For Official Use **ENV/JM/WRPR(2017)17**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **29-May-2017**

English - Or. English

ENVIRONMENT DIRECTORATE

ENV/JM/WRPR(2017)17
For Official Use

Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology

Draft Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Serious Eye Damage and Eye Irritation

The Joint Meeting is invited to agree to declassify the document by 10 July 2017.

Anne Gourmelon
Principal Adminsitrator, Test Guidelines Programme
Email: Anne.Gourmelon@oecd.org; Tel: +33(0)1 45 24 98 49

JT03414807

Complete document available on OLIS in its original format
This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

English - Or. English

DRAFT GUIDANCE DOCUMENT ON AN INTEGRATED APPROACH ON TESTING AND ASSESSMENT (IATA) FOR SERIOUS EYE DAMAGE AND EYE IRRITATION

SUMMARY:

This document has two aims:

First, it suggests an Integrated Approach on Testing and Assessment (IATA) for serious eye damage and eye irritation hazard identification, in view of replacing the *"sequential testing strategy"*, which is currently provided in the supplement to OECD TG 405 and which requires adaptation to technical progress.

Second, the document provides key information characteristics of each of the individual information sources comprising the IATA. Furthermore it provides guidance on how and when to integrate existing and/or newly generated information for decision making, including decisions on the need for further testing or final decisions on classification and labelling regarding the potential eye hazard effects of test chemicals.

This Guidance Document was approved by the 29th Meeting of the WNT in April 2017.

TABLE OF CONTENTS

1. INTRODUCTION AND SCOPE 6

2. COMPOSITION OF THE IATA FOR SERIOUS EYE DAMAGE AND EYE IRRITATION 8

3. INTEGRATION OF *IN VITRO* TESTS INTO TOP-DOWN AND BOTTOM-UP TESTING APPROACHES 14

4. DESCRIPTION OF THE ELEMENTS OF THE IATA FOR SERIOUS EYE DAMAGE AND EYE IRRITATION 18

4.1. Module 1: Existing human data on serious eye damage and eye irritation 18
4.2. Module 2: *In vivo* animal data according to OECD TG 405 on serious eye damage and eye irritation 20
4.2.1. Description and use of the in vivo rabbit eye test method (OECD TG 405) within the IATA20
4.2.2. Considerations on the intrinsic characteristics of the in vivo rabbit eye test method..... 24
4.3. Module 3: *In vitro* data from OECD adopted test methods on serious eye damage and eye irritation 25
4.3.1. Bovine Corneal Opacity and Permeability (BCOP) test method (OECD TG 437)..... 28
4.3.2. Isolated Chicken Eye (ICE) test method (OECD TG 438) 31
4.3.3. Short Time Exposure (STE) test method (OECD TG 491)..... 35
4.3.4. Reconstructed human Cornea-like Epithelium Eye Irritation Test (RhCE EIT) (OECD TG 492) 37
4.3.5. Fluorescein Leakage (FL) test method (OECD TG 460) 40
4.4. Module 4: Other existing animal data from non-OECD adopted test methods on serious eye damage and eye irritation 42
4.5. Module 5: Other data from non-OECD adopted alternative test methods on serious eye damage and eye irritation 43
4.5.1. OECD Guidance Document 160 on the use of histopathology as an additional endpoint ... 43
4.5.2. In vitro test methods that underwent validation studies 49
4.5.3. Promising optimized in vitro test methods..... 52
4.6. Module 6: Existing human, *in vivo* and *in vitro* data on skin corrosion 53
4.7. Module 7: Physico-chemical properties (existing, measured or estimated) 53
4.8. Module 8: Non-testing data on serious eye damage and eye irritation 55
4.8.1. (Q)SAR, expert systems, grouping and read-across (substances)..... 55
4.8.2. Bridging principles and theory of additivity (mixtures)..... 58
4.8.3. Overview of non-testing data on serious eye damage and eye irritation..... 59
4.9. Module 9: Phases and elements of Weight of Evidence (WoE) approaches 62
4.9.1. Place/role of WoE in the IATA..... 62
4.9.2. Coverage of relevant sources of information 62
4.9.3. Assessment of data quality..... 63
4.9.5. Adequacy of information..... 63
4.9.6. Non-testing data 63
4.9.7. Existing human data..... 64
4.9.8. Evaluation of consistency of the data..... 65
4.9.10. Assessment of the coverage of relevant parameters and observations..... 65
4.9.11. Conclusions of WoE..... 65

