Organisation for Economic Co-operation and Development

ENV/JM/MONO(2019)9

Unclassified

English - Or. English

17 April 2019

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

Streamlined Summary Document Supporting OECD Test Guideline 438 on the Isolated Chicken Eye for Eye Irritation/Corrosion

Series on Testing and Assessment
No. 188

JT03446389

SERIES ON TESTING AND ASSESSMENT
NO. 188

STREAMLINED SUMMARY DOCUMENT SUPPORTING OECD TEST GUIDELINES 438 ON THE ISOLATED CHICKEN EYE FOR EYE IRRITATION/CORROSION

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2019

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 35 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents;** and **Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This streamlined summary document (SSD) was developed to provide summary information in support of OECD Test Guideline 438 on the Isolated Chicken Eye Test Method addressing the endpoint eye irritation/serious eye damage. This SSD was developed by a Secretariat consultant in March 2013 and revised in March 2018 to be submitted to the Working Group of the National Coordinators of the Test Guidelines Programme (WNT) together with the updated version of TG 438 (originally adopted in 2009 and revised in 2013, 2017 and 2018). The SSD provides useful and more detailed information than is otherwise available from the Test Guideline itself on: 1) the scientific basis of the test method, 2) the identified limitations, weaknesses and strengths, 3) the applicability domain, 4) the sensitivity, specificity and accuracy, and 5) the within-laboratory and between-laboratory reproducibility of the method.

The SSD was approved by the WNT in May 2018.

The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to its declassification on 10 July 2018.

This document is published under the responsibility of the Joint Meeting of the Chemicals committee and the Working Party on Chemicals, Pesticides and Biotechnology.

STREAMLINED SUMMARY DOCUMENT

Description of applicability domain and performance, based on the retrospective validation studies and their revisions, of the Isolated Chicken Eye (ICE) Test Method (Test Guideline 438) for identifying i) chemicals inducing serious eye damage and ii) chemicals not requiring classification for eye irritation or serious eye damage.

INTRODUCTION AND BACKGROUND

The 2003-2006 Validation Studies

1. Between 2003 and 2006, a retrospective evaluation was carried out concerning the validation status of the Isolated Chicken Eye (ICE) test method for identifying chemicals (substances and mixtures) inducing serious eye damage ("ocular corrosives and severe irritants"), i.e., its usefulness and limitations for initiating a Top-Down approach (Scott et al., 2010). This evaluation, comprising a total of 175 chemicals, was performed by the US-Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the US-National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), in collaboration with the EU-European Centre for the Validation of Alternative Methods (ECVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM). For a full description, see the Background Review Document (BRD) (ICCVAM-NICEATM, 2006) and the ICCVAM Test Method Evaluation Report (TMER) (ICCVAM, 2007).

2. The study aimed at characterising the reproducibility and predictive capacity of the ICE for the following classification systems: UN GHS (Category 1) (UN, 2017), US EPA (Category I) (US EPA, 2011) and EU DSD (R41) (EC, 2001) (the EU CLP classification system (EC, 2008) based on UN GHS had not yet been adopted at that time). On the basis of all collected data and information the ICE was considered as scientifically valid (reliable and relevant) for identifying chemicals inducing serious eye damage (i.e., to initiate a Top-Down approach (Scott et al., 2010)) and was recommended for regulatory hazard classification and labelling purposes. Chemicals inducing serious eye damage are defined as those that produce tissue damage in the eye, or serious physical decay of vision, following application to the anterior surface of the eye in the in vivo Draize rabbit eye test (OECD, 2017a), which is not (or not expected to be) fully reversible within 21 days of application. Following these recommendations, the OECD officially adopted the ICE test method as OECD Test Guideline (TG) 438 for identifying chemicals inducing serious eye damage in September 2009 (OECD, 2018a).

The 2006-2009 Validation Studies

3. NICEATM/ICCVAM, in collaboration with ECVAM and JaCVAM, further evaluated between 2006 and 2009 the usefulness and limitations of the ICE test method for the additional identification of chemicals not causing sufficient effects on the eye to require hazard classification and labelling according to the UN GHS (UN, 2017), US EPA (US EPA, 2011), US FHSA and EU DSD (EC, 2001) classification systems (the EU CLP classification system (EC, 2008) based on UN GHS had not been adopted at that time), i.e., its usefulness and limitations for initiating a Bottom-Up approach (Scott et al., 2010). The

ICE validation database remained unchanged and comprised 175 chemicals (90 substances and 85 mixtures) collected from five individual studies (Prinsen and Koëter, 1993; Balls et al., 1995; Prinsen, 1996; Prinsen, 2000; Prinsen, 2005), which were used to determine the predictive capacity of the ICE test method (ICCVAM-NICEATM, 2009).

4. In May 2009, NICEATM/ICCVAM convened an independent international scientific peer review panel (PRP) on alternative ocular safety testing methods, composed of members from EU, USA, Japan and Canada. The PRP maintained the original recommendation for using the ICE for classification of serious eye damage (ICCVAM, 2010). At that stage however, no further recommendations were made for an expansion of the ICE applicability domain to also include other classification categories, and in particular the identification of non-classified chemicals (ICCVAM, 2010). This was due to the false negatives rates (6% or 4/62 for the UN GHS classification system) and the fact that amongst the false negatives there was one substance was classified as GHS Cat 1 based on Draize rabbit eye test data (ICCVAM, 2010).

Revisions of the Validation Dataset

5. On the basis of revisions of the validation dataset carried out in 2012 it was found that the individual in vitro and in vivo classifications of a number of chemicals deserved further considerations. In particular discrepancies were found in the final in vivo and in vitro classifications for a number of chemicals in the ICCVAM BRD from April 2009 (ICCVAM-NICEATM, 2009) which had an impact on the number of false negative chemicals (1 out of 73 false negative instead of 4 out of 62; see Appendixes 2, 3 and 4). In addition, it was felt important to recognize the limitations of the in vivo Draize rabbit eye irritation test and their implications for validation purposes (SSD, 2012). Some of the drawbacks of the Draize rabbit eye test referred in literature include (see extract from Eskes (2010) in Appendix 4 for details):

- The fact that the in vivo rabbit eye irritation/corrosion test has no standardized exposure regimen, so that the duration of exposure of the test substance with the rabbit eyes remains unknown and can vary from a few minutes to several hours. In addition, for solids and sticky chemicals it is unclear how much of the compound (solid, paste or liquid) stays in contact with the eye (Prinsen, 2006);
- The limited reproducibility of the Draize rabbit eye test method;
- The subjectivity in the allocation of the rabbit ocular tissue scores;
- The type of exposure which does not reflect a potential human accidental exposure;
- The differences in physiology and sensitivity to tested chemicals between rabbit and human eyes;
- Ethical issues and the fact that the Draize test can be very painful to the rabbits.

Adoption of TG 438 for the Identification of Chemicals Not Classified for Eye Irritation or Serious Eye Damage

6. In April 2012, following a proposal from the Netherlands and the European Commission, a project for updating TG 438 was included in the work plan of the OECD Test Guidelines Programme. The aim of the project was, taking into account the review of the individual in vivo and in vitro data, to reassess the performances of ICE and address a possible update of TG 438 to allow its use also for the identification of chemicals not

requiring classification for eye irritation or serious eye damage under the UN GHS classification system. An initial re-evaluation of the in vitro and in vivo ICE dataset was carried out by the Netherlands, followed by the preparation of an Issue Paper and its addendum by a consultant to the OECD Secretariat with the aim to review existing ICE data and make a recommendation on the use of TG 438 for the identification of chemicals not requiring classification for eye irritation or serious eye damage (Appendixes 4 and 5). The ICE Issue Paper and its addendum reviewed both the ICE data presented in the ICCVAM-NICEATM ICE BRDs (ICCVAM-NICEATM 2006, 2009) as well as the evaluation carried out by The Netherlands in the first draft version of this SSD (2012). The Issue Paper and its addendum were discussed by the OECD Expert Group on Eye Irritation during a meeting held on 6-7 December 2012. The present SSD represents a compilation of all relevant data on the ICE test method evaluation and takes into account all comments received from the OECD Expert Group on Eye Irritation regarding the revisions of TG 438.

