[8]Total number of substances for which volatility information was obtained.

[9]VP = Vapor pressure

Table 6-7 Effect of Exclusion of Discordant Classes on False Negative and False Positive Rates of the STE test, for the EPA Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	74.5	73/98	4.2	1/24	32.4	24/74
w/o Salts (Organic/Inorganic)	75.3	70/93	4.0	1/24	31.9	22/69
w/o Esters	76.5	65/85	4.8	1/21	30.0	19/64
w/o Alcohols	73.7	56/76	5.3	1/19	33.3	19/59
w/o Salts & Esters	77.5	62/80	4.8	1/21	28.8	17/59
w/o Salts & Alcohols	74.6	53/71	5.3	1/19	32.7	17/52
w/o Esters & Alcohols	76.2	48/63	6.3	1/16	29.8	14/47
w/o Salts & Esters & Alcohols	77.6	45/58	6.3	1/16	28.6	12/42

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.

Table 6-8 Effect of Exclusion of Discordant Volatile Chemicals on False Negative and False Positive Rates of the STE test, for the EPA Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall with vapor pressure (VP) data	71.1	59/83	0.0	0/20	38.1	24/63
w/o VP > 15 kPa	71.4	55/77	0.0	0/18	37.3	22/59
w/o VP > 10 kPa	69.9	51/73	0.0	0/17	39.3	22/56
w/o VP > 8 kPa	70.4	50/71	0.0	0/16	38.2	21/55
w/o VP> 7kPa	71.4	50/70	0.0	0/16	37.0	20/54
w/o VP > 6kPa	72.5	50/69	0.0	0/16	35.8	19/53
w/o VP > 4kPa	73.1	49/67	0.0	0/15	34.6	18/52
w/o VP > 3 kPa	74.2	47/66	0.0	0/15	33.3	17/51
w/o VP > 2kPa	75.8	47/62	0.0	0/15	31.9	15/47

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.
VP, vapor pressure

6.2 Accuracy of the STE test for Identifying Ocular Irritant - Summary of Results

A little difference in performance of STE test among three hazard classification system evaluated (i.e, GHS [UN 2003] and EPA [EPA 1996]) were identified; the accuracy analysis

revealed that the STE test had a best performance in predicting GHS classification. As can be seen in **Table 6-1**, the over all accuracy of STE test in predicting GHS classification was 85%. Sensitivity and specificity were 84% and 86% respectively. The false positive rate and the false negative rate were 14% and 16% respectively. However, the accuracy analysis revealed that the performance of the STE test in predicting EPA classification was relatively low compared to that in GHS classification. As can be seen in **Table 6-5**, the over all accuracy of STE test in predicting EPA classification was 74%. The sensitivity was 68%. However, that in predicting EPA classification was high (96%). Resulting false positive rates in predicting EPA classification were 4%. The false negative rates in predicting EPA classifications were 32%.

6.2.1 Discordance Among Chemical Classes

For GHS classification, the accuracy analysis indicated that organic/inorganic salts, esters and ketones/lactones are often overpredicted (100% [1/1], 22% [2/9] and 25% [1/4] false positive rate, respectively) in the STE test. In contrast, organic/Inorganic salts were most often underpredicted by the STE test (50% [2/4] false negative rate). Moreover, hydrocarbons (50% [2/4]), esters (25% [1/4]), and alcohols (25% [4/16]) also had high false negative rates. The numbers of substances among the remaining chemical classes were too few to resolve any definitive trends in false prediction by the STE test. For the purposes of these analyses, we considered three substances to be the threshold number per chemical class for consideration, and thus chemical classes represented by fewer than three substances were not considered.

For EPA classification, the accuracy analysis indicated that only one overpredicted (Polyethyleneglycol monolaurate (10E.O.)) substance was identified in the STE test. In contrast, hydrocarbons (77.8% [7/9] false negative rate) and esters (50% [5/10]) were most often underpredicted by the STE test. Salts (40% [2/5]), ketones/lactones (33.3% [2/6]) and alcohols (31.3% [5/16]) also had high false negative rates.

6.2.2 Discordance Among Physical or Chemical Properties of Interest

With regard to physical form of the substances overpredicted by the STE test, one to five was liquids and zero or one was solids. Although the minor differences were existed depending on the classification system used, considering the proportion of the total available data, the rate of overprediction of liquids (74/99 or 73/98) and solids (23/99 or 23/98) were generally seems to be equivalent by the STE test.

With regard to physical form of the substances underpredicted by the STE test, five or six were solids and four to 18 were liquid. Although the minor differences were existed depending on the classification system used, despite the proportion of the total available database indicated above, the rate of underprediction of solids (27% or 25% false negative rate) and liquids (37% or 12% false negative rate) were generally seems to be equivalent by the STE test.

Exclusion of three discordant classes (i.e., organic/inorganic salts, esters and alcohols) from the data set resulted in an increased accuracy (from 85% to 90% [GHS] and from 75% to 78% [EPA]), and a decreased false negative rate (from 16% to 7% [GHS] and 32% to 29% [EPA]).

It's notable that 17 and 16 substances labeled as surfactants were not underpredicted by the STE test regardless the classification system used.

Among the underpredicted substances for which pH information was available for GHS classification, as two was acidic (pH ≤ 7.0) and zero was basic (pH > 7.0) regardless the classification system used. Therefore, the acidic substances (26/35; 74% or 25/34; 74%) may have a tendency to underestimate. However, it is noted that pH information was available for only 35 substances in all substances with *in vivo* data. The numbers of substance among the overpredicted substances were too few to resolve any definitive trends in overprediction by STE test.

Finally, with regard to volatility of the substances underpredicted by the STE test, the accuracy analysis indicated that the chemicals with the vapor pressure between 10kPa and 1kPa often underpredicted for GHS and EPA classification systems (50% [4/8] and 67% [8/12] false negative rate, respectively) in the STE test. When the substance with vapor pressure over 6kPa was excluded from the data set, changes performance statistics were noted regardless the classification system used. When the substances with vapor pressure around over 6kPa were excluded from the data set; accuracy increased (from 86% to 88% [GHS], and from 71% to 73% [EPA]), and the false negative rate decreased (from 20% to 16% [GHS], and from 38% to 36% [EPA])

6.3 Applicability domain

As described previously (Section 1.2.3), substances with insoluble either in saline, saline with

5% DMSO and mineral oil are enable to assay in the STE test, however, the colored test substances may be problematic as they could interfere with the optical density measured in MTT assay. In addition, if the test substances can reduce MTT directly, the measurement of OD_{570} would be affected by the direct reduction by test substances. In this case, STE test can be conducted basically. If obtained data was “non-irritant” in STE test for these substances, it can not be judged whether the STE test result is correct or not. Therefore, it would be assigned as "inconclusive". In 119 data set of STE test, there are 16 substances, which effect to MTT reduction as shown in **Appendix B-5**. Among 16 substances, 2-Ethylhexylthioglycolate (not classified by GHS and category IV by EPA classification) was assigned as “non irritant” in STE test. Therefore, this substance would be finally assigned as “inconclusive” due to insufficient information. In the **Appendix B-5**, the effect to direct reduction of MTT for each substance was listed.

Moreover, as a result of the findings regarding the predictive capacity of the STE test, the false negative rates of inorganic/organic salts, volatile substances including alcohols and esters were relatively high compared to that of other substances regardless the classification system used. False positive substances for GHS and EPA classification system were listed in **Table 6-9** and **Table 6-11**, respectively, and false negative substances for GHS and EPA classification system were listed in **Table 6-10** and **Table 6-12** respectively.

Table 6-9 False Positive Substances, for the GHS Classification System

Substance	GHS	Chemical class[1]	Form	Vapour pressure[2] (Pa, 25°C)
Cyclohexanone	Not classified	Ketones	Liquid	539
Ethyl acetate	Not classified	Esters (acetate)	Liquid	13100
Glycidyl Methacrylate	Not classified	Ester (methacrylate)	Liquid	82.9
Polyethyleneglycol monolaurate (10E.O.)	Not classified	Surfactants (nonionic)	Liquid	n.c.
Sodium 2-naphthalenesulfonate	Not classified	Organic salts	Solid	n.c.
Tween 20	Not classified	Surfactants (nonionic)	Liquid	n.c.

[1]Chemical classes and properties included in this table are represented by at least five substances tested in the STE test and assignments are based on the MeSH categories

[2]Vapore pressure values were culculated Estimation Program Interface (EPI) - Suite™ program developed by the Environmental Protection Agency (EPA) and ACD Labs

[2]n.c.=Not calculated (due to any case as followes; miscellaneous, aqueous solution, polydisperse substance, have a large molecular weight >500 Da, or have a coordinate bond)

Table 6-10 False Negative Substances, for the GHS Classification System

Substance	GHS	Chemical class[1]	Form	Vapour pressure[2] (Pa, 25°C)
2,5-Dimethyl-2,5-hexanediol	Category 1	Alcohols	Solid	0.578
Sodium salicylate	Category 1	Organic salts	Solid	4.84E-09
Ammonium nitrate	Category 2B	Inorganic salts	Solid	4.48E-13
Camphene	Category 2B	Hydrocarbons (cyclic)	Solid	237
Ethanol	Category 2A	Alcohols	Liquid	8120
Isopropanol	Category 2A	Alcohols	Liquid	6610
Methyl acetate	Category 2A	Esters (acetate)	Liquid	7030
Myristyl alcohol	Category 2A	Alcohols	Solid	0.0269
Toluene	>Category 2B	Hydrocarbons (cyclic)	Liquid	3160

[1]Chemical classes and properties included in this table are represented by at least five substances tested in the STE test and assignments are based on the MeSH categories

[2]Vapore pressure values were culculated Estimation Program Interface (EPI) - Suite™ program developed by the Environmental Protection Agency (EPA) and ACD Labs

Table 6-11 False Positive Substance, for the EPA Classification System

Substance	GHS	Chemical class[1]	Form	Vapour pressure[2] (Pa, 25°C)
Polyethyleneglycol monolaurate (10E.O.)	Not classified	Surfactants (nonionic)	Liquid	n.c.

[1]Chemical classes and properties included in this table are represented by at least five substances tested in the STE test and assignments are based on the MeSH categories

[2]Vapore pressure values were culculated Estimation Program Interface (EPI) - Suite™ program developed by the Environmental Protection Agency (EPA) and ACD Labs

[2]n.c.=Not calculated (due to any case as followes; miscellaneous, aqueous solution, polydisperse substance, have a large molecular weight >500 Da, or have a coordinate bond)

Table 6-12 False Negative Substances, for the EPA Classification System

Substance	GHS	Chemical class[1]	Form	Vapour pressure[2] (Pa, 25°C)
1,5-Hexadiene	Category III	Hydrocarbons	Liquid	28600
2,2-Dimethyl-3-pentanol	Category III	Alcohols	Liquid	413
2,5-Dimethyl-2,5-hexanediol	Category I	Alcohols	Solid	0.578
2-Ethoxyethyl acetate	Category III	Esters (acetate)	Liquid	397
Ammonium nitrate	Category III	Inorganic salts	Solid	4.48E-13
Butyl acetate	Category III	Esters (acetate)	Liquid	1586.5
Camphene	Category III	Hydrocarbons (cyclic)	Solid	237
Dimethyl sulfoxide	Category III	Sulphoxides	Liquid	82.9
Dodecane	Category III	Hydrocarbons	Liquid	31.5
Ethanol	Category III	Alcohols	Liquid	8120
Ethyl trimethyl acetate	Category III	Esters (acetate)	Liquid	2240
Ethylhexyl salicylate	Category III	Esters (salicylate)	Liquid	0.000951
Isopropanol	Category III	Alcohols	Liquid	6610
Methyl acetate	Category III	Esters (acetate)	Liquid	7.03E+03
Methyl amyl ketone	Category III	Ketones	Liquid	655
Methyl cyclopentane	Category III	Hydrocarbons (cyclic)	Liquid	17800
Methyl isobutyl ketone	Category III	Ketones	Liquid	2900
Myristyl alcohol	Category III	Alcohols	Solid	0.0269
n,n-Dimethylguanidine sulfate	Category III	Amidines	Solid	4040
Sodium salicylate	Category I	Organic salts	Solid	4.84E-09
Styrene	Category III	Hydrocarbons (cyclic)	Liquid	673
Toluene	Category III	Hydrocarbons (cyclic)	Liquid	3160
Triethanolamine	Category III	Amines	Liquid	0.000451
Xylene	Category II	Hydrocarbons (cyclic)	Liquid	882.6

[1]Chemical classes and properties included in this table are represented by at least five substances tested in the STE test and assignments are based on the MeSH categories

[2]Vapore pressure values were culculated Estimation Program Interface (EPI) - Suite™ program developed by the Environmental Protection Agency (EPA) and ACD Labs

6.3.1 Define applicability domain

6.3.1.1 *Define applicability domain excluding organic/inorganic salts and high-volatile substances*

For GHS classification, when salts was excluded out from over all substances, the accuracy was improved from 85% to 87%, and the number of false negative substances and false positive substances were decreased from nine to seven and six to five, respectively. In addition, when the substances with vapor pressure over 6 kPa were excluded out from the substance that have vapor pressure data except for salts, the number of false positive was finally decreased to two (glycidyl Methacrylate, cyclohexanone), and the number of false negative was finally decreased to four (2,5-dimethyl-2,5-hexanediol, camphene, toluene and myristyl alcohol). The highest accuracy (91%) was obtained when the substances that have vapor pressure over 6kPa were excluded out with the exclusion of salts (**Table 6-13**).

Table 6-13 Effect of Exclusion of Discordant Chemicals and Properties on False Negative and False Positive Rates of the STE test method, for the GHS Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	84.8	84/99	13.6	6/44	16.4	9/55
w/o Salts (Inorganic/Organic)	87.2	82/94	11.6	5/43	13.7	7/51
w/o Salts with vapor pressure (VP) data	87.3	69/79	7.9	3/38	17.1	7/41
w/o Salts and VP > 6 kPa	90.8	59/65	6.5	2/31	11.8	4/34

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.
VP, vapor pressure

For EPA classification, when organic/inorganic salts were excluded out from overall substances, the accuracy and the false negative rate were not changed contrary to the GHS classification. In contrast, when the substances with vapor pressure over 6 kPa were excluded out from the substance that have vapor pressure data except for salts, the number of false negative was finally decreased to 17, although the overall accuracy for EPA classification was not increased by excluding the salts and high volatile substances (**Table 6-14**).

Table 6-14 Effect of Exclusion of Discordant Chemicals and Properties on False Negative and False Positive Rates of the STE test method, for the EPA Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	74.5	73/98	4.2	1/24	32.4	24/74
w/o Salts (Inorganic/Organic)	75.3	70/93	4.2	1/24	31.9	22/69
w/o Salts with vapor pressure (VP) data	72.2	57/79	0.0	0/20	37.3	22/59
w/o Salts and VP > 6 kPa	73.8	48/65	0.0	0/16	34.7	17/49

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.
VP, vapor pressure

6.3.1.2 *Define applicability domain excluding several solids, and high-volatile substances*

Although the exclusion mentioned section 6.3.1.1 was applied, four false negative results remains (2,5-dimethyl-2,5-hexanediol, camphene, toluene and myristyl alcohol) for GHS classification. Among these 4 substances, three out of four are solids. **Table 6-15** shows the accuracy, false positive rate, and false negative rate for the surfactants, salts, amines/amidines, alcohols, hydrocarbons, or the others in solid substances. Although there were small number of substances represented, false negative rate of solid alcohols and solid hydrocarbons were relatively higher than other chemical classes (**Table 6-15**). When the substances in several chemical classes, which higher false negative rates were obtained, were excluded from the data set, the false negative was decreased to seven or eight. In addition, when the substances that were solid salts, solid alcohols, and solid hydrocarbons were all excluded out from the data set, the highest accuracy (90.0%) were obtained and the false negatives were decreased to four (toluene, isopropanol, methyl acetate, ethanol).