5. REFERENCES 66

ANNEX I – EXAMPLE OF MATRIX FOR WEIGHT OF EVIDENCE ANALYSES..... 77

ANNEX II – EXAMPLES OF WEIGHT OF EVIDENCE EVALUATIONS 79

List of acronyms

A.I.S.E. International Association for Soaps, Detergents and Maintenance Products
AOP Adverse Outcome Pathway
BCOP Bovine Corneal Opacity and Permeability (test)
BfR German Federal Institute for Risk Assessment
CAMVA Chorio-Allantoic Membrane Vascular Assay
CAS Chemical Abstracts Service (number)
Cat. Category
C&L Classification and Labelling
CV Coefficient of Variation
DA Defined Approach
DIP Data interpretation procedure
ET_{50} Time of exposure to reduce tissue viability of 50%
GD Guidance Document
GHS Globally Harmonized System for classification and labelling (UN GHS, 2015)
ECHA European Chemicals Agency
EIT Eye Irritation Test
EPAA European Partnership for Alternative Approaches to Animal Testing
EURL ECVAM European Union Reference Laboratory for Alternatives to Animal Testing
ESAC EURL ECVAM Scientific Advisory Committee
EVEIT *Ex Vivo* Eye Irritation Test
EU European Union
FL Fluorescein Leakage (test)
GLP Good Laboratory Practices
HCE Human Corneal Epithelium
HET-CAM Hen's Egg Test on the Chorio-Allantoic Membrane
HPLC High Performance Liquid Chromatography
IATA Integrated Approach on Testing and Assessment
ICCVAM US Interagency Coordinating Committee on Validation of Alternative Methods
ICE Isolated Chicken Eye (test)
IRE Isolated Rabbit Eye (test)
JaCVAM Japanese Centre for the Validation of Alternative Methods
JRC European Commission – Directorate General Joint Research Centre
LVET Low Volume Eye Test
MAGAM Multinational Analysis of data from Poison control centres on corrosive Eye lesions of Automatic dishwashing detergent and other detergent and cleaning products
MDCK Madin-Darby Canine Kidney (cells)
MoA Mode of Action
MTT 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide / Thiazolyl blue tetrazolium bromide
NRR Neutral Red Release (test)
OECD Organisation for Economic Co-operation and Development
PorCORA Porcine Ocular Cornea Opacity/Reversibility Assay

QMRF (Q)SAR Model Reporting Format
QPRF (Q)SAR Prediction Reporting Format
(Q)SAR (Quantitative) Structure-Activity Relationship
RBC Red Blood Cell (test)
REACH EU Regulation 1907/2006 on the Registration, Evaluation, Authorisation and restriction of Chemicals
RhCE Reconstructed human Cornea-like Epithelium
SAR Structure-Activity Relationship
SIRC Statens Seruminstitut Rabbit Cornea (cells)
SMI Slug Mucosal Irritation (test)
STE Short Time Exposure (test)
TG Test Guideline
UN United Nations
UN GHS Cat. 1 Serious eye damage/irreversible effects on eye
UN GHS Cat. 2/2A Eye irritation/reversible effects on the eye
UN GHS Cat. 2B Mildly irritating to eyes
UN GHS No Cat. No category/no need for classification
UPLC Ultra Performance Liquid Chromatography
US United States
US EPA US Environmental Protection Agency
UVCB Substances of Unknown and Variable Composition and Biologicals
WNT OECD Working Group of the National Coordinators of the Test Guidelines Programme
WoE Weight of Evidence

1. INTRODUCTION AND SCOPE

1. The objective of the present Guidance Document (GD) is to establish an Integrated Approach on Testing and Assessment (IATA) for hazard identification of serious eye damage and eye irritation potential of test chemicals[1] (or the absence thereof) that provides adequate information for classification and labelling according to the United Nations Globally Harmonised System (UN GHS, 2015).