7. In 2013, the OECD approved the updated version of TG 438 allowing the use of ICE for identifying chemicals inducing serious eye damage as well as for identifying chemicals not requiring classification for eye irritation or serious eye damage.

Updated ICE Decision Criteria for the Identification of Chemicals Not Classified for Eye Irritation or Serious Eye Damage

8. Since the revisions from 2013, the criteria for acceptance of test methods to identify UN GHS No Category chemicals have changed. This is based on the work performed by Adriaens et al. (2014) who showed that with the Draize in vivo eye irritation test itself, at least 11% of chemicals classified in vivo as UN GHS / EU CLP Category 1 could equally be identified as Category 2, and that about 12% of the Category 2 chemicals could equally be identified as non-classified chemicals. Such values take into account nevertheless only the in vivo within-test variability, and not variability between tests and between laboratories, suggesting that the actual values might be higher when considering also potential variability between tests and between laboratories for the same chemical. In 2015, two New Test Guidelines were adopted (OECD TG 492 and OECD TG 491), having higher false negative rates as compared to TG 438 revised in 2013 (4% and 2% respectively versus 1%). Based on this it was agreed to reconsider the Decision Criteria of the ICE test method for the identification of UN GHS No Category chemicals taking into account the original criteria proposed by the test developer (i.e., criteria '2xII, 1xI' proposed to identify UN GHS No Category (and not 'No Prediction Can Be Made) (see also paragraph 11)). An evaluation was conducted in 2017 to assess the impact of changing these decision criteria, taking into account new available ICE in vitro data having parallel Draize in vivo data (Appendixes 1 and red-shadowed rows of Appendix 2). The results of this evaluation led to the revision of OECD TG 438 and the corresponding updates to the OECD GD 160 in 2018 where the criteria '2xII, 1xI' is used to identify UN GHS No Category (instead of 'No Prediction Can Be Made').

Use of ICE Histopathology to Identify Non-Extreme pH (2 < pH < 11.5) Detergents and Surfactants Inducing Serious Eye Damage

9. Since the revisions from 2013, the International Association for Soaps, Detergents and Maintenance Products (A.I.S.E.) showed that specific ICE histopathological effects (i.e. at least moderate level of epithelial erosion and/or epithelial vacuolation in the mid and/or lower layers) were found to correlate with serious eye damage classification induced

by non-extreme pH (2<pH<11.5) detergents and surfactants in vivo (Cazelle et al., 2014, 2015). Use of these histopathology criteria for non-extreme pH detergents allowed to decrease the false negative rates (increase in sensitivity) obtained with the standard ICE prediction model for the identification of UN GHS Cat. 1 non-extreme pH (2<pH<11.5) detergents and surfactants, increase the accuracy and maintain an acceptable specificity (see paragraphs 29 to 32). In particular, use of ICE histopathology allowed to correctly identify a majority of non-extreme pH detergents classified as UN GHS Cat. 1 based on in vivo persistence of effects i.e., having tissue effects that do not reverse 21 days after treatment and that do not lead to severity of effects that would warrant a UN GHS Cat. 1 classification (Cazelle et al., 2014). Furthermore, appropriate reproducibility of the histopathological evaluation by the pathologists (and peer-reviewers) from three independent laboratories and over time was found provided that i) an internal peer-review system was in place; ii) original slides were assessed (and not microphotographs) in order to enable the evaluation of three dimensional effects; and iii) appropriate training & proficiency appraisal was conducted. Based on those findings, the ICE histopathology was included in TG 438 in 2018, for the specific applicability domain of non-extreme pH (2<pH<11.5) detergents and surfactants. However, based on the dataset currently available, histopathology cannot be used in a stand-alone manner to identify UN GHS Cat. 2 and UN GHS No Cat. test chemicals. Furthermore, histopathology did not improve the predictive capacity of the ICE test method for extreme pH detergents where a majority of in vivo UN GHS Cat. 1 were classified in vivo due to severity of effects (and not persistence as in the case of non-extreme pH detergents), suggesting that the applicability domain of the ICE histopathology is likely based on the mode of action of the tested chemicals (see paragraph 29). Additional appropriate and relevant data are needed to verify and expand the applicability of the ICE histopathology decision criteria to chemistries other than non-extreme pH (2<pH<11.5) detergents and surfactants.

SCIENTIFIC BASIS FOR THE ICE TEST METHOD

10. The ICE test method (TG 438) is an organotypic model that provides short-term maintenance of the chicken eye in vitro. In this test method, damage by the test chemical is assessed by determination of corneal swelling, opacity, and fluorescein retention. Furthermore, histopathology can be used to increase the sensitivity of the method for identifying UN GHS Category 1 non-extreme pH (2 < pH < 11.5) detergents and surfactants (see paragraphs 29-32 and Appendix 7). Whilst measurement of corneal swelling provides for a quantitative assessment, corneal opacity, fluorescein retention and histopathological changes each involve a qualitative assessment. Each measurement is either converted into a quantitative score used to assign an ICE Class (I to IV), or assigned a qualitative categorization that is used to assign an in vitro ocular hazard classification, either as UN GHS Category 1 or as UN GHS No Category. Either of these outcomes can then be used to predict the potential in vivo serious eye damage or no requirement for eye hazard classification of a test chemical (see TG 438, 2018a). However, no prediction can be made for chemicals not identified as causing serious eye damage (UN GHS Category 1) or as UN GHS No Category with the ICE test method.

Data Interpretation Procedures

11. If the criteria used to attribute the ICE classes (I to IV) for the three endpoints (corneal swelling, corneal opacity and fluorescein retention) remained unchanged, the Data

Interpretation Procedure (DIP) proposed within the 2013 revision of OECD TG 438 was the one proposed by ICCVAM (IICVAM-NICEATM, 2009) which differed slightly from the one proposed by The Netherlands. However, following revisions to the TG 438 undertaken from 2015 to 2018, it was agreed to make use of the DIP originally proposed by The Netherlands as described in Table 1.

Table .1. Data Interpretation Procedures applied for the ICE test method for identifying i) chemicals inducing serious eye damage and ii) chemicals not requiring classification for eye irritation or serious eye damage.

UN GHS Classification	Combinations of the 3 Endpoints*
No Category	3 x I
	2 x I,1 x II
	2 x II,1 x I
No prediction can be made	Other combinations
Category 1	3 x IV
	2 x IV, 1 x III
	2 x IV, 1 x II**
	2 x IV, 1 x I**
	Corneal opacity = 3 at 30 min (in at least 2 eyes)
	Corneal opacity = 4 at any time point (in at least 2 eyes)
	Severe loosening of the epithelium (in at least 1 eye)

Notes:
* Based on the criteria proposed in the latest updated TG 438 (OECD, 2018a) and in the Guidance Document 160 (OECD, 2018b)
**Combinations less likely to occur.