As described in section 6.3.1.1, when the substances with vapor pressure over 6 kPa were excluded out from the substance that have vapor pressure data except for salts, the number of false negative was decreased. Therefore, both of the substances with vapor pressure over 6 kPa and the substances that were solid salts, solid alcohols, and solid hydrocarbons should be excluded. When the substances of solid salts, solid alcohols, and solid hydrocarbons, and high volatile substances (>6kPa) were excluded out from the dataset, the accuracy was finally 92.2% and false negative were decreased to one substance (toluene). When adopting the applicability domain to exclude the solid salts, solid alcohols, solid hydrocarbons, and high volatile substances (>6kPa), four false positives (glycidyl methacrylate, cyclohexanone, tween 20, and polyethyleneglycol monolaurate (10E.O.)) and one false negatives (toluene)

remain. The mechanisms for the over-prediction of these substances still remain to be established. On the other hands, regarding the toluene, it has been confirmed not to affect to change the corneal opacity from *in vivo* data (unpublished data from TSCA).

Table 6-15 Effect of Exclusion of Discordant Chemicals and Properties on False Negative and False Positive Rates of the STE test method, for the GHS Classification System by Adopting to the Selection by Chemical Class

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	84.8	84/99	13.6	6/44	16.4	9/55
Solids	73.9	17/23	100.0	1/3	25.0	5/20
- Surfactants	100.0	9/9	0.0	0/1	0.0	0/8
- Organic/inorganic salts (A)	40.0	2/5	100.0	1/1	50.0	2/4
- Amines/Amidines (B)	100.0	3/3	0.0	0/1	0.0	0/2
- Alcohols (C)	0.0	0/2	-	0/0	100.0	2/2
- Hydrocabrbons (D)	50.0	1/2	-	0/0	50.0	1/2
- Others	100.0	2/2	-	0/0	0.0	0/2
w/o (A)	87.2	82/94	11.6	5/43	13.7	7/51
w/o (C)	86.6	84/97	13.6	6/44	13.2	7/53
w/o (D)	85.6	83/97	13.6	6/44	15.1	8/53
w/o (A) and (C)	89.1	82/92	10.9	5/43	10.2	5/49
w/o (A) and (D)	88.0	81/92	10.9	5/43	12.2	6/49
w/o (C) and (D)	86.4	83/95	13.6	6/44	11.8	6/51
w/o (A), (C), and (D)	90.0	81/90	11.6	5/43	8.5	4/47
w/o solids	88.2	67/76	12.2	5/41	11.4	4/35

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.

Table 6-16 Effect of Exclusion of Discordant Chemicals and Properties on False Negative and False Positive Rates of the STE test method, for the GHS Classification System by Adopting to the Selection by Chemical Class and Vapor Pressure

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	84.8	84/99	13.6	6/44	16.4	9/55
w/o salts and VP >6 kPa	88.9	72/81	13.5	4/36	9.1	4/44
w/o solid salts, solid alcohols, and VP>6 kPa	91.1	72/79	13.5	4/36	4.8	2/42
w/o solid salts, solid alcohols, and solid hydrocarbons, and VP >6 kPa	92.2	71/77	13.5	4/36	2.5	1/40

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.
VP, vapor pressure

For EPA classification, when the substances, which were solid salts, alcohols, hydrocarbons, and high volatile substances with high vapor pressure over 6 kPa, were excluded out from the data set, the accuracy was 80.0% (60/75). Although only one false positive was obtained, 14 false negatives remain (false negative rate: 25.5%) (**Table 6-17**).

To improve the rate of underprediction, the viability threshold in STE test has been changed from 70% to 75 or 80%. **Table 6-18** shows that not only the overall accuracy but also the rate of underprediction and overprediction were completely same among the three thresholds (70, 75, and 80%). As a result, the change of threshold in STE test would not insufficient for improving the rate of underprediction against EPA classification.

After excluding the solid salts, alcohols, hydrocarbons, and high volatile substances, it remains 14 false negatives for EPA classification listed in **Table 6-19**. Among the 14 substances, 5 substances did not induce the effect to corneal opacity *in vivo*. For other 9 substances, the mechanisms for the underprediction of these substances still remain to be established.

Table 6-17 Effect of Exclusion of Discordant Chemicals and Properties on False Negative and False Positive Rates of the STE test method, for the EPA Classification System by Adopting to the Selection by Chemical Class and Vapor Pressure

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Overall	74.5	73/98	4.2	1/24	32.4	24/74
w/o salts and VP >6 kPa	77.2	61/79	5.0	1/20	28.8	17/59
w/o solid salts, solid alcohols, and VP>6 kPa	79.2	61/77	5.0	1/20	26.3	15/57
w/o solid salts, solid alcohols, and solid hydrocarbons, and VP >6 kPa	80.0	60/75	5.0	1/20	25.5	14/55

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.
VP, vapor pressure

Table 6-18 Effect of Change of Threshold in STE and Properties on False Negative and False Positive Rates of the STE test method, for the EPA Classification System

Data Set	Accuracy		False Positive Rate[1]		False Negative Rate[2]	
	%	No.[3]	%	No.	%	No.
Threshold = 70% (Original)	74.5	73/98	4.2	1/24	32.4	24/74
Threshold = 75%	74.5	73/98	4.2	1/24	32.4	24/74
Threshold = 80%	74.5	73/98	4.2	1/24	32.4	24/74

[1]False Positive Rate = The proportion of all negative substances that are falsely identified as positive in vitro.
[2]False Negative Rate = The proportion of all positive substances that are falsely identified as negative in vitro.
[3]Data used to calculate the percentage.

Table 6-19 False Negative Substances after Excluding the Solid Salts, Solid Alcohol, Solid Hydrocarbons, and High Volatile Substances, for the EPA Classification System

Substance	GHS	Chemical class[1]	Form	Vapour pressure[2] (Pa, 25°C)	In vivo effect to Corneal opacity
2,2-Dimethyl-3-pentanol	Category III	Alcohols	Liquid	413	Yes
2-Ethoxyethyl acetate	Category III	Esters (acetate)	Liquid	397	Yes
Butyl acetate	Category III	Esters (acetate)	Liquid	1586.5	Yes
Dimethyl sulfoxide	Category III	Sulphoxides	Liquid	82.9	Yes
Dodecane	Category III	Hydrocarbons	Liquid	31.5	No
Ethyl trimethyl acetate	Category III	Esters (acetate)	Liquid	2240	Yes
Ethylhexyl salicylate	Category III	Esters (salicylate)	Liquid	0.000951	No
Methyl amyl ketone	Category III	Ketones	Liquid	655	Yes
Methyl isobutyl ketone	Category III	Ketones	Liquid	2900	Yes
n,n-Dimethylguanidine sulfate	Category III	Amidines	Solid	4040	No
Styrene	Category III	Hydrocarbons (cyclic)	Liquid	673	Yes
Toluene	Category III	Hydrocarbons (cyclic)	Liquid	3160	No
Triethanolamine	Category III	Amines	Liquid	0.000451	Yes
Xylene	Category II	Hydrocarbons (cyclic)	Liquid	882.6	No

[1]Chemical classes were assigned from the study report or the other source (ex. MeSH system)
[2]Vapore pressure values were culculated Estimation Program Interface (EPI) - SuiteTM program developed by the Environmental Protection Agency (EPA) and ACD Labs (available at: http://www.acdlabs.com).

Finally, the vapor pressure of most surfactants was not calculated. Moreover, taking sodium lauryl sulfate as the example, this substance is known to be also categorized in the solid salts or solid alcohols. However, as described in Section 6.1.1.1 or 6.1.2.1, there was no false negative result for the surfactant for not only GHS classification but also EPA classification. Therefore, the substances categorized into surfactants would be able to be adopted for STE test, even if it is also categorized into solid salts, solid alcohols, and solid hydrocarbons or its vapor pressure is not calculated.

In summary, the BRD analysis indicated that solid salts, solid alcohols, solid hydrocarbons except for surfactants and high volatile sample (VP>6 kPa) seems to be out of applicability domain of the STE test, regardless the classification system used.

6.3.2 Effective application a decision tree

The predictive potential of a tiered approach combining the STE test and the BCOP assay for assessing GHS eye irritation categories has been examined. As a result, the good prediction accuracy of the GHS eye irritation categories was confirmed using the tiered approach combination of STE test and BCOP assay. (Hayashi et al., 2012a). However, insoluble substances in any vehicles for STE test (i.e., saline, 5% DMSO in saline, or mineral oil) were unable to assay according to this tiered approach.

On the other hands, recently, several 3D reconstituted tissue models (e.g. EpiOcular™, MatTek Co.; SkinEthic HCE, SkinEthic laboratories; LabCyte CORNEA-MODEL, J-TEC Co., Ltd.) capable of exposure with neat substances without any vehicle were developed all over the world.

Therefore, to solve these issues, the possibility of a tiered approach combining the STE test, the EpiOcular assay, and the BCOP assay (**Figure 6-1**) for predicting the eye irritation potential of substances not soluble in saline or mineral oil were further assessed. The possibility of achieving accurate estimation of irritation potential for substances, which are solid salts, solid alcohols, solid hydrocarbons, and high volatile substances with vapor pressure over 6 kPa, was also examined by EpiOcular (**Appendix I**). Twenty-three substances, which are solid salts, alcohols, hydrocarbons, and high volatile substances with vapor pressure over 6 kPa, were examined by EpiOcular. Among 23 substances, the accuracy for the GHS and EPA classification were both 87%. Only one false negative (myristyl alcohol) for GHS classification was obtained. For the EPA classification, three false negatives (myristyl alcohol, methyl cyclopentane, 1,5-hexadiene) were obtained. As a result, almost false negatives out of applicability domain against GHS classification in STE test were able to be estimated accurately by EpiOcular. Therefore, the tiered approach combining the STE test, the EpiOcular assay, and also the BCOP assay was allowed to estimate the eye irritation potential of not only insoluble substance but also the substances categorized into solid salts, solid alcohols, solid hydrocarbons and the substances with vapor pressure over 6 kPa accurately. Accordingly, the good prediction accuracy and under prediction rate of the GHS eye irritation categories was confirmed using this approach combining the STE test, the EpiOcular assay, and the BCOP assay (Hayashi et al. 2012b, submitted for publication).

From these findings, the tiered approach combining the STE test, the EpiOcular assay, and the BCOP assay might be a promising alternative eye irritation testing strategy capable of testing for wide range of test substances regardless of solubility and volatility with minimum under prediction.

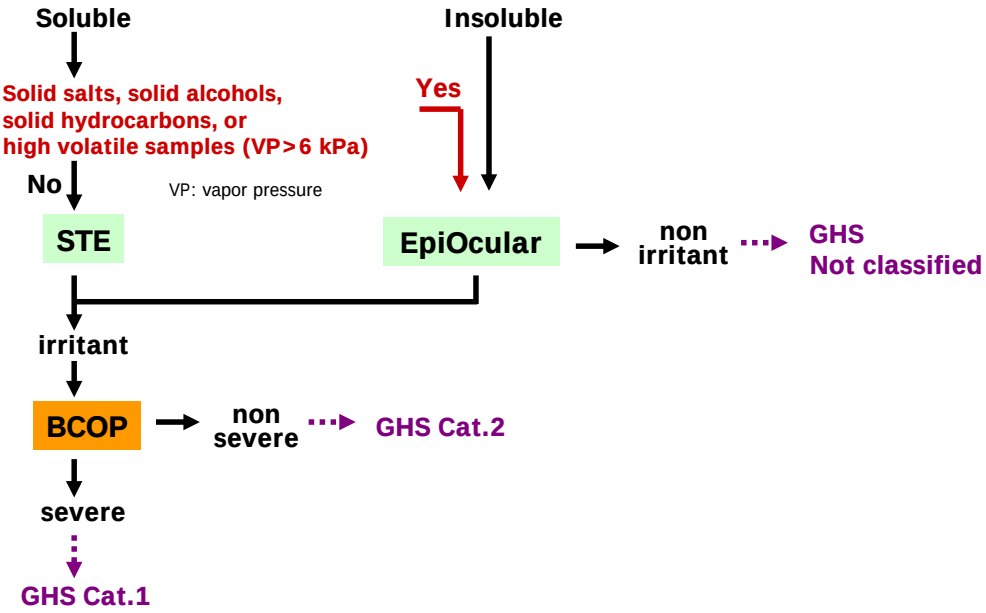

Figure 6-1 Tiered (bottom-up) Approach for Identification Ocular Irritant (Hayashi et al. 2012b, submitted for publication)

6.3.3 Conclusion

The STE test using cultured cells have the advantage of being simple, a quick procedure, and a low evaluation cost. Furthermore, poorly water-soluble chemicals like octanol and hexanol could be evaluated in the STE test by using mineral oil as the vehicle. Therefore, the STE test could be considered a building block assay in a tiered approach to establishing an ocular irritation animal alternative testing paradigm.

7.0 STE test Reliability

Quantitative STE test data were available for replicate experiments within an individual laboratory for four studies (Takahashi et al. 2009, Takahashi et al. 2010, Sakaguchi et al. 2011 and Kojima et al. 2012). The author-supplied individual STE test data were applied for an evaluation of the repeatability and/or intralaboratory reproducibility of the STE test. Additionally, comparable STE data were available for multiple laboratories within four comparative validation studies mentioned above, which allowed for an evaluation of the interlaboratory reproducibility of the STE test.

7.1 Selection Rationale for Substances Used to Evaluate the Reliability of STE test

The rationale for substance selection used in the various intralaboratory and multilaboratory studies was previously discussed in Section 3.0. The rationale for the chemicals selected was not known for two studies (Takahashi et al. 2009 and Takahashi et al. 2010). Meanwhile, the rationale for substance selection was described for other two studies. In brief, substances were selected for inclusion based on available *in vivo* rabbit eye data for comparison, to cover the range of ocular irritation potential.

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Assessment of Intralaboratory Repeatability and Reproducibility

All of the studies discussed in Section 6.0 included intralaboratory data (Takahashi et al. 2009, Takahashi et al. 2010, Sakaguchi et al. 2011, and Kojima et al. 2012). The STE test data were available for replicate within individual experiments repeated three times for each test substance in two to five different laboratories. CV analyses were performed on within-experiment and between-laboratory STE data, using the viability value obtained for each test substance within each of the two to five testing laboratories.

7.2.1.1 *Takahashi et al. (2009)*

Intralaboratory Repeatability:

In this study, 44 substances were evaluated in three laboratories multiple times (two or three experiments). A %CV value was calculated for the viability value for each test (**Appendix E1**). **Tables 7-1**, **7-2**, and **7-3** summarize the mean and the %CV values of the viability for each test conducted in Laboratory A, Laboratory B, and Laboratory C, respectively. The results for each laboratory are sorted by %CV values from lowest to highest value.