2. **Serious eye damage** refers to the production of tissue damage in the eye, or serious physical decay of vision, which is not fully reversible (i.e., within 21 days of application in the rabbit test according to OECD TG 405), occurring after exposure of the eye to a test chemical. Test chemicals that have the potential to induce serious eye damage/irreversible effects on the eye are classified as UN GHS Category 1 (UN, 2017). **Eye irritation** refers to the production of changes in the eye, which are fully reversible (i.e., within 21 days in the rabbit test according to OECD TG 405), occurring after exposure of the eye, to a test chemical (UN, 2017). Test chemicals that have the potential to induce eye irritation/reversible effects on the eye are classified as UN GHS Category 2 (UN, 2015). For regulatory authorities requiring more than one classification for reversible eye irritation, Categories 2A and 2B are used, where Category 2A uses the same classification criteria as Category 2 but in which a Category 2B is assigned when the irritant effects triggering Category 2A effects are fully reversible within 7 days of observation (UN, 2015). Finally, **test chemicals not classified** for eye irritation or serious eye damage are defined as those that do not meet the requirements for classification as UN GHS Category 1 or 2 (2A or 2B), and are referred to as UN GHS No Category (No Cat.) (UN, 2015). A test chemical can be an individual (mono- or multi-constituent) substance or a mixture, and represents what is tested without *a priori* defining the applicability domain for a specific test method.

3. Since 2002, the OECD Test Guideline (TG) 405 on *in vivo* acute eye irritation and corrosion contains a supplement describing a sequential testing and evaluation strategy for eye irritation/corrosion (OECD, 2012a). While this supplement is not covered by the OECD Council decision on Mutual Acceptance of Data (MAD), it has provided valuable guidance on how to consider existing information and organise the generation of new testing data on acute eye hazard effects. In its revised version from 2012, the sequential testing and evaluation strategy calls for the use of validated and accepted *in vitro* and/or *ex vivo* test methods for identification of serious eye damage (UN GHS Cat. 1), eye irritation (UN GHS Cat. 2 or UN GHS Cat. 2A and 2B), and insufficient eye hazard effects to require classification (i.e., UN GHS No Cat.), before conducting an *in vivo* animal test. The use of an *in vivo* animal test is recommended only as a last resort with the purpose of minimising animal use.

4. Since the adoption in 2002 and revision in 2012 of this sequential testing strategy within OECD TG 405, a number of Test Guidelines on *in vitro* methods have been adopted and/or revised for the identification of test chemicals inducing serious eye damage (UN GHS Cat. 1) or for the identification of test chemicals not requiring classification for eye irritation and serious eye damage hazards (UN GHS No Cat.), notably OECD TG 437, TG 438, TG 460, TG 491 and TG 492 (OECD 2012b, 2013a, 2013b, 2015a,

[1] A test chemical can be an individual (mono- or multi-constituent) substance or a mixture, and represents what is tested without a priori defining the applicability domain for a specific test method.

2015b). In addition, methods not adopted by the OECD (i.e., not yet validated, not yet accepted by the OECD or implemented within specific country regulatory requirements) may provide further information required by some authorities, e.g. on specific mechanistic insights such as reversibility of effects and effects on the vascular system. The suitability of such data for regulatory purposes needs to be judged on a case by case basis.

5. Updates to the sequential testing and evaluation strategy supplement within OECD TG 405 are therefore required in view of providing guidance on the use, combination and generation of new data, where required. Furthermore, based on the growing experience with the composition and use of IATAs for this specific human health endpoint (UN, 2015; ECHA, 2015), and the adoption in 2014 of the Guidance Document No. 203 on an Integrated Approach on Testing and Assessment for Skin Corrosion and Irritation (OECD, 2014a), such revision is timely in order to incorporate current scientific and regulatory considerations and practices for the identification of eye hazards.

6. For these reasons, the OECD Working Group of the National Coordinators for the Test Guidelines (WNT) approved in 2015, a project jointly proposed by the US and the European Commission to develop a Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for serious eye damage and eye irritation. The IATA is composed of well described and characterised "Modules", each of which contain one to several individual information sources of similar type. The strengths and limitations as well as the potential role and contribution of each Module and their individual information sources in the IATA for the identification of serious eye damage, eye irritation and no need for classification are described with the purpose of minimizing the use of animals to the extent possible, while ensuring human safety.

2. COMPOSITION OF THE IATA FOR SERIOUS EYE DAMAGE AND EYE IRRITATION

7. The IATA groups the various individual information sources in "modules" according to the type of information provided. Nine modules were identified as relevant elements of the IATA for eye hazard identification, which can be grouped in three major parts as described in Table 1. The different individual information sources associated with each module are described in chapter 4 in a detailed and consistent manner in terms of their applicability, limitations and performance characteristics.