Comparison of the ICE Test Method with the In Vivo Rabbit Eye Test Method

12. In contrast to ICE, the in vivo rabbit eye test involves only qualitative evaluations based mainly on visual observations 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 ICE, liquids and solids are typically tested undiluted and are applied to evenly cover the entire surface of the cornea. In the in vivo rabbit eye test, liquid and solid test substances are also tested usually undiluted, however they are applied to the conjunctival sac of the rabbit eyes. Due to rabbits blinking and/or tearing, exposure of the cornea to the test substance will be affected by these factors in terms of coverage or duration. The neurogenic components that drive tear film production are not present in the ICE. When compared with an in vivo rabbit eye study, application of a test substance in the absence of this protective barrier might be expected to cause an increase in false positive outcomes. One of the conclusions from a workshop on mechanisms of eye irritation highlighted the need for additional research on the impact of chemicals on tear film and the consequences of tear film disruption. However, for some test substances (e.g., solids), blinking can also induce mechanical damage in vivo,

contributing to a higher degree of irritation. Thus, the ICE test method differs from the in vivo rabbit eye test method in the following significant ways:

- The ICE evaluates only corneal effects and does not assess effects on the iris and the conjunctiva as performed in the in vivo rabbit eye test. Measurements are performed quantitatively and qualitatively with the help of a slit-lamp in the ICE assay, while they are assessed only qualitatively based mainly on visual observations in the in vivo rabbit eye test.
- Corneal exposure conditions, including test substance concentration and exposure duration, are well defined in the ICE assay, whereas subject to potentially large variations in vivo due to the ill-defined exposure conditions, blink response and natural tearing of the eye in a live animal. Moreover, based on the unrealistic accidental in vivo exposure conditions, solids may lead to variable and extreme responses in the in vivo Draize eye irritation test, which may not reflect their true irritation potential in humans (Prinsen, 2006).
- The observation period of the ICE assay is typically of 4 hours, whereas ocular effects are typically evaluated in the in vivo rabbit eye test for a minimum of 72 hours and can extended up to 21 days.
- Reversibility/irreversibility of corneal effects induced by a test substance cannot be observed in the ICE assay per se, but histopathological evaluation of the exposed eyes may provide additional information about the depth and type of injury that could aid in some cases predicting whether damage is irreversible or not as e.g. described in paragraph 9. It has been proposed, based on rabbit eye studies, that an assessment of the initial depth of corneal injury may be used to identify some types of irreversible effects (Maurer et al., 2002) although further scientific knowledge is required to understand how irreversible effects not linked with initial high level injury occur.
- Protective mechanisms of the eye, such as tear production and blinking (e.g., against drying and infection), are built into in vivo testing, but are absent in in vitro / ex vivo testing. However, if regeneration of the tear film might be important for the in vivo healing process, it may play a minor role in the ICE test method where the initial damage is measured rather than recovery.
- The ICE assay does not account for systemic effects following ocular instillation that may be noted with the in vivo rabbit eye test (e.g., toxicity or lethality as in the case of certain pesticides). However, these effects are typically predicted from other acute toxicity test methods, and may not be relevant for the many consumer products that are formulated with well characterized raw materials of known systemic toxicity.

IDENTIFIED LIMITATIONS, WEAKNESSES AND STRENGTHS

13. The potential shortcomings of the ICE test method when used to identify chemicals inducing serious eye damage (UN GHS Category 1) in e.g., a Top-Down approach, are based on the high false positive rate for alcohols and the high false negative rates for solids and surfactants, as observed in the 2003-2006 retrospective validation study (ICCVAM-NICEATM, 2006; ICCVAM, 2007). Moreover, test chemicals inducing persistent non severe effects in vivo may also risk underprediction (Barroso et al., 2017). However, false negative rates in this context (UN GHS Category 1 identified as not being UN GHS

Category 1) are not critical since all test chemicals that come out negative would be subsequently tested with other adequately validated in vitro test(s), or as a last option in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight-of-evidence approach. Furthermore, histopathology was found to be a useful additional endpoint to decrease the false negative rates when used to identify UN GHS Category 1 non-extreme pH ($2 < pH < 11.5$) detergents shown to induce mainly persistent non severe effects in vivo and surfactants (Cazelle et al, 2014, 2015; Appendix 7). Regarding solids, it should be noted that these may lead to variable and extreme exposure conditions in the in vivo Draize eye irritation test, which may result in irrelevant predictions of their true irritation potential (Prinsen, 2006). Investigators could consider using this test method for all types of chemicals, whereby a positive result should be accepted as indicative of serious eye damage, i.e., UN GHS Category 1 classification without further testing. However, positive results obtained with alcohols should be interpreted cautiously due to risk of over-prediction.

14. When used to identify chemicals not requiring classification for eye irritation or serious eye damage under the UN GHS classification system, in e.g., a Bottom-Up approach, anti-fouling organic solvent containing paint were found to risk under-prediction (1 out of 2 classified anti-fouling solvent containing paint was found to be under-predicted as non-classified). When chemicals within this product classes are excluded from the database, the accuracy of the ICE test method is only slightly improved (see Table 7 below). In addition, the only underpredicted material (TNO-94) was classified in vivo as GHS Cat 1 due to unusual effects, i.e., a residue of paint got attached to the cornea most probably caused by grooming/scratching of the eye by the rabbit and the type of exposure of the rabbit eyes (i.e., adding the paint in the conjunctival sac of the eye and holding the eye lids together). This was observed in only one animal, whereas the two other animals only showed GHS Cat 2-type effects (see Appendix 6). In 2013, it was agreed to add a warning sentence for this category of materials, but not to exclude them from the applicability domain of the ICE test method for the following reasons: i) there was insufficient evidence to exclude those types of formulations (only two classified chemicals from this category), ii) the type of exposure to this material is unlikely to occur in humans, and iii) sticky materials present similar difficulties to test either in vitro and in vivo. Regarding the false positive rates, surfactants (6 out of 9) may risk overprediction. However due to i) the fact that not all surfactants were overpredicted, ii) that some surfactants were correctly predicted, and that iii) surfactants not predicted as GHS non-classified would need to be subsequently tested with other suitable test method(s) in a sequential testing strategy (ICCVAM-NICEATM, 2006) (as none of the in vivo GHS non-classified chemicals (n=83) was overpredicted as GHS Cat. 1), surfactants are not considered to be out of the applicability domain of the ICE test method.

15. Whilst the ICE test method does not directly address conjunctival and iridial injuries as evaluated in the rabbit ocular irritancy test method, it addresses corneal effects which are the major driver of classification in vivo when considering the UN GHS Classification (Barroso et al., 2017). In this respect, it should be noted that effects on the iris are of lesser importance for classification of chemicals according to UN GHS (Adriaens et al., 2014; Barroso et al., 2017). Prinsen (1996) reported a high correlation between conjunctival reactions and the endpoints assessed in the ICE test method after parallel testing of test substances in vivo and in vitro. Furthermore, although the reversibility of corneal lesions cannot be evaluated per se in the ICE test method, it has been shown that histopathological observations can help in identifying test chemicals causing irreversible effects not linked with initial high level injury such as those caused by non-extreme pH (2

< pH < 11.5) detergents (Cazelle et al., 2014). Finally, the ICE test method does not allow for an assessment of the potential for systemic toxicity associated with ocular exposure

16. The ICE test method is not recommended for the identification of test chemicals that should be classified as irritating to eyes (UN GHS Category 2 or Category 2A) or test chemicals classified as mildly irritating to eyes (UN GHS Category 2B) due to the considerable number of UN GHS Category 1 chemicals underclassified as UN GHS Category 2, 2A or 2B and UN GHS No Category chemicals overclassifed as UN GHS Category 2, 2A or 2B. For this purpose, further testing with another suitable method may be required.

Table 2. Physicochemical properties and compatibility with the ICE

Physicochemical property	Is a material with this property compatible with the ICE assay?
Fixative	Yes
Solvent	Yes
Extreme pH	Yes
Gases	No
Liquids	Yes
Solid materials	Yes
Emulsions	Yes
Granular materials	Yes
Suspensions	Yes
Coloured materials	Yes
Diluted concentrations of chemicals	Yes
Highly viscous materials	Yes
Volatile materials	Yes
Reactive chemistries	Yes
Hydrophobic/lipophilic chemicals	Yes
Neat concentrations of chemicals	Yes

APPLICABILITY DOMAIN

17. The standard ICE test method can be used for testing all types of substances and mixtures, provided there is no evidence that the method is not valid for the chemical tested. The ICE histopathology in contrast, can be used for the specific applicability domain of non-extreme pH (2<pH<11.5) detergents and surfactants. However, histopathology did not improve the predictive capacity of the ICE test method for extreme pH detergents where a majority of in vivo UN GHS Cat. 1 were classified in vivo due to severity of effects (and not persistence as in the case of non-extreme pH detergents), suggesting that the applicability domain of the ICE histopathology is likely based on the mode of action of the

tested chemicals (see paragraph 29).Additional appropriate and relevant data are needed to verify and expand the applicability of the ICE histopathology decision criteria to other chemistries.