Table 7-1 Intralaboratory Repeatability of Viability Values for Replicate Laboratory A, Takahashi et al. 2009[1]

Substance	Mean Viability (n=2-3)	% CV[2]	Category Classification
Methyl amylketone	91.7	0.3	Non Irritant
Polyethylene glycol 400	97.8	1.0	Non Irritant
Methyl isobutyl ketone	88.5	2.8	Non Irritant
Toluene	101.3	3.1	Non Irritant
Isopropyl alcohol	101.6	3.6	Non Irritant
Tween20	21.1	3.6	Irritant
3-Methoxy-1,2-propanediol	109.5	3.8	Non Irritant
Methyl cyclopentane	102.2	4.8	Non Irritant
Acetone	9.6	5.8	Irritant
Glycerin	99.3	6.3	Non Irritant
Polyoxyethylene Hydrogenated Castor Oil (60E.O.)	83.5	7.2	Non Irritant
Triethanolamine	90.6	7.2	Non Irritant
Sodium salicylate	89.7	8.2	Non Irritant
3,3-Dimethylpentane	92.6	8.8	Non Irritant
Ethanol	98.2	9.7	Non Irritant
Lactic acid	4.4	12.7	Irritant
Butanol	8.4	16.3	Irritant
m-Phenylenediamine	6.4	18.7	Irritant
2-Ethyl-1-hexanol	44.0	21.8	Irritant
Silicic Anhydride	81.9	23.5	Non Irritant
Benzalkonium Chloride	2.1	24.4	Irritant
Acetic acid	4.5	30.1	Irritant
Glycolic acid	3.4	30.5	Irritant
Stearyltrimethylammonium Chloride	1.1	32.8	Irritant
2-Benzyloxyethanol	2.0	48.2	Irritant
Domiphen Bromide	3.6	49.3	Irritant
Cyclohexanol	1.4	52.8	Irritant
Benzyl Alcohol	3.4	61.7	Irritant
Sodium hydroxide	1.4	109.8	Irritant
2-Methyl-1-pentanol	1.8	123.7	Irritant
Cetylpyridinium bromide	0.6	143.3	Irritant
Di(2-Ethylhexyl) SodiumSulfosuccinate	2.5	146.3	Irritant
Potassium Laurate	0.8	236.7	Irritant
Cetyltrimethylammonium Bromide	0.4	254.2	Irritant
2-Ethylhexyl p-Dimethylamino Benzoate	104.8	?	Non Irritant
Isopropyl Myristate	105.7	?	Non Irritant
1-Octanol	-0.5	NA	Irritant
Acid Red 92	-0.1	NA	Irritant
Cetylpyridinium Chloride	-0.1	NA	Irritant
Diisopropanolamine	-1.1	NA	Irritant
Monoethanolamine	-1.0	NA	Irritant
n-Hexanol	-0.3	NA	Irritant
Sucrose Fatty Acid Ester	-0.6	NA	Irritant
Triton X-100	-0.2	NA	Irritant
Mean %CV	44.5		
Median %CV	17.5		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

[2]A "?" indicats that a CVs could not be calculated due to the lack of the number of experiment for this substance.

Table 7-2 Intralaboratory Repeatability of Viability Values for Replicate Laboratory B, Takahashi et al. 2009[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Isopropyl alcohol	94.8	1.8	Non Irritant
3-Methoxy-1,2-propanediol	103.6	2.6	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	99.6	2.9	Non Irritant
Methyl cyclopentane	94.5	3.9	Non Irritant
Ethanol	95.2	4.3	Non Irritant
Silicic anhydride	72.5	4.4	Non Irritant
Triethanolamine	93.6	4.6	Non Irritant
Toluene	95.6	4.8	Non Irritant
Glycerol	97.8	5.1	Non Irritant
Methyl amyl ketone	87.7	5.9	Non Irritant
3,3-Dimethylpentane	89.1	6.0	Non Irritant
Methyl isobutyl ketone	89.5	6.1	Non Irritant
Polyethylene glycol 400	97.4	6.5	Non Irritant
Isopropyl myristate	92.5	7.0	Non Irritant
Polyoxyethylene hydrogenated castor Oil (60E.O.)	91.4	8.4	Non Irritant
2-Ethyl-1-hexanol	69.9	8.9	Irritant
Sodium salicylate	88.1	10.7	Non Irritant
Acetone	35.6	14.7	Irritant
Monoethanolamine	1.7	20.6	Irritant
Butanol	6.0	21.2	Irritant
Tween 20	29.8	32.1	Irritant
Acetic acid	3.7	33.2	Irritant
Glycolic acid	4.0	35.4	Irritant
Lactic acid	4.7	44.8	Irritant
Acid red 92	1.4	55.4	Irritant
Benzyl alcohol	3.2	64.1	Irritant
m-Phenylenediamine	5.0	69.5	Irritant
2-Benzyloxyethanol	3.6	69.5	Irritant
Cyclohexanol	2.6	78.9	Irritant
Triton X-100	1.2	100.7	Irritant
n-Hexanol	1.0	105.1	Irritant
2-Methyl-1-pentanol	1.6	122.6	Irritant
Sodium hydroxide	0.6	150.1	Irritant
Potassium laurate	0.7	161.5	Irritant
Diisopropanolamine	0.6	233.3	Irritant
1-Octanol	0.7	270.7	Irritant
Cetyltrimethylammonium bromide	0.6	404.5	Irritant
Benzalkonium chloride	-0.4	NA	Irritant
Cetylpyridinium bromide	-1.0	NA	Irritant
Cetylpyridinium chloride	-0.6	NA	Irritant
Di(2-Ethylhexyl) sodium sulfosuccinate	-1.7	NA	Irritant
Domiphen bromide	-0.9	NA	Irritant
Stearyltrimethylammonium chloride	-0.5	NA	Irritant
Sucrose fatty acid ester	-3.9	NA	Irritant
Mean %CV	59.0		
Median %CV	20.6		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

Table 7-3 Intralaboratory Repeatability of Viability Values for Replicate Laboratory C, Takahashi et al. 2009[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Ethanol	98.9	1.3	Non Irritant
Isopropyl alcohol	106.4	1.9	Non Irritant
Triethanolamine	109.6	1.9	Non Irritant
3-Methoxy-1,2-propanediol	94.4	2.3	Non Irritant
Silicic Anhydride	78.4	2.3	Non Irritant
Polyethylene glycol 400	98.1	3.3	Non Irritant
Glycerin	108.0	3.8	Non Irritant
Methy amylketone	85.6	4.5	Non Irritant
2-Ethylhexyl p-Dimethylamino Benzoate	100.8	4.8	Non Irritant
3,3-Dimethylpentane	97.4	5.1	Non Irritant
Polyoxyethylene Hydrogenated CastorOil (60E.O.)	122.6	5.2	Non Irritant
Benzyl Alcohol	3.6	5.8	Irritant
2-Ethyl-1-hexanol	48.0	5.9	Irritant
Toluene	96.4	6.6	Non Irritant
Sodium Salicylate	84.5	7.1	Non Irritant
Isopropyl Myristate	100.4	9.3	Non Irritant
Methyl cyclopentane	98.1	9.9	Non Irritant
Methyl isobutyl ketone	98.2	10.1	Non Irritant
Butanol	7.0	10.2	Irritant
Domiphen Bromide	3.9	20.0	Irritant
Benzalkonium Chloride	1.7	21.1	Irritant
Lactic acid	4.3	21.4	Irritant
Acetone	13.6	27.3	Irritant
m-Phenylenediamine	11.3	34.3	Irritant
2-Benzyloxyethanol	2.8	37.8	Irritant
n-Hexanol	1.8	40.2	Irritant
Acetic acid	3.7	42.9	Irritant
Glycolic acid	3.8	46.4	Irritant
Tween20	3.5	72.8	Irritant
1-Octanol	4.7	76.6	Irritant
2-Methyl-1-pentanol	2.1	79.5	Irritant
Diisopropanolamine	0.4	95.8	Irritant
Monoethanolamine	0.4	246.2	Irritant
Di(2-Ethylhexyl) SodiumSulfosuccinate	1.3	394.0	Irritant
Cyclohexanol	0.2	462.2	Irritant
Sodium hydroxide	0.2	771.8	Irritant
Acid Red 92	-0.2	NA	Irritant
Cetylpyridinium bromide	-1.6	NA	Irritant
Cetylpyridinium Chloride	-1.3	NA	Irritant
Cetyltrimethylammonium Bromide	-0.3	NA	Irritant
Potassium Laurate	-0.4	NA	Irritant
Stearyltrimethylammonium Chloride	-1.4	NA	Irritant
Sucrose Fatty Acid Ester	-2.5	NA	Irritant
Triton X-100	-0.2	NA	Irritant
Mean %CV	72.0		
Median %CV	10.2		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

The ranges of %CV values for substances classified as non irritants (i.e., viability value > 70) are 0.3 to 23.5 for Laboratory A, 1.8 to 10.7 for Laboratory B, and 1.3 to 10.1 for Laboratory C. The ranges of %CV for substances classified as irritant (i.e., viability value ≤ 70) are 3.6 to 254.2 for Laboratory A, 8.9 to 404.5 for Laboratory B, and 5.8 to 771.8 for Laboratory C. The within experiment, the mean and the median %CV values for the three laboratories for all substances ranged from 72.0 to 44.5 and 20.6 to 10.2 respectively. Substances classified *in vitro* as irritants (i.e., viability value ≤ 70) tended to have greater %CV values. The CVs for seven to eight chemicals in all laboratories were not possible to calculate due to the existence of a negative value (noted in the table as NA; Not applicable). The CVs for two chemicals in laboratory A were also not possible to calculate due to lack of the number of experiment for this chemical (noted in the table as "?").

7.2.1.2 *Takahashi et al. (2010)*

Intralaboratory Repeatability:

In this study, 70 substances were evaluated in two laboratories multiple times (two or three experiments). A %CV value was calculated for the viability value for each test (**Appendix E1**). **Tables 7-4** and **7-5** summarize the mean and the %CV values of the viability for each test conducted in Laboratory A, and Laboratory B, respectively. The results for each laboratory are sorted by %CV values from lowest to highest value.

Table 7-4 Intralaboratory Repeatability of Viability Values for Replicate Laboratory A, Takahashi et al. 2010[1]

Substance	Mean Viability (n=2-3)	% CV	Category Classification
Methyl amylketone	91.7	0.3	Non Irritant
Dimethyl sulfoxide	95.3	2.3	Non Irritant
Propylene glycol	96.4	2.6	Non Irritant
Methyl isobutyl ketone	88.5	2.8	Non Irritant
Toluene	101.3	3.1	Non Irritant
Mineral oil	97.1	3.3	Non Irritant
Glycerin	95.7	3.4	Non Irritant
Isopropyl alcohol	101.6	3.6	Non Irritant
Isopropyl Myristate	106.0	3.6	Non Irritant
Tween20	21.1	3.6	Irritant
2-Ethylhexyl p-Dimethylamino Benzoate	106.4	4.4	Non Irritant
Butyl cellosolve	5.6	4.5	Irritant
Polyoxyethylene Hydrogenated CastorOil (60E.O.)	117.9	4.7	Non Irritant
Methyl cyclopentane	102.2	4.8	Non Irritant
Polyethylene glycol 400	92.1	4.9	Non Irritant
Acetone	9.6	5.8	Irritant
Cyclopentasiloxane	106.2	6.6	Non Irritant
Silicic Anhydride	79.5	7.3	Non Irritant
Physiologocal saline	91.6	7.9	Non Irritant
3,3-Dimethylpentane	92.6	8.8	Non Irritant
Diisopropanolamine	0.6	9.1	Irritant
Tween80	114.1	9.2	Non Irritant
2-Ethoxyethyl acetate	102.0	9.4	Non Irritant
Ethanol	98.2	9.7	Non Irritant
Sodium Salicylate	88.8	10.5	Non Irritant
Isobutyl alcohol	6.1	10.5	Irritant
Methyl ethyl ketone (2-butanone)	44.7	11.7	Irritant
Ethylhexyl salicylate	107.7	11.7	Non Irritant
Polyoxyethylene 23 lauryl ether	92.3	11.7	Non Irritant
Triethanolamine	101.6	12.1	Non Irritant
Lactic acid	4.4	12.7	Irritant
3-Methoxy-1,2-propanediol	93.6	13.7	Non Irritant
Myristyl alcohol	81.0	14.1	Non Irritant
Distearyldimethylammonium chloride	57.6	14.2	Irritant
Ethyl acetate	7.8	14.9	Irritant
Propylene carbonate	67.2	15.4	Irritant
Acetic acid	4.9	16.2	Irritant
Polyoxyethylene 8 stearate	33.5	16.2	Irritant
Butanol	8.4	16.3	Irritant
Myristic acid	10.9	17.3	Irritant

Table 7-4 (continued)

Substance	Mean Viability (n=2-3)	% CV	Category Classification
Calcium Thioglycollate	7.0	17.4	Irritant
m-Phenylenediamine	6.4	18.7	Irritant
2-Ethyl-1-hexanol	44.0	21.8	Irritant
Sodium 2-naphthalenesulfonate	3.1	23.3	Irritant
Benzalkonium Chloride	2.1	24.4	Irritant
Glycolic acid	3.4	30.5	Irritant
Stearyltrimethylammonium Chloride	1.1	32.8	Irritant
2-Benzyloxyethanol	2.0	48.2	Irritant
Domiphen Bromide	3.6	49.3	Irritant
Cyclohexanol	1.4	52.8	Irritant
Polyethylene Glycol Monolaurate (10E.O.)	2.3	53.6	Irritant
Benzyl Alcohol	3.4	61.7	Irritant
Sodium hydroxide	1.4	109.8	Irritant
Polyoxyethylene 10 lauryl ether	1.2	110.5	Irritant
Sodium lauryl sulfate	0.3	113.6	Irritant
2-Methyl-1-pentanol	1.8	123.7	Irritant
Cetylpyridinium bromide	0.6	143.3	Irritant
Monoethanolamine	0.5	144.2	Irritant
Di(2-Ethylhexyl) SodiumSulfosuccinate	2.5	146.3	Irritant
Potassium Laurate	0.8	236.7	Irritant
Cetyltrimethylammonium Bromide	0.4	254.2	Irritant
Sodium Polyoxyethylene LauryletherSulfate (2E.O.) (27%)	0.2	462.2	Irritant
1-Octanol	-0.5	NA	Irritant
Acid Red 92	-0.1	NA	Irritant
Cetylpyridinium Chloride	-0.1	NA	Irritant
n-Hexanol	-0.3	NA	Irritant
n-Lauroylsarcosinatesodium salt (30% solution)	-0.3	NA	Irritant
Polyoxyethylene 5 lauryl ether	-1.4	NA	Irritant
Sucrose Fatty Acid Ester	-0.6	NA	Irritant
Triton X-100	-0.1	NA	Irritant
Mean %CV	42.3		
Median %CV	13.2		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