8. The three Parts that guide the assessment of serious eye damage and eye irritation hazards are Part 1 on existing and non-testing data, Part 2 on a weight of evidence analysis, and Part 3 on the generation of new testing data. Under Part 1 of the IATA (existing and non-testing data), existing and available information is retrieved from literature and databases and other reliable sources for Modules 1 to 6, while Module 7 covers physico-chemical properties (primarily pH, which can be existing, measured or estimated) and Module 8 covers non-testing methods, including (Q)SAR, expert systems, grouping and read-across (for substances), and bridging principles and theory of additivity (for mixtures). Part 2 is equivalent to Module 9 and consists of the phases and elements of a weight of evidence (WoE) approach. If the WoE analysis is inconclusive regarding the identification or non-identification of serious eye damage and eye irritation hazard potential, new testing, starting with *in vitro* methods, needs to be conducted in Part 3 (testing data), in which animal testing is foreseen only as a last resort and after considering the newly obtained *in vitro* data together with other available information in a second WoE evaluation.

9. A schematic outline of the IATA for eye hazard identification focusing on classification and labelling (C&L) is presented in Figure 1. Briefly, the collected existing and non-testing information from Part 1 is evaluated in a WoE approach. If the WoE is conclusive, decision for C&L can be taken accordingly. If it is inconclusive, all available information from the WoE should be considered to formulate a hypothesis of the most likely classification for eye hazard potential of the test chemical, i.e. classified (UN GHS Cat. 1, Cat. 2, Cat. 2A or Cat. 2B), no need for classification (UN GHS No Cat.), or high certainty of not inducing serious eye damage (Non-Cat. 1) (see also chapter 3). This hypothesis will then guide the sequence of prospective testing to e.g. a Top-Down or Bottom-Up approach (Scott *et al*., 2010).

Table 1: Parts and Modules of the IATA.

Part (*)	**Modules**
Part 1: Existing information, physico-chemical properties and non-testing methods	1. Existing human data on serious eye damage and eye irritation
	2. Existing *in vivo* animal data according to OECD TG 405 on serious eye damage and eye irritation
	3. Existing *in vitro* data from OECD adopted test methods on serious eye damage and eye irritation a) OECD TG 437 on the BCOP test method b) OECD TG 438 on the ICE test method c) OECD TG 491 on the STE test method d) OECD TG 492 on the RhCE test method e) OECD TG 460 on the FL test method
	4. Other existing animal data from non-OECD adopted test methods on serious eye damage and eye irritation
	5. Other data from non-OECD adopted alternative test methods on serious eye damage and eye irritation
	6. Existing data on skin corrosion (human, animal and *in vitro*)
	7. Physicochemical properties (existing, measured or estimated) such as pH and acid/alkaline reserve
	8. Non-testing data on serious eye damage and eye irritation a) Substances: (Q)SAR, expert systems, grouping and read-across b) Mixtures: bridging principles and theory of additivity
Part 2: WoE analysis	9. Phases and elements of WoE approaches
Part 3: New testing	4.Testing on OECD adopted *in vitro* test methods for serious eye damage and eye irritation
	6. Testing on other non-OECD adopted alternative test methods for serious eye damage and eye irritation
	3. As a last resort, testing on *in vivo* animal test method according to OECD TG 405 for serious eye damage and eye irritation

(*) While the three Parts are considered as a sequence, the order of Modules 1 to 8 of Part 1(here shown in decreasing order of complexity) might be arranged as appropriate. Furthermore, if sufficient and adequate data exist, each module may lead on its own to a classification decision or the absence of classification where relevant, as described Figure 1.

BCOP: Bovine Corneal Opacity and Permeability; FL: Fluorescein Leakage; ICE: Isolated Chicken Eye;; RhCE: Reconstructed human Cornea-like Epithelium; STE: Short Time Exposure; WoE: Weight of Evidence.