Categories of Irritancy

18. Based on the conclusions of the 2003-2006 and 2006-2009 retrospective validation studies (ICCVAM, 2007, 2010), TG 438 was adopted in 2009 for classification of chemicals inducing serious eye damage (UN GHS Category 1). In addition, following the re-evaluations carried out in 2012 and their review by the OECD Expert Group on Eye Irritation, TG 438 was also approved in 2013 for the identification of chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Category), under the UN GHS classification system.

Potential Role in Integrated Approaches to Testing and Assessment (IATA)

19. It is currently generally accepted that, in the foreseeable future, no single in vitro eye irritation test will be able to fully replace the in vivo Draize eye test to predict across the full range of irritation for different chemical classes. However, strategic combinations of alternative test methods within a (tiered) testing strategy and/or IATA may be able to replace the Draize eye test (OECD GD 263, 2017b; Scott et al., 2010). The Top-Down approach is designed to be used when, based on existing information, a test chemical is expected to have high irritancy potential, while the Bottom-Up approach is designed to be used when, based on existing information, a test chemical is expected not to cause sufficient eye irritation to require a classification (OECD GD 263, 2017b; Scott et al., 2010). The ICE test method is an in vitro test method that can be used, under certain circumstances and with specific limitations as described in paragraphs 13 to 16 for eye hazard classification and labelling of chemicals. While it is not considered valid as a stand-alone replacement for the in vivo rabbit eye test, the ICE test method is recommended as an initial step within a testing strategy such as the Top-Down approach recommended within the OECD GD 263 (OECD, 2017b) to identify chemicals inducing serious eye damage, i.e., chemicals to be classified as UN GHS Category 1 without further testing (UN, 2017). The ICE test method is also recommended to identify chemicals that do not require classification for eye irritation or serious eye damage as defined by the UN GHS (No Category) (UN, 2017), and may therefore be used as an initial step within a Bottom-Up testing strategy approach (OECD GD 263, 2017b). However, a chemical that is not predicted as causing serious eye damage or as not classified for eye irritation/serious eye damage with the ICE test method would require additional information to establish a definitive classification. Choice of the most appropriate test method(s) should be seen in the context of the OECD Guidance Document on an Integrated Approach on Testing and Assessment for Serious Eye Damage and Eye irritation (OECD GD 263, 2017b).

Mode of Action (MoA)

20. An expert meeting held at EC-ECVAM in 2005 (Scott et al., 2010) recommended to expand the concept of defining the applicability domain as not only chemical classes, but also as a function of the mechanism of eye irritation. The four identified MoA that were discussed included: (i) cell membrane lysis (breakdown of membrane integrity as might occur from exposure to membrane active materials, e.g., surfactants), (ii) saponification (breakdown of lipids by alkaline action), (iii) coagulation (precipitation/denaturation of macromolecules, particularly protein, characteristic of acid, alkali, or organic solvent exposure), and (iv) actions on macromolecules (chemicals that react with cellular

constituents/organelles that may or may not lead to overt lysis or coagulation, e.g., peroxides, mustards and bleaches). The ICE test method addresses the three first MoAs. In addition, it may also address the fourth MoA (actions on macromolecules) when histopathological information is available.

Table 3. Summary of events involved in chemical-induced eye irritation in vivo

Events involved in chemical-induced eye irritation	Modelled by the ICE assay?
Chemical interaction with tear film	No
Chemical binding to the conjunctival epithelium	No
Adhesion molecules compromised	Yes
Corneal epithelial damage	Yes
Inhibition of receptor-mediated transport	Yes
Compromise of cell membrane integrity of upper corneal epithelium	Yes
Cell **membrane lysis** of all corneal epithelial layers	Yes
Hydration of corneal stroma	Yes
Cross-linking *of proteins in corneal stroma*	Yes
Erosion of corneal stroma	Yes
Cell damage to corneal epithelium and limbus	Yes
Dilation and increased lymphatic leakage from scleral vasculature	No
Stimulation of nerve endings, i.e*., enhanced blinking, tearing*	No
Erosion of nerve endings in corneal and sclera	No
Duration of the response, *i.e.*, length of time cell responses deteriorate. Duration of response covers the effects of **reactive chemicals**, which can cause **coagulation**, **saponification**, that are effects, which develop and increase over time.	Yes
Recovery from response, *i.e.*, length of time for cell responses to return to control levels	No

Chemical Classes

21. The ICE validation dataset comprised a total of 207 test chemicals (93 substances and 114 mixtures) collected from Prinsen and Koëter (1993), Balls et al. (1995), Prinsen (1996), Prinsen (2000) and Prinsen (2005) as described in the retrospective validation studies (ICCVAM-NICEATM, 2006, 2009), and including the additional dataset as described in Appendix 1. Tables 4 and 5 show the chemical and product classes representation within the ICE validation database. Although the single components from the mixtures used in the validation dataset could not be disclosed due to proprietary reasons they represented relevant to current commerce mixtures and formulations. In addition, details on the product categories and chemical classes of most of the tested mixtures and substances were available as described in the ICCVAM BRD from April 2009 (ICCVAM-NICEATM, 2009). Out of the 207 test chemicals, 85 (including mixtures) could not be assigned a specific chemical class and 23 (including substances) could not be assigned a specific product category. Detailed information, including chemical name, Chemical Abstracts Service Registry Number (CASRN), chemical class, product category, physical state, purity, concentration(s) tested, in vivo GHS classification, in vitro GHS

classification, in vitro raw data, in vitro categories and literature reference using the chemical are provided in Appendixes 1 and 2.

Table4. Chemical classes tested in the ICE test method*.

Chemical Class	# of Chemicals	Chemical Class	# of Chemicals
Acetate	1	Inorganic Salt	3
Acid	5	Inorganic Silver / Nitrogen Compound	1
Acyl halide	1	Ketone	4
Alcohol	15	Lactone	1
Aldehyde	2	Lipid	1
Alkali	3	Nitrile	1
Amide /Amidine	7	Nitro Compound	1
Amino acid	1	Not Classified	85
Boron compound	1	Onium Compound	8
Carbohydrate	2	Organic Silicon Compound	2
Carboxylic acid	12	Organic Sulfur Compound	3
Ester	10	Organometallic	2
Ether	1	Organophosphrous Compound	1
Heterocyclic	9	Polycyclic	4
Hydrocarbon	5	Polyether	5
Imide	2	Surfactant	3
Inorganic Chemical	1	Urea Compound	1
Inorganic Chloride Compound	1		

Note: * Revised from (ICCVAM-NICEATM, 2009) based on new information presented in Appendix 1 and revised information presented in Appendix 2.

Table 5. Product classes tested in the ICE test method*.

Product Class	# of Chemicals	Product Class	# of Chemicals
Adhesive	2	Fertilizer	1
Antifungal	3	Food additive	1
Antihistamine	1	Fungicide / Germicide / Bactericide	8
Anti-infective	3	Industrial Chemical, Intermediate or Formulation	19
Antiseptic	2	Not Classified	23
Caustic Agent	4	Optical Resolution Agent	1
Chlorination by-product	1	Paint	4
Cleaner	16	Pesticide / Herbicide / Agrochemical	27
Copolymer	8	Preservative	6
Cosmetic Ingredient	1	Pharmaceutical Compounds / Intermediates	6
Detergent	11	Raw Material	9
Developer	1	Reagent	4
Disinfectant	5	Resin	2
Dyes & Stains	10	Silicon Resin	1
Elastomer	2	Soap	9
Enzyme Inhibitor	1	Surfactant	34
Enzyme Solution	3	Solvent	37

Note: * Revised from (ICCVAM-NICEATM, 2009) based on new information presented in Appendix 1 and revised information presented in Appendix 2.