Table 7-5 Intralaboratory Repeatability of Viability Values for Replicate Laboratory B, Takahashi et al. 2010[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Polyoxyethylene 23 lauryl ether	106.3	0.4	Non Irritant
Mineral oil	96.0	1.4	Non Irritant
Isopropyl alcohol	94.8	1.8	Non Irritant
3-Methoxy-1,2-propanediol	103.6	2.6	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	99.6	2.9	Non Irritant
Methyl cyclopentane	94.5	3.9	Non Irritant
Physiological saline	94.9	4.2	Non Irritant
Ethanol	95.2	4.3	Non Irritant
Silicic anhydride	72.5	4.4	Non Irritant
Triethanolamine	93.6	4.6	Non Irritant
Toluene	95.6	4.8	Non Irritant
Tween 80	91.5	4.9	Non Irritant
Glycerol	97.8	5.1	Non Irritant
Cyclopentasiloxane	101.5	5.2	Non Irritant
Ethylhexyl salicylate	94.6	5.4	Non Irritant
2-Ethoxyethyl acetate	93.4	5.5	Non Irritant
Methyl amyl ketone	87.7	5.9	Non Irritant
Propylene glycol	98.9	6.0	Non Irritant
3,3-Dimethylpentane	89.1	6.0	Non Irritant
Methyl isobutyl ketone	89.5	6.1	Non Irritant
Polyethylene glycol 400	97.4	6.5	Non Irritant
Dimethyl sulfoxide	99.1	6.9	Non Irritant
Isopropyl myristate	92.5	7.0	Non Irritant
Myristyl alcohol	83.4	7.3	Non Irritant
Polyoxyethylene hydrogenated castor Oil (60E.O.)	91.4	8.4	Non Irritant
2-Ethyl-1-hexanol	69.9	8.9	Irritant
Propylene carbonate	81.1	10.7	Non Irritant
Sodium salicylate	88.1	10.7	Non Irritant
Polyoxyethylene 8 stearate	35.7	13.8	Irritant
Acetone	35.6	14.7	Irritant
Methyl ethyl ketone (2-Butanone)	46.9	15.8	Irritant
Isobutyl alcohol	5.2	18.2	Irritant
Monoethanolamine	1.7	20.6	Irritant
Butanol	6.0	21.2	Irritant
Myristic acid	16.9	23.4	Irritant
Calcium thioglycolate	5.2	25.9	Irritant
Tween 20	29.8	32.1	Irritant
Ethyl acetate	10.2	33.1	Irritant
Acetic acid	3.7	33.2	Irritant
Glycolic acid	4.0	35.4	Irritant

Table 7-5 (continued)

Substance	Mean Viability (n=3)	% CV	Category Classification
Distearyldimethylammonium chloride	19.2	38.0	Irritant
Lactic acid	4.7	44.8	Irritant
Acid red 92	1.4	55.4	Irritant
Benzyl alcohol	3.2	64.1	Irritant
n-Lauroylsarcosine sodium salt (30%)	0.4	66.1	Irritant
m-Phenylenediamine	5.0	69.5	Irritant
2-Benzyloxyethanol	3.6	69.5	Irritant
Butyl cellosolve	1.2	73.7	Irritant
Cyclohexanol	2.6	78.9	Irritant
Sodium 2-naphthalenesulfonate	1.8	85.5	Irritant
Triton X-100	1.2	100.7	Irritant
n-Hexanol	1.0	105.1	Irritant
Polyethyleneglycol monolaurate (10E.O.)	1.3	110.6	Irritant
2-Methyl-1-pentanol	1.6	122.6	Irritant
Sodium hydroxide	0.6	150.1	Irritant
Potassium laurate	0.7	161.5	Irritant
Sodium lauryl sulfate	0.4	173.2	Irritant
Polyoxyethylene 10 lauryl ether	0.3	195.7	Irritant
Diisopropanolamine	0.6	233.3	Irritant
1-Octanol	0.7	270.7	Irritant
Cetyltrimethylammonium bromide	0.6	404.5	Irritant
Benzalkonium chloride	-0.4	NA	Irritant
Cetylpyridinium bromide (10%)	-1.0	NA	Irritant
Cetylpyridinium chloride	-0.6	NA	Irritant
Di(2-Ethylhexyl) sodium sulfosuccinate	-1.7	NA	Irritant
Domiphen bromide	-0.9	NA	Irritant
Polyoxyethylene 5 lauryl ether	-0.8	NA	Irritant
Sodium polyoxyethylene laurylether sulfate (2 E.O.) (27%)	-0.1	NA	Irritant
Stearyltrimethylammonium chloride	-0.5	NA	Irritant
Sucrose fatty acid ester	-3.9	NA	Irritant
Mean %CV	51.0		
Median %CV	15.8		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

The ranges of %CV values for substances classified as non irritants (i.e., viability value > 70) are 0.3 to 14.1 for Laboratory A, and 0.4 to 10.7 for Laboratory B. The ranges of %CV values for substances classified as irritants (i.e., viability value ≤ 70) are 3.6 to 462.2 for Laboratory A, and 8.9 to 404.5 for Laboratory B. The within experiment, the mean and the median %CV values for the two laboratories for all substances ranged from 42.3 and to 51.0 and 13.2 and 15.8 respectively. Substances classified *in vitro* as irritants (i.e., viability value ≤ 70) tended to have greater %CV values. The CVs for seven to eight chemicals in all laboratories were not possible to calculate due to the existence of a negative value (noted in the table as NA; Not applicable).

7.2.1.3 *Sakaguchi et al. (2011)*

Intralaboratory Repeatability:

In this study, 25 substances were evaluated in five laboratories multiple times (three experiments). A %CV value was calculated for the viability value for each test (**Appendix E1**). **Tables 7-6, 7-7, 7-8, 7-9** and **7-10** summarize the mean and the %CV values of the viability for each test conducted in Laboratory 1, Laboratory 2, Laboratory 3, Laboratory 4, and Laboratory 5, respectively. The results for each laboratory are sorted by %CV values from lowest to highest value.

Table 7-6 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 1, Sakaguchi et al. 2011[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Glycerol	90.3	2.2	Non Irritant
Ethanol	99.8	3.0	Non Irritant
Methyl ethyl ketone	71.0	3.2	Non Irritant
3-Methoxy-1,2-propanediol	95.7	3.3	Non Irritant
Toluene	89.3	3.5	Non Irritant
2-Ethylhexyl p-dimethyl-amino benzoate	84.5	5.2	Non Irritant
Gluconolactone	79.1	6.2	Non Irritant
Methyl cyclopentane	97.1	6.7	Non Irritant
Polyethyleneglycol 400	92.1	6.9	Non Irritant
Acetone	80.8	7.1	Non Irritant
3,3-Dimethylpentane	101.2	7.9	Non Irritant
Methyl isobutyl ketone	100.5	8.7	Non Irritant
n,n-Dimethylguanidine sulfate	83.4	9.8	Non Irritant
Propylene glycol	94.4	10.7	Non Irritant
Tween 20	77.0	11.4	Non Irritant
Methyl amylketone	85.1	11.6	Non Irritant
2-Ethyl-1-hexanol	64.3	12.9	Irritant
2-Methyl-1-pentanol	10.6	17.0	Irritant
Sodium hydroxide	4.0	19.9	Irritant
Benzalkonium chloride	3.9	25.1	Irritant
Cyclohexanol	5.3	35.2	Irritant
n-Hexanol	11.1	35.4	Irritant
1-Octanol	7.7	39.5	Irritant
Cetylpyridinium bromide	1.6	39.9	Irritant
Triton X-100	3.9	61.3	Irritant
Mean %CV	15.8		
Median %CV	9.8		

[1]Substances organized by increasing %CV.

Table 7-7 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 2, Sakaguchi et al. 2011[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Ethanol	95.8	0.6	Non Irritant
Gluconolactone	78.7	1.8	Non Irritant
Methyl cyclopentane	102.4	2.2	Non Irritant
Glycerol	103.4	4.5	Non Irritant
n,n-Dimethylguanidine sulfate	99.2	4.5	Non Irritant
Polyethyleneglycol 400	98.0	5.1	Non Irritant
Acetone	96.9	5.3	Non Irritant
3-Methoxy-1,2-propanediol	96.6	5.3	Non Irritant
Methyl amylketone	93.0	5.9	Non Irritant
3,3-Dimethylpentane	89.3	7.4	Non Irritant
Methyl isobutyl ketone	95.9	8.0	Non Irritant
2-Ethyl-1-hexanol	70.2	8.3	Non Irritant
Propylene glycol	96.6	8.5	Non Irritant
Toluene	103.1	8.9	Non Irritant
Tween 20	101.4	10.9	Non Irritant
Methyl ethyl ketone	36.0	12.3	Irritant
2-Ethylhexyl p-dimethyl-amino benzoate	88.7	13.4	Non Irritant
n-Hexanol	1.0	50.5	Irritant
Cyclohexanol	8.8	52.8	Irritant
Benzalkonium chloride	2.8	56.4	Irritant
Triton X-100	2.7	72.8	Irritant
2-Methyl-1-pentanol	2.1	81.6	Irritant
Cetylpyridinium bromide	1.5	86.6	Irritant
1-Octanol	4.4	149.2	Irritant
Sodium hydroxide	0.9	154.3	Irritant
Mean %CV	32.7		
Median %CV	8.5		

[1]Substances organized by increasing %CV.

Table 7-8 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 3, Sakaguchi et al. 2011[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Propylene glycol	96.2	0.9	Non Irritant
Ethanol	99.9	1.4	Non Irritant
Acetone	91.8	1.5	Non Irritant
3,3-Dimethylpentane	95.7	4.5	Non Irritant
Methyl amylketone	63.0	5.5	Irritant
Toluene	83.0	6.9	Non Irritant
Glycerol	104.7	7.1	Non Irritant
n,n-Dimethylguanidine sulfate	92.8	7.2	Non Irritant
Polyethyleneglycol 400	108.1	7.4	Non Irritant
Gluconolactone	73.0	8.3	Non Irritant
Tween 20	82.0	9.0	Non Irritant
Methyl isobutyl ketone	86.5	10.3	Non Irritant
2-Ethylhexyl p-dimethyl-amino benzoate	85.1	10.4	Non Irritant
3-Methoxy-1,2-propanediol	101.0	11.2	Non Irritant
Methyl cyclopentane	91.2	15.8	Non Irritant
Methyl ethyl ketone	55.8	17.5	Irritant
2-Ethyl-1-hexanol	27.4	25.3	Irritant
2-Methyl-1-pentanol	3.3	72.0	Irritant
Benzalkonium chloride	1.4	72.2	Irritant
Cyclohexanol	3.5	75.1	Irritant
n-Hexanol	2.1	87.6	Irritant
1-Octanol	3.0	91.5	Irritant
Sodium hydroxide	0.6	97.9	Irritant
Triton X-100	0.4	173.2	Irritant
Cetylpyridinium bromide	0.0	-	Irritant
Mean %CV	34.2		
Median %CV	10.4		

[1]Substances organized by increasing %CV.

[2]A "-" indicats that a CVs could not be calculated in order that all viability values for this substance were zero.

Table 7-9 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 4, Sakaguchi et al. 2011[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
3-Methoxy-1,2-propanediol	88.7	4.1	Non Irritant
Glycerol	97.2	4.3	Non Irritant
Polyethyleneglycol 400	99.1	4.5	Non Irritant
Propylene glycol	96.0	4.9	Non Irritant
n,n-Dimethylguanidine sulfate	81.9	5.0	Non Irritant
Methyl isobutyl ketone	108.5	5.0	Non Irritant
Methyl amylketone	73.3	5.3	Non Irritant
Acetone	84.6	7.7	Non Irritant
3,3-Dimethylpentane	98.5	8.2	Non Irritant
Tween 20	101.3	8.3	Non Irritant
Methyl cyclopentane	109.2	8.7	Non Irritant
Ethanol	92.7	8.8	Non Irritant
Gluconolactone	74.8	9.4	Non Irritant
2-Ethylhexyl p-dimethyl-amino benzoate	101.1	12.5	Non Irritant
Methyl ethyl ketone	43.2	12.8	Irritant
2-Ethyl-1-hexanol	56.6	12.9	Irritant
1-Octanol	8.3	14.0	Irritant
Toluene	92.8	15.4	Non Irritant
2-Methyl-1-pentanol	9.1	17.6	Irritant
n-Hexanol	9.0	31.8	Irritant
Cyclohexanol	6.1	41.8	Irritant
Benzalkonium chloride	1.7	109.1	Irritant
Cetylpyridinium bromide	0.9	173.2	Irritant
Triton X-100	0.7	173.2	Irritant
Sodium hydroxide	0.3	173.2	Irritant
Mean %CV	34.8		
Median %CV	9.4		

[1]Substances organized by increasing %CV.

Table 7-10 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 5, Sakaguchi et al. 2011[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Methyl cyclopentane	99.3	3.2	Non Irritant
2-Ethyl-1-hexanol	80.4	3.4	Non Irritant
Gluconolactone	65.8	3.6	Irritant
3,3-Dimethylpentane	93.9	3.8	Non Irritant
Ethanol	101.3	4.0	Non Irritant
Polyethyleneglycol 400	96.7	4.2	Non Irritant
Glycerol	99.0	5.0	Non Irritant
3-Methoxy-1,2-propanediol	91.1	6.0	Non Irritant
Acetone	98.6	6.0	Non Irritant
Propylene glycol	100.0	6.5	Non Irritant
Tween 20	94.0	7.7	Non Irritant
Methyl isobutyl ketone	105.1	7.9	Non Irritant
Methyl amylketone	77.8	9.5	Non Irritant
Toluene	96.1	9.8	Non Irritant
Methyl ethyl ketone	51.8	10.8	Irritant
2-Ethylhexyl p-dimethyl-amino benzoate	88.7	12.0	Non Irritant
n,n-Dimethylguanidine sulfate	92.8	12.6	Non Irritant
2-Methyl-1-pentanol	3.7	33.2	Irritant
n-Hexanol	2.0	46.0	Irritant
Triton X-100	0.4	86.7	Irritant
Cyclohexanol	4.0	87.3	Irritant
Sodium hydroxide	0.2	139.3	Irritant
Benzalkonium chloride	0.6	173.2	Irritant
1-Octanol	2.4	173.2	Irritant
Cetylpyridinium bromide	0.0	-	Irritant
Mean %CV	35.6		
Median %CV	8.7		

[1]Substances organized by increasing %CV.

[2]A "-" indicats that a CVs could not be calculated in order that all viability values for this substance were zero.

The ranges of %CV values for substances classified as non irritants (i.e., viability value > 70) are 2.2 to 11.6 for Laboratory 1, 0.6 to 13.4 for Laboratory 2, 0.9 to 15.8 for Laboratory 3, 4.1 to 15.4 for Laboratory 4, and 3.2 to 12.6 for Laboratory 5. The ranges of %CV values for substances classified irritants (i.e., viability value ≤ 70) are 12.9 to 61.3 for Laboratory 1, 12.3 to 154.3 for Laboratory 2, 5.5 to 173.2 for Laboratory 3, 12.8 to 173.2 for Laboratory 4, and 3.6 to 173.2 for Laboratory 5. The within experiment, the mean and the median %CV values for the five laboratories for all substances ranged from 15.8 to 35.6 and 8.5 to 10.4 respectively. Substances classified *in vitro* as irritants (i.e., viability value ≤ 70) tended to have greater %CV values. The CVs for one chemicals in laboratory 1 and laboratory 5 were not possible to calculate in order that all viability values for these chemicals were zero (noted in the table as "-").

7.2.1.4 *Kojima et al. (2012)*

Intralaboratory Repeatability:

In this study, 40 substances were evaluated in two or three laboratories multiple times (three experiments). A%CV value was calculated for the viability value for each test (**Appendix E1**). **Tables 7-11, 7-12,** and **7-13** summarize the mean and the %CV values of the viability for each test conducted in Laboratory 1, Laboratory 2, and Laboratory 3, respectively. The results for each laboratory are sorted by %CV values from lowest to highest value.