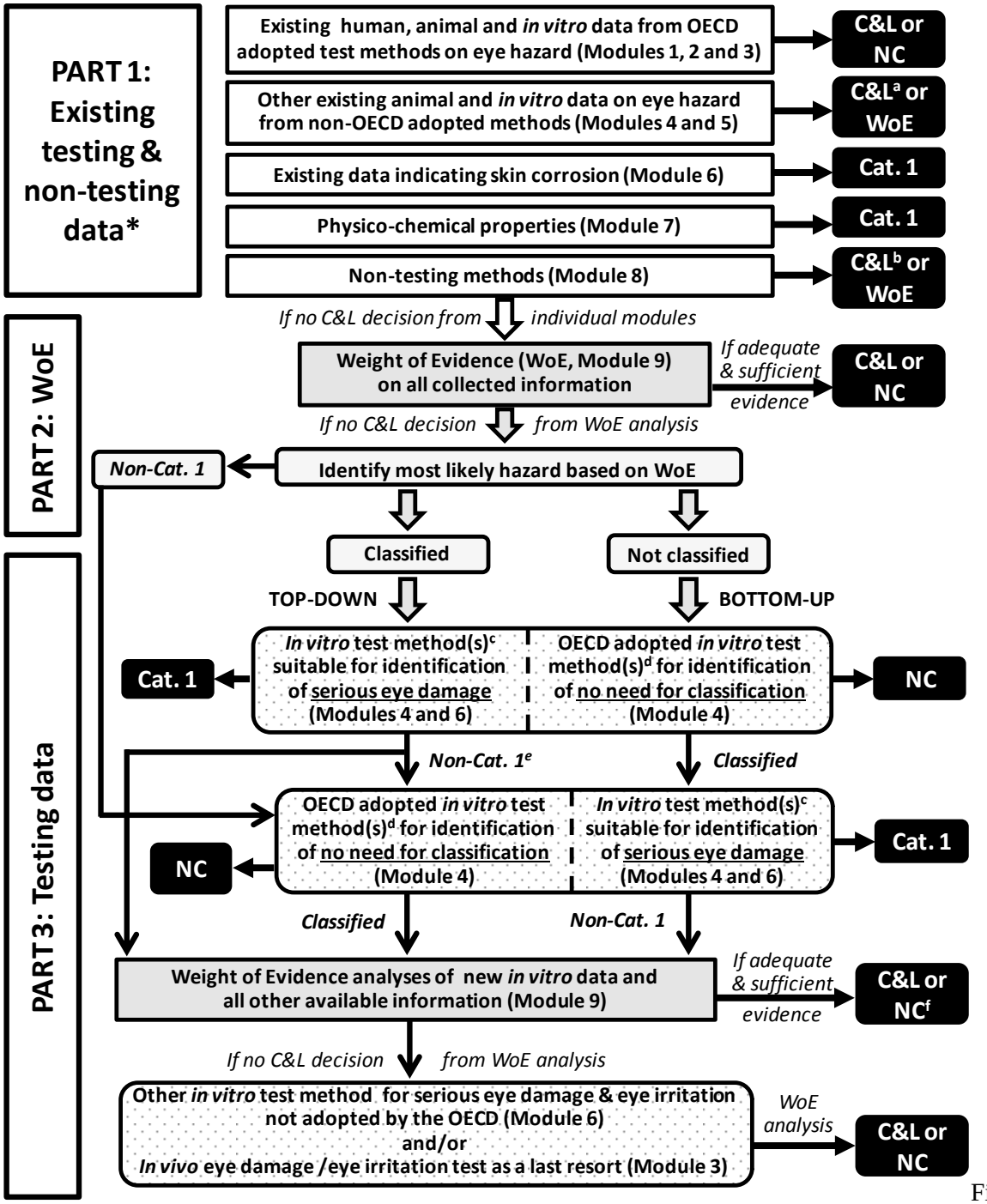

Figure 1: Detailed IATA for serious eye damage and eye irritation. C&L: Classification and labelling (i.e., UN GHS Cat. 1 or Cat. 2); NC: UN GHS No Category.

* While the three Parts are considered as a sequence, the order of Modules 1 to 8 of Part 1 (here shown in decreasing order of complexity) might be arranged as appropriate. Furthermore, if sufficient and adequate data exist, each module may lead on its own to a classification decision or the absence of classification where relevant, as described in the figure.