SENSITIVITY, SPECIFICITY AND ACCURACY

22. Within the ICE validation database a total of 184 individual chemicals (75 substances and 109 mixtures) had sufficient in vivo and in vitro data to assess the ICE predictive capacity (Appendixes 1 and 2). Their distribution according to the UN GHS classification categories are described below.

- Identification of GHS NC (Bottom-Up approach):
  184 chemicals (75 substances + 109 mixtures)
  83 NC + 101 classified chemicals (48 Cat 1 + 6 Cat 1/2 + 35 Cat 2A + 10 Cat 2B + 2 Cat 2A/2B)
- Identification of GHS Cat 1 (Top-Down approach):
  172 chemicals (68 substances + 104 mixtures)
  45 Cat 1 + 127 non-Cat 1 (34 Cat 2A + 10 Cat 2B + 83 Non-Classified (NC))

23. Out of the chemicals from the validation dataset (n=175), a number of chemicals (n=15) had no raw in vivo data to allocate a UN GHS Classification. In addition, a number of chemicals (n=13) had Study Criteria Not Met (SCNM) to assign an in vivo classification (e.g., incomplete dataset to assess reversibility / irreversibility of effects at day 21). For a large number of them (n=11), the in vivo scores suggested the need for classification even if not possible to allocate a specific classification category (i.e., GHS Cat 2B versus 2A versus 1). These chemicals were used for the evaluation of the predictive capacity of the ICE test method in a bottom-up approach, but not for the top-down approach due to uncertainty as to which classification category to assign (i.e., GHS Cat 1 versus GHS Cat 2). Chemicals that had a SCNM and were estimated to be non-classified based on expert judgement (n=2), were not included in any of the analyses for precautionary reasons (although in the original evaluation they were considered as NC). A total of 5 chemicals that were classified as Eye GHS Cat 1 based on data from skin corrosion studies were not included for the purposes of the Test Guideline, in order to consider only chemicals for which high quality in vivo ocular data was available. Finally, two chemicals had a borderline GHS Cat 1 / Cat 2 classification so that they could only be used for the evaluation of the predictive capacity of the ICE test method in a bottom-up approach, and not in a top-down approach.

Overall Predictive Capacity

24. Due to discrepancies found in a number of in vitro and in vivo classifications from previous validation studies (for details see Appendixes 2, 3 and 4), the predictive capacities of the ICE test method were re-calculated for i) the identification of GHS Category 1 chemicals (Top-Down approach) and ii) the identification of non-classified chemicals (Bottom-up approach) as shown in Tables 6 and 7. The analyses were based on the outcome of individual test substances (and not on individual laboratory outcome), as recommended by the Expert Group on Eye Irritation, in order to be in alignment with previous ICCVAM evaluations and with the analyses carried out in the context of the revisions of the BCOP Test Guideline. Furthermore, the additional data obtained between 2013 and 2015 were added to the dataset (see Appendix 1).

Table 6. Predictive capacity of the ICE test method for distinguishing chemicals (substances and mixtures) inducing serious eye damage (UN GHS[1] Category 1) from all other categories.

Top-.Down Approach	No.	Accuracy		Sensitivity		False Negatives		Specificity		False Positives	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	172	83	142/172	53	24/45	47	21/45	93	118/127	7	9/127
Without alcohols, solids and surfactants	93	95	88/93	75	9/12	25	3/12	98	79/81	3	2/81

Note: Abbreviations: No. = data used to calculate the percentage.
[1]UN GHS classification system (UN, 2017): Category 1 vs. Non-Category 1 (No Category + Cat. 2B + Cat. 2A).

Table 7. Predictive capacity of the ICE test method for distinguishing chemicals (substances and mixtures) not requiring classification for eye irritation or serious eye damage (UN GHS[1] Non-Classified) from all other irritant categories.

Bottom-Up Approach	No.	Accuracy		Sensitivity		False Negatives		Specificity		False Positives	
		%	No.	%	No.	%	No.	%	No.	%	No.
Overall	184	88	161/184	97	98/101	3	3/101	76	63/83	24	20/83
Without anti-fouling organic-solvent containing paints	181	88	159/181	98	97/99	2	2/99	76	62/82	24	20/82

Note: Abbreviations: No. = data used to calculate the percentage.
[1]UN GHS classification system (UN, 2017): Category 1 vs. Non-Category 1 (No Category + Cat. 2B + Cat. 2A).

25. For the Top-Down approach, Table 8 shows the false positive and false negative rates obtained for specific chemical classes and properties of interest, including mixtures and substances based on the revised dataset (Appendixes 1 and 2). Alcohols were found to risk over-prediction (4 alcohols out of 10 non-Category 1 were over-predicted as Category 1) whereas solids and surfactants were found to risk under-prediction (6 out of 12 Category 1 solids were found to be under-predicted, and 9 out of 17 Category 1 surfactants were found to be under-predicted). Substances, mixtures, liquids and solids all showed false positive rates below or equal to 17% suggesting an appropriate identification of GHS Category 1. The rate of false negatives was found to be particularly high (i.e., higher than 50% for 5 chemicals or more), for solids, surfactants and mixtures in particular for pesticides, herbicides and agrochemicals. However due to the fact that the underpredicted

solids and surfactants would need to be subsequently tested with other suitable test method(s) in a sequential testing strategy (as none of the GHS Cat 1 solid and surfactant materials in the validation database are underpredicted as GHS non-classified), solids and surfactants are not considered to be out of the applicability domain of the ICE test method. Regarding mixtures, histopathology was found to be a useful additional endpoint to decrease the false negative rates when used to identify UN GHS Category 1 non-extreme pH ($2 < pH < 11.5$) detergents and surfactants shown to induce mainly persistent non severe effects in vivo (Cazelle et al., 2014; Appendix 7). Finally, pesticides, herbicides, agrochemicals showed a high rate of both false positives (5/11) and false negatives (7/11). This could be due to the fact that these formulations are complex mixtures in which ingredient(s) may target specific adverse effects on biological processes/pathways (e.g. neurotoxicity), which may lead to an in vivo Cat. 1 classification. An US project is ongoing for more in depth evaluation of agrochemicals. It was agreed by the OECD Expert Group on Eye Irritation to take into consideration the outcome of this study before making final conclusions regarding the applicability of the OECD TG 438 to these types of mixtures.

Table 8. False positive and false negative rates of the ICE test method, by properties of interest, chemical class and product categories, for distinguishing chemicals (substances and mixtures) inducing serious eye damage (UN GHS[1] Category 1) from all other categories.

Top-Down Approach	N[2]	False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.[3]
Overall	172	7	9/127	47	21/45
Properties of interest					
Substances	68	17	7/41	41	11/27
Mixtures	104	3	3/86	56	10/18
Liquids[5]	125	9	8/94	45	14/31
Solids[5]	35	0	0/23	50	6/12
Emulsions and gels[5]	8	14	1/7	0*	0/1*
Chemical Classes[6]					
Alcohol	12	40	4/10	50*	1/2*
Amine/Amidine	5	0*	0/1*	50*	2/4*
Carboxylic acid	10	0*	0/3*	43	3/7
Ester	9	13	1/8	0*	0/1*
Heterocyclic	9	0*	0/3*	50	3/6
Onium compound	8	0*	0/2*	50	3/6
Polyether	5	25*	1/4*	100*	1/1*
Product categories					
Cleaners	12	33*	2/6*	33*	2/6*
Copolymer	8	13	1/8	n.a.	n.a.
Detergent	10	0*	0/7	100*	3/3*
Dyes & stains	8	0	0/7	0*	0/1*
Fungicide / Germicide / Bactericide	7	0*	0/1*	50	3/6
Industrial Chemical, Intermediate or Formulation	16	0	0/14	50*	1/2*
Pesticide / Herbicide / Agrochemical	22	45	5/11	64	7/11
Preservative	5	0*	0/1*	25*	1/4*
Pharmaceutical compound or intermediate	4	50*	1/2*	50*	1/2*
Raw material	8	0	0/8	n.a.	n.a.
Soap	7	0	0/6	100*	1/1*
Solvent	34	19	6/31	0*	0/3*
Surfactant – Total[7]	32	0	0/150/3*	53	9/17
-cationic	14	0*	0/4*	36	4/11
-nonionic	6	0*	0/2*	100*	2/2*
-anionic	4	0*	n.a.	50*	1/2*
- amphoteric	1	n.a.		100*	1/1*