Table 7-11 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 1, Kojima et al. 2012[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Methyl acetate	86.4	1.9	Non Irritant
Isopropyl bromide	88.6	4.3	Non Irritant
Ammonium nitrate	102.2	4.4	Non Irritant
Camphene	102.0	5.0	Non Irritant
Myristyl alcohol	104.7	6.0	Non Irritant
2-Benzyloxyethanol	5.1	6.7	Irritant
Hexyl cinnamic aldehyde	101.2	8.8	Non Irritant
Isopropyl myristate	99.1	9.4	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	104.6	10.9	Non Irritant
Cyclohexanone	8.5	12.7	Irritant
Distearyldimethylammonium chloride	33.1	13.3	Irritant
Pyridine	4.6	14.5	Irritant
Cyclopentanol	8.0	16.6	Irritant
Methyl cyanoacetate	51.0	17.0	Irritant
Butanol	6.8	30.1	Irritant
2-Methylbutyric acid	3.4	32.6	Irritant
Calcium thioglycolate	5.4	36.4	Irritant
Isobutanal	3.9	41.8	Irritant
Sodium monochloroacetate	25.8	48.4	Irritant
Promethazine hydrochloride	0.9	62.6	Irritant
m-Phenylenediamine	3.9	68.0	Irritant
Ethyl acetate	16.9	79.5	Irritant
Isobutyl alcohol	2.7	82.2	Irritant
Ethyl 2-methylacetoacetate	1.5	87.7	Irritant
Monoethanolamine	1.2	91.9	Irritant
Di(propylene glycol) propyl ether	5.9	152.8	Irritant
n-Lauroylsarcosine sodium salt	0.5	173.2	Irritant
1-Octanol	0.4	173.2	Irritant
Sodium lauryl sulfate	0.0	NA	Irritant
Imidazole	0.0	NA	Irritant
Mean %CV	46.1		
Median %CV	23.6		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

Table 7-12 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 2, Kojima et al. 2012[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Isopropyl alcohol	88.7	1.1	Non Irritant
Pyridine	9.0	3.3	Irritant
Isopropyl myristate	88.6	3.4	Non Irritant
Potassium sorbate	93.7	3.4	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	95.8	4.7	Non Irritant
2,5-Dimethyl-2,5-hexanediol	85.3	6.0	Non Irritant
Sodium salicylate	94.4	6.9	Non Irritant
Methoxyethyl acrylate	7.7	6.9	Irritant
Methyl cyanoacetate	87.5	8.4	Non Irritant
Camphene	75.1	9.0	Non Irritant
Isopropyl bromide	95.9	10.2	Non Irritant
Butanol	7.4	10.2	Irritant
Calcium thioglycolate	18.5	12.2	Irritant
Butyrolactone	46.0	13.5	Irritant
Distearyldimethylammonium chloride	70.1	14.5	Non Irritant
m-Phenylenediamine	7.5	15.8	Irritant
Ammonium nitrate	76.0	16.5	Non Irritant
Ethyl 2-methylacetoacetate	44.4	17.1	Irritant
Ethyl acetate	17.2	21.7	Irritant
Citric acid	6.5	22.9	Irritant
Cyclopentanol	9.7	23.3	Irritant
n-Butanal	10.4	26.3	Irritant
2-Methylbutyric acid	4.4	31.9	Irritant
1-Octanol	4.7	38.9	Irritant
Propasol solvent P	19.1	61.9	Irritant
Isobutanal	8.8	74.9	Irritant
Imidazole	0.8	116.3	Irritant
Triton X-100 (5%)	0.1	121.4	Irritant
n-Lauroylsarcosine sodium salt	0.3	173.2	Irritant
Sodium lauryl sulfate	0.0	NA	Irritant
Mean %CV	30.2		
Median %CV	14.5		

[1]Substances organized by increasing %CV.

[2]NA=Not applicable (i.e., CVs could not be calculated due to the existence of a negative value).

Table 7-13 Intralaboratory Repeatability of Viability Values for Replicate Laboratory 3, Kojima et al. 2012[1]

Substance	Mean Viability (n=3)	% CV	Category Classification
Sodium salicylate	95.3	0.8	Non Irritant
Potassium sorbate	101.2	1.4	Non Irritant
Isopropyl alcohol	93.4	4.6	Non Irritant
Hexyl cinnamic aldehyde	90.0	7.9	Non Irritant
Myristyl alcohol	75.3	8.2	Non Irritant
Methyl acetate	84.4	8.7	Non Irritant
2,5-Dimethyl-2,5-hexanediol	73.6	9.1	Non Irritant
Butyrolactone	63.2	9.3	Irritant
2-Ethylhexyl p-dimethylamino benzoate	89.5	10.8	Non Irritant
Isopropyl myristate	100.8	12.8	Non Irritant
Isobutanal	5.5	13.5	Irritant
Triton X-100 (5%)	1.8	21.0	Irritant
Cyclohexanone	6.4	22.4	Irritant
Butanol	3.7	28.7	Irritant
m-Phenylenediamine	4.0	30.8	Irritant
Di(propylene glycol) propyl ether	1.2	40.2	Irritant
Methoxyethyl acrylate	4.8	43.6	Irritant
Propasol solvent P	4.4	44.4	Irritant
2-Benzyloxyethanol	3.1	65.0	Irritant
Sodium monochloroacetate	6.7	86.0	Irritant
Promethazine hydrochloride	1.5	86.8	Irritant
1-Octanol	1.7	91.1	Irritant
Monoethanolamine	1.7	92.9	Irritant
n-Butanal	1.0	96.2	Irritant
n-Lauroylsarcosine sodium salt	1.0	97.8	Irritant
Ethyl 2-methylacetoacetate	4.8	102.8	Irritant
Isobutyl alcohol	3.4	107.2	Irritant
Citric acid	3.0	108.6	Irritant
Imidazole	0.6	110.5	Irritant
2-Methylbutyric acid	0.6	166.2	Irritant
Mean %CV	51.0		
Median %CV	35.5		

[1]Substances organized by increasing %CV.

The ranges of %CV values for substances classified as non irritants (i.e., viability value > 70) are 1.9 to 10.9 for Laboratory 1, 1.1 to 16.5 for Laboratory 2, and 0.8 to 12.8 for Laboratory 3. The ranges of %CV values for substances classified as irritants (i.e., viability value ≤ 70) are 6.7 to 173.2 for Laboratory 1, 3.3 to 173.2 for Laboratory 2, and 9.3 to 166.2 for Laboratory 3. The within experiment, the mean and the median %CV values for the three laboratories for all substances ranged from 30.2 to 51.0 and 14.5 to 35.5 respectively. Substances classified *in vitro* as irritants (i.e., viability value ≤ 70) tended to have greater %CV values. The CVs for one and two chemicals in laboratory 1 and 2 were not

possible to calculate due to the existence of a negative value (noted in the table as NA; Not applicable).

7.2.2 Evaluation of Interlaboratory Reproducibility

All of four studies discussed in Section 6.0 included interlaboratory data for at least a subset of the substances evaluated. The ability of the STE test to reproducibly identify ocular irritants versus non irritants was evaluated using two approaches.

In the first approach, a qualitative assessment of reproducibility was conducted. In this evaluation, the ocular irritation category classification for each substance obtained from individual laboratory was used to evaluate the extent of agreement among the participating laboratories in their ability to identify ocular irritants versus non irritants. The reliability of STE was assessed separately for each study reviewed in Sections 4.0 and 5.0. Substances classified, based on STE data, as irritants or non irritants were further classified according to the GHS (UN 2003) and the EPA (EPA 1996), and ocular irritancy classification systems. If the detailed *in vivo* data was not available from ECETOC reference chemicals data bank (ECETOC 1998), the ocular irritancy classification for GHS and EPA described either in the publicly-available documents or related literatures were adopted without any modification. Because the focus of this reliability assessment is on the interlaboratory reproducibility of STE in identifying irritants versus non irritants, considerable variability could exist among laboratories in their classification of substances as irritants or non irritants that would not be apparent from this analysis. Therefore, in the second approach, a quantitative assessment of reproducibility was determined by calculating the CV for test substance data. The reproducibility of STE was assessed for the studies reviewed in Sections 4.0 and 5.0 where individual testing laboratory data were available. When there was any test substances that had no *in vivo* data (e.g., individual irritation score of cornea, iris, and conjunctiva at observation period up to 21 days) and could not determinate the classifications of GHS or EPA these substances were excluded out of the interlaboratory Reproducibility analysis.

7.2.2.1 *Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System*

Reliability analyses for the STE test were evaluated for the following four studies: Takahashi et al. 2009, Takahashi et al. 2010, Sakaguchi et al. 2011, and Kojima et al. 2012. The agreement of classification calls among participating laboratories and the relationship to the *in vivo* classification (GHS; UN 2003) for the substances tested in each validation in each study is provided in **Table 7-14**.

Table 7-14 Evaluation of the Reliability of STE test in Predicting Ocular Irritants as Defined by the GHS Classification System, by Study

Report	Classification (*In Vivo*/*In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 80% Agreement among Labs	Substances with 67-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
Takahashi et al. 2009	I/I	3	24	24 (100%)			
	I/NI	3	4	4 (100%)			
	NI/I	3	1	1 (100%)			
	NI/NI	3	11	11 (100%)			
	Total		40	40 (100%)			
Takahashi et al. 2010	I/I	2	30	30 (100%)			
	I/NI	2	5	5 (100%)			
	NI/I	2	4	4 (100%)			
	NI/NI	2	16	16 (100%)			
	Total		55	55 (100%)			
Sakaguchi et al. 2011	I/I	5	10	8 (80%)	1 (10%)	1 (10%)	
	I/NI	5	3	3 (100%)			
	NI/I	5	0				
	NI/NI	5	11	10 (91%)	1 (9%)		
	Total		24	20 (83%)	2 (8%)	1 (4%)	
Kojima et al. 2012	I/I	2	16	14 (88%)			2 (13%)
		3	7	7 (100%)			
	I/NI	2	7	7 (100%)			
		3	0				
	NI/I	2	2	2 (100%)			
		3	0				
	NI/NI	2	1	1 (100%)			
		3	2	2 (100%)			
	Total		35	33 (94%)			2 (6%)

[1]A "I" for in vivo indicates that the substance was assigned an overall classification of irritant (Category 1, 2A, and 2B); a "NI" for in vivo indicates that classification of non irritant (Not classified);
[2]n indicates number of substances.
[3]Number in parentheses indicates percentage of tested chemicals.

For the study by Takahashi et al. (2009), there was 100% agreement in regard to the ocular irritancy classification for 40 (100%) of the 40 substances, which were tested in three laboratories. No discordance in the classification results was presented.

For the study by Takahashi et al. (2010), there was also 100% agreement in regard to the ocular irritancy classification for 55 (100%) of the 55 substances, which were tested in two laboratories.

For the study by Sakaguchi et al. (2011), there was 100% agreement in regard to the ocular irritancy classification for 20 (83%) of the 24 substances, which were tested in five laboratories. Discordant results were observed for substances that were correctly identified as irritant or non irritant.

For the study by Kojima et al. (2012), there was 100% agreement in regard to the ocular irritancy classification for 33 (94%) of the 35 substances, which were tested in two or three laboratories. Discordance in the classification results were presented for two substances that was classified as irritant in one laboratory and as non irritant in another laboratory.

7.2.2.2 *Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System*

Reliability analyses for the STE test were evaluated for the following four studies: Takahashi et al. 2009, Takahashi et al. 2010, Sakaguchi et al. 2011, and Kojima et al. 2012. The agreement of classification calls among participating laboratories and the relationship to the *in vivo* classification (EPA 1996) for the substances tested in each validation in each study is provided in **Table 7-15**.

Table 7-15 Evaluation of the Reliability of STE test in Predicting Ocular Irritants as Defined by the EPA Classification System, by Study

Report	Classification (*In Vivo*/*In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 80% Agreement among Labs	Substances with 67-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
Takahashi et al. 2009	I/I	3	24	24 (100%)			
	I/NI	3	8	8 (100%)			
	NI/I	3	0				
	NI/NI	3	7	7 (100%)			
	Total		39	39 (100%)			
Takahashi et al. 2010	I/I	2	32	32 (100%)			
	I/NI	2	12	12 (100%)			
	NI/I	2	1	1 (100%)			
	NI/NI	2	9	9 (100%)			
	Total		54	54 (100%)			
Sakaguchi et al. 2011	I/I	5	9	7 (78%)	1 (11%)	1 (11%)	
	I/NI	5	8	7 (88%)	1 (13%)		
	NI/I	5	0				
	NI/NI	5	6	6 (100%)			
	Total		23	20 (87%)	2 (9%)	1 (4%)	
Kojima et al. 2012	I/I	2	19	17 (89%)			2 (11%)
		3	6	6 (100%)			
	I/NI	2	7	6 (100%)			
		3	0				
	NI/I	2	0				
		3	0				
	NI/NI	2	1	1 (100%)			
		3	2	2 (100%)			
	total		35	33 (94%)			2 (6%)

[1]A "I" for in vivo indicates that the substance was assigned an overall classification of irritant (Category I, II, and III); a "NI" for in vivo indicates that classification of non irritant (Category IV);
[2]n indicates number of substances.
[3]Number in parentheses indicates percentage of tested chemicals.

For the study by Takahashi et al. (2009), there was 100% agreement in regard to the ocular irritancy classification for 39 (100%) of the 39 substances, which were tested in three laboratories. No discordance in the classification results was presented.

For the study by Takahashi et al. (2010), there was also 100% agreement in regard to the ocular irritancy classification for 54 (100%) of the 54 substances, which were tested in two laboratories.

For the study by Sakaguchi et al. (2011), there was 100% agreement in regard to the ocular irritancy classification for 20 (87%) of the 23 substances, which were tested in five laboratories. Discordant results were observed for substances that were correctly identified as irritant or non irritant nonirritant, as well as for false negatives.

For the study by Kojima et al. (2012), there was 100% agreement in regard to the ocular irritancy classification for 33 (94%) of the 35 substances, which were tested in two or three laboratories. Discordance in the classification results were presented for two substances that was classified as irritant in one laboratory and as non irritant in another laboratory (could not be identified due to the even number of different ocular irritation category for test substance). Discordance in the classification results was also presented for one substance of which EPA classification of this substance could not be made due to the lack of appropriate *in vivo* data.

7.2.2.3 *Common Chemical Classes Among Test Substances with Discordant Interlaboratory Results*

For the Sakaguchi et al. (2011) study, four substances showed interlaboratory differences in *in vitro* classification (**Table 7-16**). Of these, three (75%) are ketones/lactones, and one is alcohol. Of the three liquid substances that produced discordant interlaboratory results in this study, two are ketones.

For the Kojima et al. (2012) study, two substances showed interlaboratory differences in *in vitro* classification (**Table 7-17**). Of these, one is surfactant (cationic), and another one is ester (acetate) compound.