[a] For example results obtained with other existing *in vivo* test methods (e.g., the FHSA method 16CFR 1500,42 (CPSC. 2003)) might be used to derive a final classification, which might include also identification of UN GHS No Category. Furthermore, results obtained with optimized non-OECD adopted test methods (e.g., Isolated Rabbit Eye

Test) might be used to identify UN GHS Cat. 1 test chemicals. Finally, negative results obtained with optimized non-OECD adopted test methods might be used in a WoE approach.

[b] For example, the application of bridging principles might be used to derive a classification of the tested mixture, which might include also identification of UN GHS No Category. In contrast, results obtained from (Q)SARs might be used in a WoE approach.

[c] The use of additional *in vitro* test methods suitable for identifying UN GHS Cat. 1, based if possible on different mechanisms of action, may be considered in case a negative result is obtained with a first *in vitro* test method used for this purpose. This is due to the fact that a single *in vitro* test method aiming at the identification of UN GHS Cat. 1 may not cover all mechanisms of action resulting in serious eye damage (e.g. persistence of effects) and may therefore produce a certain amount of false negatives (see chapters 3 and 4.3).

[d] The use of additional OECD adopted *in vitro* test methods for identifying UN GHS No Cat. may be considered in case a positive result is obtained with a first *in vitro* test method used for this purpose. This is due to the fact that the currently OECD adopted *in vitro* test method aiming at the identification of UN GHS No Cat. produce a significant amount of false positives (see chapters 3 and 4.3).

[e] In cases where the WoE evaluation in Part 2 indicates that a classification is warranted with a high degree of certainty, testing with an *in vitro* test method for identification of UN GHS No Cat. may be waived, and the next steps in the strategy should be undertaken.

[f] UN GHS Cat. 2 classification is to be considered only in cases where the WoE evaluation indicates that the test chemical is not UN GHS Cat. 1 with a high degree of certainty.

10. The structure provided by the three Parts and the information on the nine Modules described in Table 1 allow for composing an IATA. Ideally, this IATA should be universally applicable and ensure human safety, while making maximum use of existing data, being resource efficient and eliminating or at least minimising the requirement for animal testing.

11. While the three Parts are considered as a sequence, Modules 1 to 8 of Part 1 might be arranged as appropriate. This will be especially helpful in cases in which information on one or a few Modules cannot be outweighed by any other information, so that a conclusion on the eye hazard potential can be drawn without considering further Modules. Existing information on Modules 1 to 6 can be retrieved by a comprehensive literature and database search. Indeed, in recent years, large databases have become available on the internet, e.g., the European C&L Inventory[2] and the dissemination site for chemicals registered under REACH[3]. The search should be performed systematically using search terms such as CAS (Chemical Abstract Service) number or chemical name. Note that in case relevant information is identified, rights to use this information for regulatory purposes may need to be obtained. Whereas Modules 1 to 5 directly relate to eye hazard, Module 6 requires a different search for *in vitro* and *in vivo* skin corrosion data following e.g. its recommended IATA (OECD GD 203, 2014a) that can also impact the final classification of the test chemical.

12. In case the existing information (Modules 1 to 6 within Part 1) does not allow for an unequivocal decision regarding the serious eye damage and eye irritation potential (or the absence thereof) of the substance/mixture, the relevant physico-chemical data and/or non-testing data (i.e., (Q)SAR, expert systems, grouping and read across for substances as well as bridging principles and additivity approach for mixtures) should be considered. If not retrieved from database searches or available estimates are doubtful, pH and potentially acidity and alkalinity reserve, as well as other physico-chemical parameters may be measured. Regarding Module 8 (non-testing methods), the OECD QSAR Toolbox[4] may be considered as a starting point to retrieve information as it allows for (i) the retrieval of a first set of existing experimental (physico-chemical and toxicological) data on the target substance(s), (ii) the identification of analogues

[2] http://echa.europa.eu/web/guest/information-on-chemicals/cl-inventory-database

[3] http://echa.europa.eu/de/information-on-chemicals/registered-substances

[4] http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm

(for read-across) and retrieval of their existing experimental (physico-chemical and toxicological) data and (iii) the characterisation of these substances with mechanistic and other profilers, including structural alerts for serious eye damage and eye irritation. Further existing data on analogues identified with the Toolbox can then be retrieved by repeating the above literature and database search for these compounds. If data from several (Q)SAR models on a substance are already available and are known to disagree, it may not be helpful to generate other (Q)SAR predictions. If, however no (Q)SAR analysis has been performed, the generation of (Q)SAR information might be helpful to supplement the existing data and come to a conclusion on C&L. Importantly, it is always necessary to carefully consider how well the prediction from each (Q)SAR model falls within the applicability domain of that model.