Note: * Too small dataset to make definitive conclusions; n.a.: not applicable.
[1]GHS = Globally Harmonized System (UN, 2017).
[2]N = Number of chemicals.
[3]Data used to calculate the percentage.
[4] Only few formulations having severe effects are available.
[5]Physical form (i.e., solid or liquid) not known for 4 chemicals.
[6]Chemical classes included in this Table are represented by at least five chemicals tested in the ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as described in ICCVAM-NICETAM (2006, 2009)).
[7]Combines single substances labelled as surfactants along with surfactant-containing mixtures.

26. For the Bottom-up approach, Table 9 shows the false positive and false negative rates obtained for specific chemical classes and properties of interest, including mixtures and substances based on the revised and updated dataset (Appendixes 1 and 2).In particular, anti-fouling organic solvent containing paint were found to risk under-prediction (1 UN GHS Cat. 1 anti-fouling solvent containing paint was found to be under-predicted as non-classified). As explained in paragraph 14, a warning sentence was included in the Test Guideline for this category of materials but they were not to excluded from the applicability domain of the ICE test method for the following reasons: i) there was insufficient evidence (only two classified chemicals from this category), ii) the type of exposure is unlikely to occur in humans, and iii) sticky materials present similar difficulties to test either in vitro and in vivo. In addition, two UN GHS Cat. 2 test chemicals were underpredicted (the UN GHS Cat. 2B Ethyl-2-methylacetoacetate (CAS 609-14-3) and the UN GHS Cat. 2A Hand Dishwash Liquid #14 from the A.I.S.E. dataset) . In any case, substances, mixtures, liquids and solids all showed false negative rates below or equal to 4% suggesting an appropriate identification of GHS Non-classified chemicals based on the revised criteria discussed by the OECD Expert Group on Eye Irritation. Regarding the false positive rates, the ICE test method was found to have a low overall false positive rate as compared to other test methods accepted for this purpose (i.e., 24% for the ICE overall dataset versus 69% for BCOP, 37% for RhCE and 19% for STE). The false positive rates was found nevertheless to be particularly high (i.e., higher than 50% for 5 chemicals or more) for surfactants. However due to the fact that the overpredicted surfactants would need to be subsequently tested with other suitable test method(s) in a sequential testing strategy (ICCVAM-NICEATM, 2006) (as none of the in vivo GHS non-classified chemicals (n=83) was overpredicted as GHS Cat. 1), surfactants are not considered to be out of the applicability domain of the ICE test method.

Table 9. False positive and false negative rates of the ICE test method, by properties of interest, chemical class and product categories, for distinguishing chemicals (substances and mixtures) not requiring classification for eye irritation or serious eye damage (UN GHS1 No Category) from all other irritant categories.

Bottom-Up Approach	N[2]	False Positive Rate		False Negative Rate	
		%	No.[3]	%	No.[3]
Overall	184	24	20/83	3	3/101
Properties of interest					
Substances	75	57	12/21	2	1/54**
Mixtures	109	13	8/62	4	2/47
Liquids[4]	131	29	17/59	4	3/72
Solids[4]	39	18	3/17	0	0/22
Emulsions, gels and paste[4]	**10**	**0***	**0/4***	**0**	**0/6**
Chemical Classes[5]					
Alcohol	**13**	**100***	**4/4***	**0**	**0/9**
Amine/Amidine	**6**	**100***	**1/1***	**0**	**0/5**
Carboxylic acid	11	100*	2/2*	0	0/9
Ester	10	100*	4/4*	0	0/6
Heterocyclic	9	100*	2/2*	0	0/7
Onium compound	8	100*	1/1*	0	0/7
Polyether	5	100*	2/2*	0*	0/3*
Product categories					
Cleaners	13	100*	2/2*	0*	0/11*
Copolymer	8	33	2/6	0*	0/2*
Detergent	10	67*	2/3*	14**	1/7**
Dyes & stains	9	43	3/7	0*	0/2*
Fungicide / Germicide / Bactericide	8	100*	1/1*	0	0/7
Industrial Chemical, Intermediate or Formulation	18	50*	2/4*	0	0/14
Paints	4	0*	0/2*	50*	1/2*
Pesticide / Herbicide / Agrochemicals	24	40*	2/5*	0	0/19
Preservative	5	n.a.	n.a.	0	0/5
Pharmaceutical compound or intermediate	6	n.a.	n.a.	0	0/6
Raw material	9	20	1/5	0*	0/4*
Soap	8	25*	1/4*	0*	0/4*
Solvent	35	38	8/21	0	0/14
Surfactant – Total[6]	33	67	6/9	0	0/24
-cationic	14	100*	2/2*	0	0/12
-nonionic	6	100*	2/2*	0*	0/4*
-anionic	4	100*	1/1*	0*	0/3*
- amphoteric	1	n.a.	n.a.	0*	0/1*

Note: * Too small dataset to make definitive conclusions; n.a.: not applicable.
** Represents a Cat. 2 underpredicted as a No Cat.

[1]GHS = Globally Harmonized System (UN, 2017).
[2]N = Number of chemicals.
[3]Data used to calculate the percentage.
[4]Physical form (i.e., solid or liquid) not known for 4 chemicals.
[5]Chemical classes included in this Table are represented by at least five chemicals tested in the ICE test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as described in ICCVAM-NICETAM (2006, 2009).
[6]Combines single substances labelled as surfactants along with surfactant-containing mixtures.

WITHIN- AND BETWEEN-LABORATORY REPRODUCIBILITY OF THE STANDARD ICE TEST METHOD

27. A thorough evaluation of the ICE reproducibility was conducted in the 2003-2006 retrospective validation study (ICCVAM-NICEATM, 2006). Based on a quantitative analysis of within-laboratory reproducibility of the ICE test method endpoints, the evaluation showed CV values for the corneal thickness measurement, when results were compared within experiments, varying from 1.8% to 6.3% (ICCVAM-NICEATM, 2006) (ICCVAM, 2007). The other endpoints evaluated produced ranges of CV values that were larger, with variability most prominent with the non-irritating substance. However, this can be explained by an exaggeration of variability given the relatively small values that were produced by chemicals not requiring classification relative to chemicals inducing eye irritation and serious eye damage (i.e., corneal swelling values of 2, 0, and 3 yield a higher CV than values of 11, 14, and 18). A similar discussion also can be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively).

28. Regarding the between-laboratory reproducibility, the retrospective studies showed median/mean % CV values to be 32%/35% for the Irritation Index, 36%/39% for fluorescein retention, 37/47% for corneal opacity, and 75%/77% for corneal swelling (ICCVAM-NICEATM, 2006). All laboratories were in 100% agreement on the classification of 75% (44/59) of the substances according to the UN GHS classification system for both the top-down (ICCVAM-NICEATM, 2006) and bottom-up approaches (ICCVAM-NICEATM, 2009) according to the UN GHS classification. Finally, the EC/HO study showed the following inter-laboratory correlations between the ICE classification at TNO (lead laboratory) and the classifications obtained in three other laboratories: 0.829, 0.849 and 0.844 (Balls et al., 1995).