Table 7-16 Chemical Classes of Test Substances with Discordant Interlaboratory Results in the Sakaguchi et al. (2011) Study

Substance	Chemical Class	Physical Form	In Vitro Classification (% of Labs with Classification)
Methyl ethyl ketone (2-Butanone)	Ketones	Liquid	Irritant (4/5; 80%) Non Irritant (1/5; 20%)
2-Ethyl-1-hexanol	Alcohols	Liquid	Irritant (3/5; 60%) Non Irritant (2/5; 40%)
Gluconolactone	Lactones	Solid	Irritant (1/5; 20%) Non Irritant (4/5; 80%)
Methyl amyl ketone	Ketones	Liquid	Irritant (1/5; 20%) Non Irritant (4/5; 80%)

Table 7-17 Chemical Classes of Test Substances with Discordant Interlaboratory Results in the Kojima et al. (2012) Study

Substance	Chemical Class	Physical Form	In Vitro Classification (% of Labs with Classification)
Distearyldimethylammonium chloride	Surfactants (cationic)	Solid	Irritant (1/2; 50%) Non Irritant (1/2; 50%)
Methyl cyanoacetate	Esters (acetate)	Liquid	Irritant (1/2; 50%) Non Irritant (1/2; 50%)

7.2.2.4 *Interlaboratory Reproducibility Based on Coefficient of Variation Analysis of Viability values*

To provide a quantitative assessment of interlaboratory variability, individual laboratory STE test results were used to calculate a mean and CV for the viability value for each substance tested in Takahashi et al. 2009, Sakaguchi et al. 2011, and Kojima et al. 2012 (**Tables 7-18, 7-19, 7-20**).

For the Takahashi et al. (2009) study, a wide range of %CV values for individual substances is evident for the viability value (**Table 7-18**). The mean and median %CV values were 56.7% and 11.2%, respectively, ranging from 0.3% to 300.7% for the entire set of 44 test substances. Substances classified as irritants (i.e., viability value ≤70) tended to have much greater %CV values.

For the Sakaguchi et al. (2011) study, a wide range of %CV values for individual substances is evident for the viability value (**Table 7-19**). The mean and median %CV values were

32.3% and 8.6%, respectively, ranging from 2.1% to 131.9% for the entire set of 25 test substances. Substances classified as irritants (i.e., viability value ≤ 70) tended to have much greater %CV values.

For the Kojima et al. (2012) study, only 10 out of 40 test substances (with a balanced GHS ranking) were evaluated for three laboratories. Therefore, the %CV was calculated from the viability values of these 10 test substances only (**Table 7-20**). The mean and median %CV values were 58.8% and 51.4%, respectively, ranging from 6.9% to 141.1% for the entire set of 10 test substances. Substances classified as irritants (i.e., viability value ≤ 70) tended to have much greater %CV values.

Table 7-18 Coefficient of Validation Analysis of the Interlaboratory Variability of the STE test for Takahashi et al. (2009)[1]

Substance	Viability	No. of Labs	%CV	Consensus Classification
Polyethylene glycol 400	97.8	3	0.3	Non Irritant
Ethanol	97.4	3	1.6	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	101.7	3	2.2	Non Irritant
Sodium salicylate	87.4	3	2.5	Non Irritant
Toluene	97.8	3	2.6	Non Irritant
Methyl amyl ketone	88.3	3	2.9	Non Irritant
Methyl cyclopentane	98.3	3	3.2	Non Irritant
3,3-Dimethylpentane	93.0	3	3.7	Non Irritant
Lactic acid	4.5	3	3.8	Irritant
Glycerol	101.7	3	4.4	Non Irritant
Methyl isobutyl ketone	92.1	3	4.7	Non Irritant
Benzyl alcohol	3.4	3	4.8	Irritant
Silicic anhydride	77.6	3	5.0	Non Irritant
Isopropyl myristate	99.5	3	5.4	Non Irritant
3-Methoxy-1,2-propanediol	102.5	3	6.1	Non Irritant
Isopropyl alcohol	100.9	3	6.3	Non Irritant
Glycolic acid	3.7	3	6.7	Irritant
Triethanolamine	97.9	3	8.5	Non Irritant
Acetic acid	4.0	3	9.5	Irritant
2-Methyl-1-pentanol	1.8	3	11.2	Irritant
Butanol	7.1	3	13.8	Irritant
Polyoxyethylene hydrogenated castor Oil (60E.O.)	99.2	3	17.0	Non Irritant
2-Ethyl-1-hexanol	54.0	3	21.1	Irritant
2-Benzyloxyethanol	2.8	3	23.3	Irritant
m-Phenylenediamine	7.6	3	35.7	Irritant
Acetone	19.6	3	58.3	Irritant
Tween 20	18.1	3	60.2	Irritant
Sodium hydroxide	0.7	3	68.0	Irritant
Cyclohexanol	1.4	3	70.0	Irritant
Benzalkonium chloride	1.1	3	96.7	Irritant
Domiphen bromide	2.2	3	99.8	Irritant
n-Hexanol	0.8	3	103.8	Irritant
1-Octanol	1.6	3	136.1	Irritant
Potassium laurate	0.4	3	148.3	Irritant
Cetyltrimethylammonium bromide	0.2	3	165.4	Irritant
Acid red 92	0.4	3	199.6	Irritant
Triton X-100	0.3	3	247.5	Irritant
Di(2-Ethylhexyl) sodium sulfosuccinate	0.7	3	252.3	Irritant
Monoethanolamine	0.4	3	300.7	Irritant
Diisopropanolamine	0.0	3	NA	Irritant
Stearyltrimethylammonium chloride	-0.3	3	NA	Irritant
Cetylpyridinium bromide	-0.7	3	NA	Irritant
Cetylpyridinium chloride	-0.7	3	NA	Irritant
Sucrose fatty acid ester	-2.3	3	NA	Irritant
Mean %CV	56.7			
Median %CV	11.2			

[1]Substances organized by increasing %CV.

Table 7-19 Coefficient of Validation Analysis of the Interlaboratory Variability of the STE test for Sakaguchi et al. (2011) [1]

Substance	Viability	No. of Labs	%CV	Consensus Classification
Propylene glycol	96.6	5	2.1	Non Irritant
Ethanol	97.9	5	3.6	Non Irritant
3,3-Dimethylpentane	95.7	5	4.7	Non Irritant
3-Methoxy-1,2-propanediol	94.6	5	5.1	Non Irritant
Glycerol	98.9	5	5.8	Non Irritant
Polyethylene glycol 400	98.8	5	5.9	Non Irritant
Methylcyclopentane	99.8	5	6.7	Non Irritant
Gluconolactone	74.3	5	7.3	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	89.6	5	7.5	Non Irritant
n,n-Dimethylguanidine sulfate	90.0	5	8.0	Non Irritant
Toluene	92.9	5	8.1	Non Irritant
Acetone	90.6	5	8.5	Non Irritant
Methyl isobutyl ketone	99.3	5	8.6	Non Irritant
Tween 20	91.1	5	12.3	Non Irritant
Methyl amyl ketone	78.4	5	14.5	Non Irritant
Methyl ethyl ketone (2-Butanone)	51.6	5	25.8	Irritant
2-Ethyl-1-hexanol	59.8	5	33.6	Irritant
Cyclohexanol	5.5	5	38.1	Irritant
1-Octanol	5.2	5	52.3	Irritant
Benzalkonium chloride	2.1	5	62.2	Irritant
2-Methyl-1-pentanol	5.8	5	66.0	Irritant
n-Hexanol	5.0	5	92.4	Irritant
Cetylpyridinium bromide	0.8	5	96.8	Irritant
Triton X-100	1.6	5	99.4	Irritant
Sodium hydroxide	1.2	5	131.9	Irritant
Mean %CV	32.3			
Median %CV	8.6			

[1]Substances organized by increasing %CV.

Table 7-20 Coefficient of Validation Analysis of the Interlaboratory Variability of the STE test for Kojima et al. (2012) [1]

Substance	Viability	No. of Labs	%CV	Consensus Classification
Isopropyl myristate	96.2	3	6.9	Non Irritant
2-Ethylhexyl p-dimethylamino benzoate	96.7	3	7.9	Non Irritant
Butanol	6.0	3	32.9	Irritant
m-Phenylenediamine	5.2	3	39.9	Irritant
Isobutanal	6.1	3	40.7	Irritant
n-Lauroylsarcosine sodium salt	0.6	3	62.2	Irritant
2-Methylbutyric acid	2.8	3	69.7	Irritant
Imidazole	0.5	3	88.3	Irritant
1-Octanol	2.2	3	98.9	Irritant
Ethyl 2-methylacetoacetate	16.9	3	141.1	Irritant
Mean %CV	58.8			
Median %CV	51.4			

[1]Substances organized by increasing %CV.

[1]%CV was only calculated from the viability values of 10 test substances that were evaluated for three laboratories.

7.3 Historical Positive and Negative Control Data

Historical data for positive controls and negative controls were shown in **Table 7-21 and Table 7-22**.

Table 7-21 Historical Positive Control Data for STE test

Positive Control	Viability
Sodium lauryl sulfate (0.01%)	
Mean (n=71)	41.7
SD	10.3
CV	24.7%
Upper and lower limits	21.1 - 62.3

Table 7-22 Historical Negative Control Data for STE test

Negative Control	OD570
Medium (Eagle's MEM suppl. with 10% FBS)	
Mean (n=174)	0.707
SD	0.089
CV	12.6%
Lower limits	0.3

8.0 STE Test Data Quality

8.1 Adherence to National and International GLP Guidelines

As described in Sections 5.5, No report that identified following GLP guidelines or used data obtained according to GLP guidelines existed.

However, two experiments (Sakaguchi et al., 2011 and Kojima et al., 2012) were performed in the spirit of GLP compliance. To guarantee data quality, the following considerations were applied. To start the experiments, a total of seven record sheets were prepared for recording the necessary items. Information regarding adjustment/operation confirmation of instruments, culture media (EMEM supplemented with 10% FBS) preparation, vehicle selection for the test substances, use of reagent/test substance, preparation of test sample, preparation of 0.04 N HCl-isopropanol, passage of SIRC cells, and plate seeding was recorded on the record sheets under the authority of the experiment personnel and the person responsible for the experiment. These records were stored at each laboratory. To ensure measured values were appropriately recorded in the data sheet prepared for this study, a data audit was performed, and consistency between the values recorded in the printout from the plate reader and values inputted into the data sheet was confirmed.

8.2 Impact of Deviations from GLP Guidelines

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

8.3 Availability of Laboratory Notebooks or Other Records

All study notebooks and other supporting records are available, upon request, for an external audit, for the following studies: Takahashi et al. (2009), Takahashi et al. (2010), Sakaguchi et al. (2011), Kojima et al. (2012), and 23 additional data.

9.0 Other Scientific Reports and Reviews

9.1 Reports in the Peer Reviewed Literature

9.1.1 Takahashi et al. (2008)

In order to confirm the usefulness of this STE test in assessing ocular irritation potential of chemicals, 51 raw materials were tested and the correlation between CV in the STE test and the ocular irritation score in the Draize test was examined. For the neat raw materials tested in the Draize test, the 5% test concentration in the STE test gave irritation classes that correlated well with the irritation classes from the Draize test (accuracy: 89.6%). For those materials tested as a 10% solution in the Draize test, STE irritation classes with 0.05% test concentration corresponded well with the Draize irritation classes (accuracy: 80.0%). Moreover, using the cell viabilities at these 2 concentrations, the STE prediction model (PM) was developed. A score of 1 or 2 was given for the results from each tested concentration in the STE test and Draize test. The scores from each test were then summed to yield a 3-level (Rank 1: Minimally irritant, Rank 2: Moderate irritant, Rank3: Severe irritant) eye irritation potential classification. Rank classification in the STE test showed a good correlation mostly to that in the Draize test (irritation class correspondence rate: 70.2%, but after exclusion of data of alcoholic materials, the rate was 91.7%). The Figure and Tables excerpted from this article are shown below.

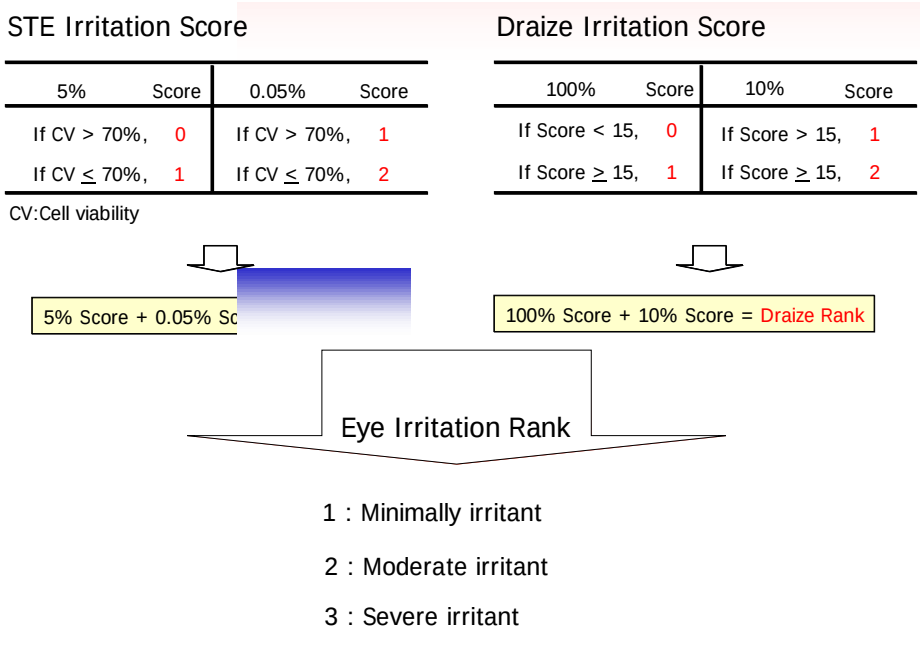

Figure 9-1 Ranking of Test Chemicals for Eye Irritation Using the STE test and Draize Test (reported by Takahashi et al., 2008)

Table 9-1 Summary of STE Test Results and Draize Eye Irritation Test Data for 51 Materials (reported by Takahashi et al., 2008)