13. In the analysis of the WoE (Module 9), each data element is characterised for its quality, relevance, coverage (e.g., serious eye damage, eye irritation and/or no need for classification) and associated uncertainty. The decision on inclusion or exclusion of each of the different pieces of existing information is to be based on these parameters (see chapter 4.9). When consistency is seen among ”qualified” data elements, WoE may reach a conclusion that the relevant endpoint or information requirement has been sufficiently covered and further testing is not necessary. When on the other hand, insufficient information remains after the "non-qualified" data have been rejected/put aside and/or when the remaining information is inconsistent or contradictory, WoE may lead to a conclusion that further testing is necessary (Part 3 of the IATA), in which case it should also inform on which test(s) to conduct to fill the identified gap(s) (see chapter 3).

14. The WoE assessment needs to be transparently explained and documented to enable a logical flow especially if leading to a final decision/conclusion on classification and labelling. While a WoE approach implies the weighing of each available piece of information on a case by case basis, the modules included in the IATA differ *a priori* with respect to their intrinsic weight e.g. based on considerations of relevance relating to the species of interest or biological and mechanistic aspects. The following relative *a priori* weights are nevertheless indicative only and will depend on the quality of the individual data in each specific case. Typically, the relative *a priori* weights of the modules can be expected to be as follows, based on regulatory acceptance of data when it is of equal quality:

- Good quality and relevant existing human data (Module 1) would be expected to carry the highest weight when the adverse ocular effect and its magnitude can be reliably attributed to the test chemical of interest, however most often such information is not available so that human data on eye hazard effects are generally rather used in a WoE approach.

- This is followed by, with equal weights, *in vivo* rabbit data according to OECD TG 405 (Module 2) and *in vitro* data from OECD adopted test methods (Module 3). In particular, it is important to critically appraise the intrinsic characteristics (e.g., uncertainty, variability, drivers of classification) of both the *in vivo* and the *in vitro* test methods of Modules 2 and 3 (see chapters 4.2.2 and 4.3).

- Other *in vivo* animal and *in vitro* data from non-OECD adopted test methods on serious eye damage and eye irritation (Modules 4 and 5), data indicating skin corrosion (Module 6), physico-chemical information (Module 7) and non-testing methods (Module 8) would typically carry less intrinsic weight.

An example for a simple approach for documenting a WoE evaluation is presented in Annex 1, and examples of evaluations are given for detergents and agrochemical mixtures in annex 2.

15. Before conducting prospective testing for serious eye damage and eye irritation hazard identification, it is strongly recommended to i) consider all existing available test data and ii) generate

information whenever possible by means of alternative methods to animal testing such as *in vitro* methods, (Q)SAR models, grouping or read-across. Evaluating existing data is key to avoid unnecessary animal testing. It can also represent a time and cost efficient way to derive a conclusion on serious eye damage and eye irritation hazard potential, if the available data allow for it.

16. Acknowledging that the applicability of the individual information sources of this IATA to mixtures may vary and that such applicability may depend on the information available in each specific case to be assessed, the IATA is considered applicable to both substances and mixtures. Indeed, data on mixtures can be used for all modules relating to the testing and/or non-testing of eye hazard effects, i.e., modules 1 to 5 and modules 7 to 8 (for details see chapter 4 and paragraph 22).

17. The individual sources of information described in Modules 1 to 8 (Table 1) have been characterised as described in chapter 4 and comprise the following information headlines:

- Regulatory use (UN GHS Classification), i.e., the UN GHS Classification that can be derived from individual information sources;
- Validation and regulatory acceptance status;
- Potential role in the IATA;
- Description;
- Scientific basis including Mode of Action (MoA);
- Protocol available;
- Strengths and weaknesses;
- Applicability domain and limitations;
- Predictive capacity, e.g., expressed as sensitivity, specificity and accuracy;
- Reliability, e.g., expressed as within- and between-laboratory reproducibility.