29. Specific issues were raised on the between-laboratory variability of the corneal swelling endpoint. This was due to the use of different slit-lamp measuring devices by the participating laboratories of the EC/HO study which, unless normalized, can contribute to the increased variability and/or the excessive values calculated for this endpoint (ICCVAM-NICEATM, 2006). In particular, out of the four participating laboratories, two (that are no longer active in the area of toxicity testing) were reported to use different slit-lamps and different slit width settings resulting in different ranges of values for corneal swelling (see Appendix 4). In order to avoid potential variability issues linked to this endpoint, the use of a specific pachymeter and appropriate slit width, together with the use of proficiency chemicals are requested in both the adopted TG 438 (Prinsen and Koeter, 1993) and the revised TG 438 (i.e., paragraph 53 of OECD TG 438 (2018a): "Corneal swelling scores shown in Table 4 are only applicable if thickness is measured with a Haag-Streit BP900 slit-lamp microscope or alternatively a Haag-Streit BQ900 slit-lamp microscope) with depth-measuring device no. I and slit-width setting at 9½, equalling 0.095 mm. Users should be aware that slit-lamp microscopes could yield different corneal thickness measurements if the slit-width setting is different.").

Considerations on variability for the Bottom-Up approach

30. As shown in Table 7 and described in paragraph 25, only three chemicals (out of 101) were identified as a false negative in the ICE test method for the identification of chemicals not requiring classification for eye irritation or serious eye damage in a Bottom-

Up approach (i.e., the UN GHS Cat. 2B Ethyl-2-methylacetoacetate (CAS 609-14-3), the UN GHS Cat. 2A Hand Dishwash Liquid #14 t and the UN GHS Cat. 1 TNO-94 anti-fouling solvent containing paint). However, a total of seven chemicals that were correctly predicted as causing ocular effects that require a UN GHS classification, were found to be false negatives in some of the participating laboratories (Table 10).

Table 10. **Chemicals showing one or more under-classification in the various participating laboratories.**

N.	Chemical name	*In vivo* GHS Cat.	Physical state	Lab 22	Lab 25	Lab 24	Lab 27	New testing (2016)		Overall *in vitro* class
15	Captan 90	Cat 1	Solid	NC	2	2B	2B	-	-	2B
16	4-Carboxybenzaldehyde	Cat 2A	Solid	2B	1	NC	2	-	-	2A
46	Maneb	Cat 2A (EJ)	Solid	NC	2A	NC	2B	-	-	2B
50	Methyl cyanoacetate	Cat 2A	Liquid	NC	2A	NC	2B	-	-	2B
62	Quinacrine	Cat 1	Solid	NC	NC	2A	NC	2A	2A	2A
71	Sodium oxalate	Cat 1	Solid	2B	2B	NC	NC	-	-	2B
72	Sodium perborate	Cat 1	Solid	NC	2B	2B	2B	-	-	2B

EJ : classification based on expert judgment

31. Over the entire dataset, these chemicals represent 6 solids out of the 22 GHS classified solids present in the ICE overall dataset (i.e., 27%), and 1 liquids out of the 72 GHS classified liquids present in the ICE validation dataset (i.e., 1%). Due to higher probability of solids to have discordant classifications and in a precautionary approach, the revised Test Guideline requires that "In the case of solid materials leading to a GHS No Category outcome, a second run of three eyes is recommended to confirm or discard the negative outcome" (paragraph 23 of OECD TG 438).

USE OF ICE HISTOPATHOLOGY FOR NON-EXTREME PH (2<PH<11.5) DETERGENTS AND SURFACTANTS

32. Currently only criteria for identification of UN GHS category 1 test chemicals have been developed and accepted (OECD TG 438, 2018a). The International Association for Soaps, Detergents and Maintenance Products (A.I.S.E.) conducted an in vitro study from 2010 to 2012 where specific ICE histopathological effects were found to be correlated with serious eye damage classification induced by non-extreme pH detergents and surfactants (Cazelle et al., 2014, 2015). The study initially comprised a total of 30 non-extreme pH (2<pH<11.5) detergents (Cazelle et al., 2014) and 18 extreme pH detergents (pH □ 2 or pH □ 11.5) (Cazelle et al., 2015). Epithelial vacuolation (in mid and lower layers) and epithelial erosion (of at least moderate level) were found to be the most typical histopathological effects induced by the non-extreme pH detergents classified in vivo as UN GHS Cat. 1. Use of these histopathology criteria for non-extreme pH detergents substantially increased the sensitivity of the standard ICE prediction model for UN GHS Cat. 1 identification (from 0% to at least 75%, n=8) whilst maintaining a good concordance (73%, n=30), and an acceptable specificity (from 100% to 73%, n=22). In particular, it allowed correctly identify 5 of 6 non-extreme pH detergents classified as UN GHS Cat. 1

based on in vivo persistence of effects i.e., having tissue effects that do not reverse 21 days after treatment and that do not lead to severity of effects that would warrant a UN GHS Cat. 1 classification (Cazelle et al., 2014). In contrast, for extreme pH detergents, 5 of the 6 tested in vivo UN GHS Cat. 1 were classified in vivo due to severity of effects and not persistence. In this case, the A.I.S.E. histopathology criteria did not improve the sensitivity of the standard ICE test method (83%, n=6), whilst it strongly decreased specificity (from 83% to 33%, n=12), and concordance (from 83% to 50%, n=18) (Cazelle et al., 2015). These data indicate that there seems to be specific applicability domains for the use of the ICE histopathology for non-extreme pH detergents and surfactants that are likely based on the mode of action of the tested detergents and surfactants. The decision criteria developed by A.I.S.E. (OECD GD 160, 2018b) were found to be applicable to non-extreme pH detergents and to surfactants but not to extreme pH detergents even though the rate of false negatives is still 2 out of 8 UN GHS Cat. 1 formulations. Appropriate and relevant data are needed to verify and expand the applicability of the ICE histopathology decision criteria to other chemistries.

33. The reproducibility of histopathological semi-quantitative scorings was conducted between pathologists and peer-reviewers from three laboratories and over time (Appendix 7). Good reproducibility was found between pathologists and peer-reviewers from three independent laboratories of (10/12 or 83%) and over time (17/18 for non-extreme pH detergents and 6/6 for surfactants) for the ICE histopathological derived predictions. However, to ensure such reproducibility, there is a need for:

- an internal peer-review system to be in place;
- assessment of the original slides in order to enable the evaluation of three dimensional effects;
- appropriate training & proficiency appraisal.
- The reproducibility of ICE histopathology between pathologists and their peer-reviewers was determined independently from the reproducibility of the standard ICE test method. This was done in order to understand the contribution from histopathology, independently from the already known reproducibility of the standard ICE test method. Although the overall reproducibility of the standard ICE together with the ICE histopathology has not been characterized in the present study, the OECD Expert Group in November 2017 expressed the opinion that such characterization was not necessary due to the fact that:
- ICE histopathology is used only to further identify UN GHS Cat. 1 non-extreme pH detergents and surfactants, whereas test chemicals predicted as No Prediction Can Be Made remain as such in case the ICE histopathology criteria are not met,
- histopathology may be conducted by a different laboratory than the laboratory conducting the standard ICE test method (i.e., without histopathology),
- that the approach undertaken by A.I.S.E. allows to evaluate the reproducibility of the standard ICE test method and the reproducibility of the ICE histopathology separately.

34. In addition, when histopathology is used as an additional endpoint to identify UN GHS Cat. 1 non-extreme pH ($2 < pH < 11.5$) detergents and surfactants, the false negative rate of the ICE test method and its accuracy (as compared to Draize in vivo data and LVET Cat. 1 data) are improved (from 64% to 27% false negatives (n=22) and from 53% to 77%

accuracy (n=30)), whilst an acceptable false positive rate is maintained (from 0% to 12.5% false positives (n=8)) (Table 11).