Test Chemicals	Draize Score[1)]			STE : Cell Viability(%)			
	100%	10%	Rank	Rank	5%	0.50%	0.05%
Polyethylene gycol monolaurate (10E.O.)	ND	3.3 (NI)	-	2	0.3 (I)	4.6 (I)	101.1 (NI)
Sodium N-lauroyl sarcosinate (30%)	ND	10.3 (NI)	-	2	-1.2 (I)	-1.0 (I)	83.5 (NI)
Sodium polyoxyethylene laurylether sulfate (2E.O.) (27%)	ND	10.0 (NI)	-	3	0.2 (I)	0.8 (I)	53.9 (I)
Domiphen bromide	EC (I)	96.3 (I)	3	3	3.6 (I)	6.8 (I)	4.0 (I)
Cetylpyridinium chloride	EC (I)	94.7 (I)	3	3	-0.1 (I)	0.5 (I)	4.2 (I)
Stearyltrimethylammonium chloride	EC (I)	91.3 (I)	3	3	1.1 (I)	5.3 (I)	7.3 (I)
Benzalkonium chloride	EC (I)	78.0 (I)	3	3	2.1 (I)	5.3 (I)	3.1 (I)
Cetyltrimethylammonium bromide	EC (I)	76.7 (I)	3	3	0.4 (I)	2.5 (I)	2.5 (I)
Di(2-Ethylhexyl) sodium sulfosuccinate *	EC (I)	57.0 (I)	3	3	2.5 (I)	1.5 (I)	2.7 (I)
Triton X-100	EC (I)	41.3 (I)	3	3	-0.2 (I)	-0.7 (I)	0.7 (I)
Potassium laurate	EC (I)	38.0 (I)	3	3	0.8 (I)	1.1 (I)	2.0 (I)
Sodium hydroxide	EC(I) 25.8[2)] (1%Data)		3	3	1.4 (I)	0.6 (I)	-1.4 (I)
Acid red 92	71.0 (I)	25 (I)	3	3	-0.1 (I)	0.4 (I)	12.8 (I)
Sodium lauryl sulfate	EC (I)	15 (I)	3	3	-0.8 (I)	-0.7 (I)	-0.8 (I)
Butanol *	60.8[2)] (I)	34.0 (I)	3	2	8.4 (I)	99.0 (NI)	94.8 (NI)
Chlorhexidine gluconate solution (20%) *	EC (I)	28.3 (I)	3	2	EC (I) 28.9 (0.2%data)		105.2 (NI)
Benzyl alcohol *	31.0 (I)	23.0 (I)	3	2	3.4 (I)	96.3 (NI)	101.6 (NI)
2-Benzyloxyethanol	32.7[3)] (I)	8.7[3)] (NI)	2	2	2.0 (I)	99.1 (NI)	101.3 (NI)
Calcium thioglycollate	79.7 (I)	4.0 (NI)	2	2	4.3 (I)	68.3 (I)	109.8 (NI)
Sucrose fatty acid ester *	28.3 (I)	11.0 (NI)	2	2	-0.6 (I)	6.0 (I)	103.3 (NI)
Isopropylalcohol	30.0[4)] (I)	1.0[4)] (NI)	2	1	101.6 (NI)	100.7 (NI)	97.6 (NI)
Sodium salicylate	83.7 (I)	0.0 (NI)	2	1	89.7 (NI)	101.1 (NI)	100.7 (NI)
Ethanol	32.7 (I)	0.0 (NI)	2	1	98.2 (NI)	108.2 (NI)	97.1 (NI)
Tween 20	4.0 (NI)	0.7 (NI)	1	2	54.7 (I)	92.1 (NI)	97.2 (NI)
Isopropyl myristate *	0.0 (NI)	0.7 (NI)	1	1	105.7 (NI)	108.2 (NI)	99.8 (NI)
Polyoxyethylene hydrogenated castorOil (60E.O.)	13.3 (NI)	0.0 (NI)	1	1	83.5 (NI)	95.7 (NI)	97.5 (NI)
Tween 80	4.7 (NI)	0.0 (NI)	1	1	96.9 (NI)	96.4 (NI)	104.6 (NI)
Glycerin	1.7 (NI)	0.0 (NI)	1	1	99.3 (NI)	100.2 (NI)	100.2 (NI)
Polyethylene glycol 400	4.0 (NI)	0.0 (NI)	1	1	97.8 (NI)	101.5 (NI)	85.9 (NI)
2-Ethylhexyl p-dimethylamino benzoate *	0.0 (NI)	0.0 (NI)	1	1	104.8 (NI)	104.3 (NI)	100.0 (NI)
3-Methoxy-1,2-propanediol	0.0[2)] (NI)	EC (NI)	1	1	109.5 (NI)	109.2 (NI)	98.1 (NI)
Silicic anhydride	2.7 (NI)	EC (NI)	1	1	81.9 (NI)	100.1 (NI)	104.7 (NI)
Physiological saline	0.0 (NI)	EC (NI)	1	1	92.0 (NI)	92.0 (NI)	92.0 (NI)
Acetic acid	EC (I)	68.0 (I)	3	2	4.5 (I)	52.1 (I)	84.1 (NI)
Glycolic acid	EC (I)	25.0 (I)	3	2	3.4 (I)	4.1 (I)	85.8 (NI)
Monoethanolamine	EC (I)	23.3 (I)	3	2	-1.0 (I)	-0.7 (I)	93.7 (NI)
Diisopropanolamine	EC (I)	23.0 (I)	3	2	-1.1 (I)	87.3 (NI)	101.6 (NI)
Lactic acid	102.7 (I)	9.7 (NI)	2	2	4.4 (I)	4.3 (I)	87.3 (NI)
***m*-Phenylenediamine**	80.7 (I)	4.3 (NI)	2	2	6.4 (I)	94.1 (NI)	96.7 (NI)
Triethanolamine	8.0 (NI)	0.0 (NI)	1	1	90.6 (NI)	99.0 (NI)	99.9 (NI)
1-Octanol **	41.0[2)] (I)	ND	-	2	-0.5 (I)	94.2 (NI)	101.4 (NI)
n-Hexanol **	64.8[2)] (I)	46.0[2)] (I)	3	2	-0.3 (I)	77.7 (NI)	100.6 (NI)
2-Ethyl-1-hexanol **	51.3[2)] (I)	39.0[2)] (I)	3	2	44.0 (I)	96.1 (NI)	99.9 (NI)
Cyclohexanol **	79.8[2)] (I)	4.0[2)] (NI)	2	2	1.4 (I)	66.8 (I)	94.3 (NI)
Acetone **	65.8[2)] (I)	3.0[2)] (NI)	2	2	9.6 (I)	102.3 (NI)	101.2 (NI)
2-Methyl-1-pentanol **	13.0[2)] (NI)	EC (NI)	1	2	1.8 (I)	85.3 (NI)	95.5 (NI)
3,3-Dimethylpentane **	0.0[2)] (NI)	EC (NI)	1	1	92.6 (NI)	102.5 (NI)	99.4 (NI)
Methyl cyclopentane **	3.7[2)] (NI)	EC (NI)	1	1	102.2 (NI)	99.0 (NI)	97.6 (NI)
Methyl iso-butyl ketone **	4.8[2)] (NI)	EC (NI)	1	1	88.5 (NI)	107.4 (NI)	102.7 (NI)
Toluene **	9.0[2)] (NI)	EC (NI)	1	1	101.3 (NI)	97.2 (NI)	99.3 (NI)
Methyl amylketone **	13.4[#2)] (NI)	11.0[2)] (NI)	1	1	91.7 (NI)	95.1 (NI)	95.5 (NI)

ND: No Data. EC: Estimated Classification (Prediction class according to another concentration.). (I): Irritant. (NI): Non-Irritant.

*: When dissolving of the raw material was difficult, DMSO was added at 5%.

**: Mineral oil used as vehicle.

Table 9-2 Accuracy of Eye Irritation Classification between STE Test and Draize Test (reported by Takahashi et al., 2008)

		STE		
		5%	0.5%	0.05%
Draize	100%	89.6%	75.0%	58.3%
	10%	76.0%	76.0%	80.0%

Table 9-3 Accuracy of Rank Classification between STE Test and Draize Test (reported by Takahashi et al., 2008)

		Draize Rank Level		
		1	2	3
STE Rank Level	1	15	3[1)]	0
	2	2[2)]	7	9[3)]
	3	0	0	11

1) Isopropylalcohol, Ethanol, Sodium salicylate
2) Tween 20, 2-Methyl-1-pentanol
3) Butanol, Benzyl alcohol, Cholorhexidine, Acetic acid, Glycolic acid, Monoethanolamine, Diisopropanolamine, n-Hexanol, 2-Ethyl-1-hexan

Rank Levels:
1: Minimally irritant
2: Moderate irritant
3: Severe irritant

9.1.2 Takahashi et al. (2011)

To assess intra-laboratory reproducibility, medium control, three vehicles (saline, saline containing 5% (w/w) dimethyl sulfoxide, mineral oil) and three standard chemicals (sodium lauryl sulfate, calcium thioglycolate, tween 80) were evaluated. Assessments were repeated 30 times for vehicles and 18 times for standard chemicals; resulting in almost the same cell viability and a low coefficient of variation value. In addition, the STE ocular irritation rankings of 3 standard chemicals, as calculated on the cell viabilities in 5% and 0.05% solutions were in agreement in all tests. In addition, the irritation category (irritant and non-irritant) was evaluated for 109 chemicals with STE test, GHS classification, and EU classification. The results of the evaluation found the accuracy with each method to be equal to or higher than 86%. The correspondence of STE rankings (1,2, and 3) based on the prediction model by STE test with the eye irritation rankings by GHS (non-irritant, Category 2, and Category 1)

and EU (non-irritant, R36, and R41) was 73% and 71%, respectively. The Figure and Tables excerpted from this article are shown below.

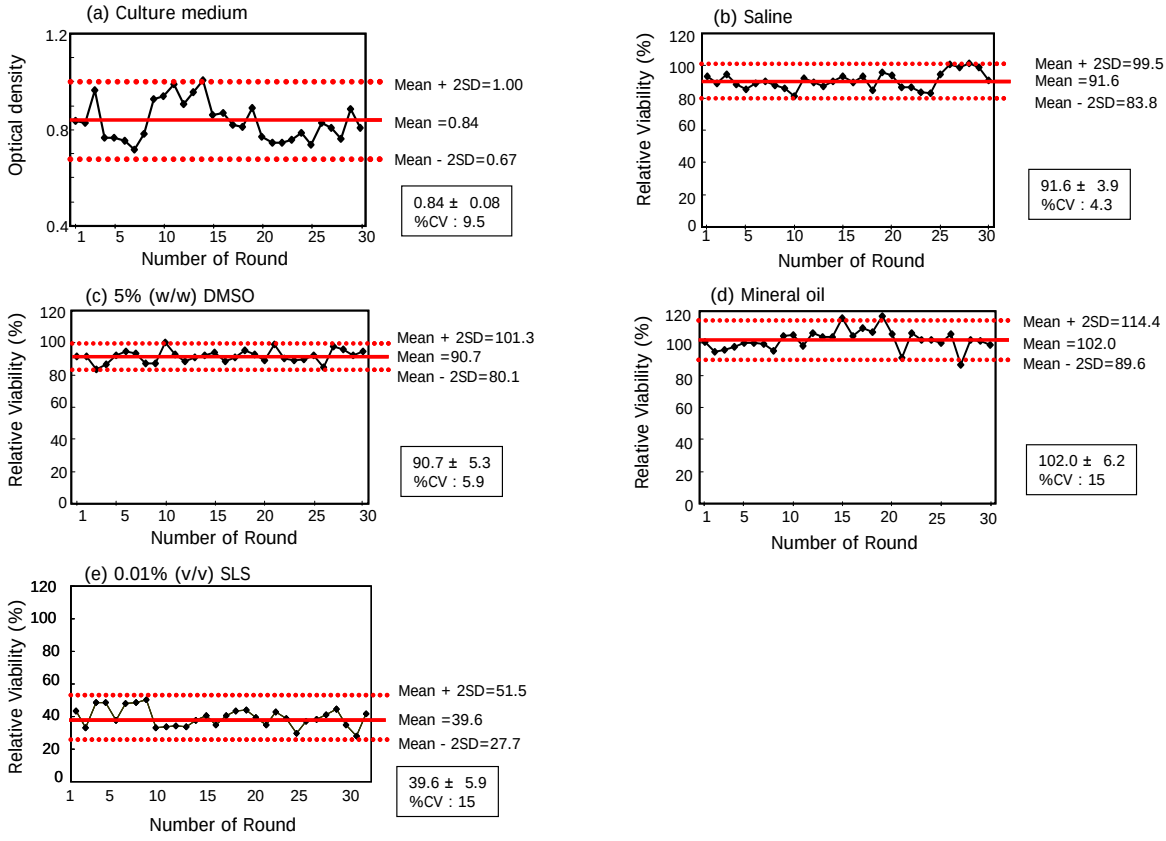

Figure 9-2 Intra-laboratory reproducibility of medium control, three vehicles and positive control (reported by Takahashi et al., 2011)

Table 9-4 Data for three standard chemicals (reported by Takahashi et al., 2011)

Round	Sodium Lauryl Sulfate			Calcium Thioglycolate			Tween 80		
	cell viability (%)			cell viability (%)			cell viability (%)		
	5%	0.05%	rank	5%	0.05%	rank	5%	0.05%	rank
1	0.3	0.4	3	8.9	101.6	2	102.2	100.4	1
2	0.3	0.9	3	12.4	109.5	2	97.8	89.3	1
3	-1.1	0.9	3	12.5	98.0	2	107.8	104.1	1
4	0.7	0.6	3	9.6	106.7	2	100.1	98.3	1
5	-0.5	-1.0	3	12.3	104.4	2	90.8	90.7	1
6	0.3	0.3	3	10.4	100.0	2	112.5	99.6	1
7	0.2	0.1	3	7.8	89.5	2	103.4	109.7	1
8	-1.4	0.3	3	16.7	93.4	2	95.8	86.9	1
9	-0.5	0.5	3	11.3	99.1	2	91.7	103.2	1
10	0.2	0.5	3	13.2	105.8	2	111.8	97.6	1
11	-0.7	-1.0	3	9.3	105.4	2	98.1	102.4	1
12	0.1	0.7	3	9.6	104.8	2	115.1	102.2	1
13	0.0	0.4	3	11.5	105.5	2	96.4	99.6	1
14	-0.6	-0.4	3	9.7	97.4	2	92.3	99.6	1
15	0.7	0.9	3	10.4	101.9	2	95.4	93.5	1
16	0.9	0.1	3	9.1	94.6	2	93.2	101.5	1
17	-0.7	-1.0	3	11.6	100.2	2	109.2	100.3	1
18	-1.0	-1.1	3	7.8	101.2	2	110.1	98.5	1
Average	-0.2	0.1		10.8	101.1		101.3	98.7	
SD	0.7	0.7		2.2	5.2		8.0	5.6	
%CV	NA	601.0		20.2	5.1		7.9	5.7	

Table 9-5 Correlation of STE and GHS eye irritation classification (reported by Takahashi et al., 2011)

a. Contingency table for STE irritation category versus GHS

		STE		
		I	NI	sum
GHS	I	48	8	56
	NI	6	47	53
	sum	54	55	109

b. Outcome (%)

Sensitivity	86	(46/56)
Specificity	89	(47/53)
Positive predictivity	89	(48/54)
Negative predictivity	85	(47/55)
Accuracy	87	(95/109)
False negatives rate	14	(8/56)
False positives rate	11	(6/53)

Table 9-6 Correlation of STE and GHS or EU eye irritation ranking (reported by Takahashi et al., 2011)

a. STE versus GHS eye irritation ranking

		STE rank			
		3	2	1	sum
GHS rank	Category 1	12	13	1	26
	Category 2	2	21	7	30
	NI	0	6	47	53
	sum	14	40	55	109
Accuracy : 73% (80/109)					

b. STE versus EU eye irritation ranking

		STE Rank			
		3	2	1	sum
EU class	R41	8	9	1	18
	R36	2	14	5	21
	NI	1	10	48	59
	sum	11	33	54	98
Accuracy : 71% (70/98)					

Table 9-7 Correlation of STE and EU eye irritation classification (reported by Takahashi et al., 2011)

a. Contingency table for STE irritation category versus EU

		STE		
		I	NI	sum
EU	I	33	6	39
	NI	11	48	59
	sum	44	54	98

b. Outcome (%)

Sensitivity	85	(33/39)
Specificity	81	(48/59)
Positive predictivity	75	(33/44)
Negative predictivity	89	(48/54)
Accuracy	83	(81/98)
False negatives	15	(6/39)
False positives	19	(11/59)

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

10.1 How the STE test Will Refine, Reduce, or Replace Animal Use

With respect to these animal welfare considerations, the STE test refines and reduces the use of laboratory animals bred specifically for the purpose of toxicity testing. This assay uses cultured cells (rabbit corneal cell line) for identification ocular irritant. Therefore, the STE test reduces the use of laboratory animals (i.e., substances that are identified as ocular irritants *in vitro* would be excluded from testing *in vivo*).