Table 11. Predictive capacity values for histopathology for UN GHS Category 1 identification

		ICE	ICE + histo.*	ICE + histo.**
Non-extreme pH (2 < pH < 11.5) detergents[A] and surfactants	False negative rate	63.6% (14/22)	**27.3% (6/22)**	**18.2% (4/22)**
	False positive rate	0.0% (0/8)	**12.5% (1/8)[B]**	**25.0% (2/8)**
	Concordance	53.3% (16/30)	**76.7% (23/30)**	**80.0% (24/30)**

Note:
[A]Includes 8 non-extreme pH detergents from the training set distributed as 5 LVET Cat. 1 (5/5 and 0/5 false negatives with ICE and ICE + histo respectively) and 3 Draize No Cat. 1 non-extreme pH detergents (0/3 and 1/3 false positives with ICE and ICE + histo respectively).
[B]The false positive observed with ICE histopathology is a Draize Cat. 2A non-extreme pH detergent.
* Histopathological criteria as described in the revised OECD GD 160 (2018b).
** Considering epithelial necrosis scores of 1 (instead of 2) as threshold based on the original score system, i.e., only necrosis of attached cells/tissues being scored.

Note about necrosis: When evaluating the histopathology results obtained with surfactants, epithelial necrosis was observed in 8 out of the 20 tested surfactants and in 6 of the 13 UN GHS Cat. 1 tested surfactants. Such findings were not observed with the 48 tested detergents, where epithelial necrosis was found only for 2 out of the 48 tested mixtures, and did not affect the predicted classification based on the ICE histopathology criteria. As a consequence, at the time when the histopathology criteria were originally developed for identification of UN GHS Cat. 1 non-extreme pH detergents, only a limited dataset was available on epithelial necrosis to make conclusive decisions. This meant that only preliminary decision criteria could be established for the epithelial necrosis effects, in contrast to epithelial erosion and vacuolation for which a considerably larger amount of data was available. However, a change of epithelial necrosis (use of threshold of 1 instead of 2) was not supported by the expert working group during the meeting from 9-10 November 2017.

35. Based on the above results, the OECD expert working group meeting from 9-10 November 2017 agreed to revise the OECD TG 438 to include the possibility of using ICE histopathology in addition to the standard ICE test method to identify UN GHS Cat. 1 non-extreme pH (2 < pH < 11.5) detergents and surfactants as described below. However that, based on the dataset currently available, histopathology cannot be used in a stand-alone manner to identify UN GHS Cat. 2 and UN GHS No Cat. test chemicals. Furthermore, additional appropriate and relevant data are needed to verify and expand the applicability of the ICE histopathology decision criteria to chemistries other than non-extreme pH (2<pH<11.5) detergents and surfactants.

Standard ICE results	**ICE histopathology criteria as defined by A.I.S.E.**	**Predicted UN GHS Classification**
No prediction can be made	Criteria met	UN GHS Category 1
	Criteria not met	No prediction can be made

REFERENCES

Adriaens E., Barroso J., Eskes C., Hoffmann S., McNamee P., Alépée N., Bessou-Touya S., De Smedt A, de Wever B., Pfannenbecker U., Tailhardat M., Zuang V. (2014). Draize test for serious eye damage / eye irritation: importance of the endpoints evaluated with regard to UN GHS / EU CLP classification. Archives of Toxicology 88, 701-723.

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

Barroso J., Pfannenbecker U., Adriaens E., Alépée N., Cluzel M., De Smedt A., Hibatallah J., Klaric M., Mewes K.R., Millet M., Templier M., McNamee P. (2017). Cosmetics Europe compilation of historical serious eye damage/eye irritation in vivo data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: the Draize eye test Reference Database (DRD). Archives of Toxicology 91, 521-547.

Cazelle E., Eskes C., Hermann M., Jones P., McNamee P., Prinsen M., Taylor H., Wijnands M.V.W. (2014). Suitability of histopathology as an additional endpoint to the isolated chicken eye test for classification of non-extreme pH detergent and cleaning products. Toxicology In vitro 28, 657-666.

Cazelle E., Eskes C., Hermann M., Jones P., McNamee P., Prinsen M., Taylor H., Wijnands M.V.W. (2015). Suitability of the Isolated Chicken Eye Test for Classification of Extreme pH Detergents and Cleaning Products. Toxicology In vitro 29, 609-616.

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

EC (2008). REGULATION (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal of the European Union L353: 1-1355.

Eskes, C., 2010. Guidance Document on the Application of Alternative methods in the Regulatory Assessment of Chemical Safety Related to Human Eye Irritation and Severe Irritation: Current Status and Future Prospects. FOPH, 65 pp. Available at: [http://www.bag.admin.ch/themen/chemikalien/00253/03225/12694/index.html?lang=fr]

ICCVAM (2007). ICCVAM Test Method Evaluation Report: In Vitro Ocular Toxicity Test Methods for Identifying Ocular Severe Irritants and Corrosives. NIH Publication No. 07-4517. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. Available at: [http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmer.htm].

ICCVAM (2010). ICCVAM Test Method Evaluation Report – Current Status of In Vitro Test Methods for Identifying Mild/Moderate Ocular Irritants: The Isolated Chicken Eye (ICE) Test Method. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No.: 10-7553A. Available at: [http://iccvam.niehs.nih.gov/methods/ocutox/MildMod-TMER.htm]

ICCVAM-NICEATM (2006). Background review document: Current Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method. NIH Publication No.: 06-4513. Research Triangle Park: National Toxicology Program. Available at: [http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_brd_ice.

htm]

ICCVAM-NICEATM (2009). Draft Background Review Document - Current Status of In vitro Test Methods for Identifying Mild/Moderate Ocular Irritants: The Isolated Chicken Eye (ICE) Test Method. National Institute of Environmental Health Sciences, Resesarch Triangle Park, North Carolina, USA.

Kolle S.N., van Cott A., van Ravenzwaay B., Landsiedel R. (2017). Lacking applicability of in vitro eye irritation methods to identify seriously eye irritating agrochemical formulations: Results of bovine cornea opacity and permeability assay, isolated chicken eye test and the EpiOcular™ ET-50 method to classify according to UN GHS. Regulatory Toxicology and Pharmacology 85, 33-47.

Maurer, J.K., Parker, R.D. and Jester, J.V. (2002). Extent of corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. Reg. Tox. Pharmacol. 36: 106-117.OECD (2017). Acute eye irritation/corrosion. OECD Guideline for Testing of Chemicals No. 405. Organisation for Economic Co-operation and Development, Paris. Available at: [http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

OECD (2018a). Isolated Chicken Eye Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage. OECD Guideline for Testing of Chemicals No. 438. Organisation for Economic Co-operation and Development, Paris.

OECD (2018b) Guidance Document on "The Bovine Corneal Opacity and Permeability (BCOP) and Isolated Chicken Eye (ICE) Test Methods: Collection of Tissues for Histological Evaluation and Collection of Data on Non-Severe Irritants. Series on Testing and Assessment no. 160. ENV/JM/MONO(2018)45.

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

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

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

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

Prinsen (2006). The Draize Eye Test and in vitro alternatives; a left-handed marriage? Toxicology In Vitro 20: 78-81.

Scott L., Eskes C., Hoffmann S., Adriaens E., Alépée 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., 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: 1-9.

SSD (2012). Draft Streamlined Summary Document (SSD) from 7 March 2012 (Document n. ENV/JM/TG(2012)20) on "Description of applicability domain and performance based on the retrospective validation studies of the Isolated Chicken Eye (ICE) Test Method (Test Guideline 438) for identifying ocular corrosives and severe irritants.

UN (2017). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Seventh revised edition, UN New York and Geneva, 2017. Available at:

[https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev07/English/03e_part3.pdf]

U.S. EPA (2011). Label Review Manual. Washington, DC: U.S. Environmental Protection Agency. Available at: [http://www.epa.gov/oppfead1/labeling/lrm/].

Weil C.S., Scala A. (1971). Study of intra- and inter- laboratory variability in the results of rabbit eye and skin irritation tests. Toxicology and Applied Pharmacoogy 19, 276-360.