11.0 Practical Considerations

11.1 Transferability of the STE test

11.1.1 Facilities and Major Fixed Equipment

The facility requirements necessary to conduct the STE test include a standard laboratory setup for cell culture under sterile condition. However the STE test is not required sterile condition during a testing period (i.e., from exposure of test substance to measurement of viability). The major equipment necessary to conduct the STE test is readily available and includes a CO_2 incubator, a clean bench, a haemocytometer and a microplate reader. Suppliers and estimated costs of this equipment are summarized in **Table 11-1** to the extent this information was available.

Table 11-1 Suppliers and Costs of Major Equipment for the STE test

Equipment	Supplier / Manufacturer	Estimated Costs
CO_2 incubator	e.g., Astec Co. Ltd, Yamato Scientific Co., Ltd.	~ $10,000
Clean bench	e.g., Airtech Japan, Ltd., Hitachi Appliances, Inc.	~ $17,000
Haemocytometer	e.g., Erma Inc.	~ $600
Microplate reader	e.g., BMG Labtech Ltd., Bio-Rad, Fisher Scientific	~ $20,000

11.1.2 General Availability of Other Necessary Equipment and Supplies

The remaining equipment and supplies necessary to conduct the STE test (e.g., culture flasks, 96-well flat bottom microplates, pipettes, and conical tubes) are readily available in most scientific laboratories or can be obtained from any of several scientific laboratory equipment suppliers.

11.2 STE test Training Considerations

11.2.1 Required Level of Training and Expertise Needed to Conduct the STE test

A training period of between one to two weeks is usually required for a technician with general laboratory skills to conduct all aspects of the STE test independently and proficiently. During the training period the technician would learn how to:

- Subculture of SIRC cells
- Seeding into 96-well microplate
- Exposure of test solution
- Washing cells
- Measurement of viability

There are currently no known proficiency criteria used to ensure that personnel are performing the test method competently. All of the tasks in the STE test are technically simple to perform. When a technician has mastered all aspects of the protocol, and can independently conduct the assay, such that the positive control falls within its historical range, the technician has essentially demonstrated proficiency in the assay.

11.3 Cost Considerations

The estimated cost of the STE test is about $100 per test substances. This cost includes both positive and negative controls.

11.4 Time Considerations

Use of the STE test would significantly reduce the time needed to assess the ability of a test substance to induce ocular irritant, when compared to the currently accepted *in vivo* rabbit eye test method. The *in vivo* Draize rabbit eye test is typically carried out for a minimum of one to three days. Depending upon the severity of ocular effects produced by a test substance, the method can be extended for up to 21 days. Comparatively, the STE test can be completed in about four hours, once the SIRC cells reach to confluent in 96-well microplate.

12.0 References

Bagley D, Booman KA, Bruner LH, Casterton PL, Demetrulias J, Heinze JE, Innis JD, McCormick WC, Neun DJ, Rothenstein AS, Sedlak RI. 1994. The SDA Alternatives Program Phase III: Comparison of *in vitro* data with animal eye irritation data on solvents, surfactants, oxidizing agents, and prototype cleaning products. Journal of Toxicology Cutaneous and Ocular Toxicology 13:127-155.

Balls M, Berg N, Bruner LH, Curren RD, De Silva O, Earl LK, Esdaile DJ, Fentem JH, Liebsch M, Ohno Y, Prinsen MK, Spielmann H and Worth AP. 1999. Eye irritation testing: the way forward,The report and recommendations of ECVAM workshop 34, Alternatives to laboratory animals, 27:53-77.

Barkman R, Germanis M, Karpe G, Malmborg AS. 1969. Preservatives in drops. Acta Ophthalmologica. 47(3):461-475.

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

Butscher P. 1953. Beitrag zur therapie von augenschadigunen durch thioglykolsaur bei der herstellung der sogenannten kaltwelle. Klinische Monatsblätter für Augenheilkunde, 122:349-350.

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

Carpenter CP, Smyth HF. 1946. Chemical burns of the rabbit cornea. American journal of ophthalmology, 29:60-73.

Cooper KR, Brown RS, Chang PW 1982. Accuracy of Blood Cytological Screening Techniques for the Diagnosis of a Possible Hematopoietic Neoplasm in the Bivalve Mollusc, *Mya arenaria*: Journal of Invertebrate Pathology 39: 281-289

Dalbey W, Rodriguez S, Wilkins K, Cope C. 1993. Reducing the number of rabbits in eye and skin irritancy tests. Journal of the American College of Toxicology 12:347-357.

DeSousa D, Rouse A, Smolon W. 1984. Statistical consequences of reducing the number of rabbits utilized in eye irritation testing: Data on 67 petrochemicals. Toxicology and Applied Pharmacology 76:234-242.

Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003, Official Journal of the European Union, L66, 26-35

Draize J, Woodard G, Calvery H. 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. The Journal of pharmacology and experimental therapeutics, 82:377-390.

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

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

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

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

Eskes C, Bessou S, Bruner L, Curren R, Harbell J, Jones P, Kreiling R, Liebsch M, McNamee P,

Pape W, Prinsen MK, Seidle T, Vanparys P, Worth A, Zuang V. 2005. Alternative (Non-Animal)

Methods for Cosmetics Testing: Current Status and Future Prospects, 3.3. Eye Irritation. Alternatives to laboratory animals, 33:47-81.

Estable JL. 1948. The ocular effect of several irritant drugs applied directly to the conjunctiva. American journal of ophthalmology, 31:837-844.

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

Gartner S. 1944. Blood vessels of the conjunctiva. Arch Opthalmol. 36:464-471.

Grant WM. 1974. Toxicology of the eye. In: Toxicology of the eye. 2nd ed. (Grant WM, ed.)

Springfield: Charles C Thomas.

Gautheron P, Dukic M, Alix D, Sina JF. 1992. Bovine Corneal opacity and permeability test: an *in vitro* assay of ocular irritancy. Fundamental and Applied Toxicology 18:442-449.

Gautheron P, Giroux J, Cottin M, Audegond L, Morilla A, Mayordomo-Blanco L, Tortajada A, Haynes G, Vericat JA, Pirovano R, Gillio Tos E,0 Hagemann C, Vanparys P, Deknudt G, Jacobs G, Prinsen M, Kalweit S, Spielmann H. 1994. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) assay. Toxicology *in vitro* 8:381-392.

Guideline for cosmetic safety evaluation. 2008 (edited by Japan Cosmetic Industry Association), Yakuji Nippo-sha, Tokyo, Japan. (in Japanese)

Hagino S, Okazaki Y and Itagaki H. 2008. An *in vitro* tier evaluation for the identification of cosmetic ingredients whitch are not ocular irritants, Alternatives to laboratory animals, 36:641-652.

Harbell JW, Koontz SW, Lewis RW, Lovell D, Acosta D. 1997. IRAG working group 4: Cell Cytotoxicity Assays. Food and Chemical Toxicology 35:79-126.

Hayashi K, Mori T, Abo T, Koike M, Takahashi Y, Sakaguchi H and Nishiyama N. A tiered approach combining the short time exposure (STE) test and the bovine corneal opacity and permeability (BCOP) assay for predicting eye irritation potential of chemicals. 2012a. The Journal of toxicological sciences (in press).

Hayashi K, Mori T, Abo T, Ooshima K, Hayashi T, Komano T, Takahashi Y, Sakaguchi H, Takatsu A and Nishiyama N. Two-stage bottom-up tiered approaches combining several *in vitro* assays for identification of eye irritation potential of chemicals including insoluble or volatile substances. 2012b. Toxicology *In Vitro* (submitted for publication).

Itagaki H, Hagino S, Kato S, Kobayashi T and Umeda M. 1991. An *in vitro* alternative to the draize eye-irritation test: Evaluation of the crystal violet staining method, Toxicology in Vitro, 5(2):139-143.

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

Triangle Park, NC:National Institute of Environmental Health Sciences.
Kaluzhny Y, Kandárová H, Hayden P, Kubilus J, D'Argembeau-Thornton L, Klausner M. 2011. Development of the epiocular™ eye irritation test for hazard identification and labelling of eye irritating chemicals in response to the requirements of the EU cosmetics directive and REACH legislation. ATLA Alternatives to Laboratory Animals 39: 339-364.

Kay JH, Calandra JC. 1962. Interpretation of eye irritation tests. Journal of the Society of Cosmetic Chemists 13:281-289.

Kojima H, Hayashi K, Sakaguchi H, Omori T, Otoizumi T, Kuwahara H, Hayashi T, Sakaguchi M, Toyoda A, Goto H, Watanabe S, Ahiko K, Nakamura T, Morimoto T, Zang V and Stoles W. The second phase Validation study of the Short Time Exposure (STE) test to assess the eye irritation potential of chemicals. 2012. (under preparation).

Leopold IH. 1945. Local toxic effect of detergents on ocular structures. Archives of ophthalmology 34:99-102.

Lewin L, Guillery H. 1913. Die Wirkungen von Arzneimitteln und Giften auf das Auge. Hirschwald, Berlin. 2nd edition.

Marsh RJ, Maurice DM. 1971. The influence of non-ionic detergents and other surfactants on human corneal permeability. Experimental eye research 11:43-48.

McDonald TO, Seabaugh V, Shadduck JA, Edelhauser HF. 1987. Eye irritation. In: Dermatotoxicology. (Marzulli FN, Maibach HI, eds). Washington:Hemisphere Publishing Corporation, 3rd edition, 641-696.

McLaughlin RS. 1946. Chemical burns of the human cornea. American journal of ophthalmology, 29:1355- 1362.

McNamee P, Hibatallah J, Costabel-Farkas M, Goebel C, Araki D, Dufour E, Hewitt NJ, Jones P, Kirst A, Le Varlet B, Macfarlane M, Marrec-Fairley M, Rowland J, Schellauf F and Scheel J. 2009. A tired approach to the use of alternative to animal testing for the safety assessment of cosmetics: Eye irritation, Regulatory toxicology and pharmacology, 54:197-209.

Mikkelson TJ, Chrai SS, Robinson J R. 1973 Altered bioavailability of drugs in the eye due to

drug-protein interaction. Journal of Pharmaceutical Sciences 62:1648-1653.

Motose K. 1984. Bioavailability of eye drops and eye lotion. Nanzando, Tokyo, 179-185. (in Japanese)

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

OECD. 1987. Test guideline 405, Acute eye irritation/corrosion, adopted February 24, 1987. In OECD Guidelines for Testing of Chemicals. OECD, Paris.

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

Okamoto Y, Kanzaki N and Tanaka N. 1990 Studies of an *in vitro* alternative method to the draize rabbit eye irritation test, Nihonkeshohingijutsushakaishi (J. Soc. Cosmet. chem. Japan), 23:272-279.(in Japanese)

Ohno Y, Kaneko T, Inoue T, Morikawa Y, Yoshida T, Fujii A, Matsuda M, Ohno T, Hayashi M, Momma J, Uchiyama T, Chiba K, Ikeda N, Imanishi Y, Itagaki H, Kakishima H, Kasai Y, Kurishita A, Kojima H, Matsukawa K, Nakamura T, Ohkoshi K, Okumura H, Saijo K, Sakamoto K, Suzuki T, Takano K, Tutsumi H, Tani N, Usami M and Watanabe R. 1999. Interlaboratory validation of the *in vitro* eye irritation tests for cosmetic ingredients. (1) Overview of the validation study and draize scores for the evaluation of the tests, Toxicology in Vitro, 13:73-98.

Sakaguchi H, Ota N, Omori T, Kuwahara H, Sozu T, Takagi Y, Takahashi Y, Tanigawa K, Nakanishi M, Nakamura T, Morimoto T, Wakuri S, Okamoto Y, Sakaguchi M, Hayashi T, Hanji T, Watanabe S. 2011. Validation study of the Short Time Exposure (STE) test to assess the eye irritation potential of chemicals, Toxicology in Vitro, 25:796-809.

Scott L, Eskes C, Hoffmann S, Adriaens E, Alepee N, Bufo M, Clothier R, Facchini D, Faller C, Guest R, Harbell J, Hartung T, Kamp H, Le Varlet B, Meloni M, McNamee P, Osborne R, Pape W, Pfannenbecker U. Prinsen M, Seaman C, Spielman H, Stokes W, Trouba K, Van den Berghe C, Van Goethem F, Vassallo M, Vinardell P and Zuang V. 2010. A proposed eye irritation

testing strategy to reduce and replace *in vivo* studies using Bottom-Up and Top-Down approaches, Toxicology in Vitro, 24(1):1-9.

Solti J, Freeman JJ. 1988. Effect of reducing the number of animals in acute toxicity/irritation tests on U.S. and European labeling requirements. The Toxicologist 8:263.

Springer J, Chambers W, Green S, Gupta K, Hill R, Hurley P, Lambert L, Lee C, Lee J, Liu P, Lowther D, Roberts C, Seabaugh V, Wilcox N. 1993. Number of animals for sequential testing. Food and Chemical Toxicology 31:105-109.

Suker GF. 1913. Injury to cornea from oxalic acid. Opthalmol Rec. 23:40-47.

Takahashi Y, Koike M, Honda H, Ito Y. Sakaguchi H, Suzuki H and Nishiyama N. 2008. Development of the short time exposure (STE) test: an *in vitro* eye irritation test using SIRC cells. Toxicology in Vitro 22:760-770.

Takahashi Y, Hayashi T, Watanabe S, Hayashi K, Koike M, Aisawa N, Ebata S, Sakaguchi H, Nakamura T, Kuwahara H and Nishiyama N. 2009. Inter-laboratory study of short time exposure (STE) test for predicting eye irritation potential of chemicals and correspondence to globally harmonized system (GHS) classification. The Journal of toxicological sciences 34:611-626.

Takahashi Y, Hayashi T, Koike M, Sakaguchi H, Kuwahara H and Nishiyama N. 2010. An interlaboratory study of the short time exposure (STE) test using SIRC cells for predicting eye irritation potential. Cutaneous and ocular toxicology 29:77-90.

Takahashi Y, Hayashi K, Abo T, Koike M, Sakaguchi H and Nishiyama N. 2011. The Short Time Exposure (STE) test for predicting eye irritation potential: intra-laboratory reproducibility and correspondence to globally harmonized system (GHS) and EU eye irritation classification for 109 chemicals. Toxicology in Vitro, 25:1425-1434.

Talsma D, Leach C, Hatoum N, Gibbons R, Roger J-C, Garvin P. 1988. Reducing the number of rabbits in the Draize eye irritancy test: A statistical analysis of 155 studies conducted over 6 years. Fundamental and Applied Toxicology 10:146-153.

Tani N, Kinoshita S, Okamoto Y, Kotani M, Itagaki H, Murakami N, Sugiura S, Usami M, Kato K,

Kojima H, Ohno T, Saijo K, Kato M, Hayashi M and Ohno Y. 1999. Interlaboratory validation of the *in vitro* eye irritation etsts for cosmetic ingredients. (8) Evaluation of cytotoxicity tests on SIRC cells, Toxicology in Vitro, 13:175-187.

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

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

Wilhelm KP, Bottjer B and Siegers CP. 2001 Quantitative assessment of primary skin irritants *in vitro* in a cytotoxicity model: comparison with *in vitro* human irritation tests, The British journal of dermatology 145:709-715.

13.0 Supporting Materials (Appendices)