3. INTEGRATION OF *IN VITRO* TESTS INTO TOP-DOWN AND BOTTOM-UP TESTING APPROACHES

18. It is generally acknowledged that a single *in vitro* test method is not able to cover the criteria for injury and inflammation addressed by the regulatory adopted *in vivo* animal test method, i.e. the *in vivo* rabbit eye test as described in OECD TG 405. Therefore, in order to replace or to reduce the use of the *in vivo* rabbit eye test, it is recommended to make use of testing strategies that combine the strengths of individual *in vitro* test methods to address the required ranges of irritation potential and/or chemical classes (Scott *et al*., 2010). In particular, two tiered testing approaches as shown in Figure 1 are recommended for serious eye damage and eye irritation hazard identification:

- **A Top-Down approach**, starting with *in vitro* test methods that can identify test chemicals causing serious and/or irreversible eye damage (UN GHS Cat. 1) with low false positive predictions and the highest possible accuracy.
- **A Bottom-Up approach**, starting with *in vitro* test methods that can identify test chemicals not requiring classification for eye hazard (UN GHS No Cat.) with low false negative predictions and the highest possible accuracy.

19. These tiered testing approaches can be considered as Defined Approaches (DAs) to Testing and Assessment and can be used as a component within the IATA. According to the OECD GD 255 a Defined Approach to testing and assessment is a standardised strategy consisting of a defined set of information sources (*in silico* and/or *in vitro*) and a fixed Data Interpretation Procedure (DIP) that is applied to the combined data obtained from the information sources to derive predictions of toxicological effects that can be used either on their own, or together with other information sources within an IATA, to satisfy a specific regulatory need (OECD, 2016). The concept of DIP is taken from the OECD guidance document 34 (OECD, 2005), and is defined in this context as any algorithm for interpreting data from one or more information sources. The defined approach to testing and assessment can be used to support the hazard identification, hazard characterisation and/or safety assessment of chemicals and can be used either on its own to reach a conclusion, or together with other sources of information within an IATA (OECD, 2016). With a view to facilitating the evaluation of IATA in regulatory decision-making, the OECD GD 255 provides with a set of principles and a template for reporting defined approaches to testing and assessment. Such template enables a transparent, structured and harmonised approach to document the defined approaches to testing and assessment. These templates should be used alongside the reporting formats for other IATA components, such as QSARs (OECD, 2007), grouping and read-across strategies (OECD, 2014b) and non-guideline test methods (OECD, 2014c).

20. All available information and the WoE assessment should be used to formulate a hypothesis of the most likely eye hazard potential of the test chemical, e.g. likelihood to induce serious eye damage (UN GHS Cat. 1) or likelihood of no need for classification for eye hazard (UN GHS No Cat.). This hypothesis and the regulatory context under which a decision must be taken should then guide the choice of the prospective testing approach and test methods to be used. The Bottom-Up approach should be followed only when all available collected information and the WoE assessment result in a high *a priori* probability that the test chemical does not require classification for eye hazard (UN GHS No Cat.). The Top-Down approach, on the other hand should be used when all available collected information and the WoE assessment result in a high *a priori* probability that the test chemical may induce serious eye damage (UN GHS Cat. 1) or a likelihood for the test chemical to be an eye irritant (UN GHS Cat. 2). Independently of the strategy undertaken, further *in vitro* testing will depend on the results obtained in the first test following

the strategies as shown in Figure 1. Only in case of a high likelihood for the test chemical to be an eye irritant (UN GHS Cat. 2) but not to induce serious eye damage (non-Cat. 1), the initial *in vitro* test method in a top-down approach for identification of UN GHS Cat. 1 may be waived. In this case though, the next step in the tiered strategy should be undertaken (i.e., testing with an OECD adopted test method for the identification of UN GHS No Cat. test chemical), followed in case of a positive result by a second WoE analyses to determine the most likely eye hazard classification, or the further testing with additional *in vitro* test methods (see paragraphs 24 to 26).

21. Recommended testing options include the OECD adopted *in vitro* test methods as described in Module 3 (OECD TG 437 on the BCOP test method, OECD TG 438 on the ICE test method, OECD TG 460 on the FL test method, OECD TG 491 on the STE test method and OECD TG 492 on the RhCE test method). It is generally acknowledged that when the applicability and limitations of the *in vitro* test methods adopted by the OECD are adequately considered, these methods can, irrespective of the starting point, be used to identify chemicals i) inducing serious eye damage (UN GHS Cat. 1); or ii) chemicals not requiring classification for eye hazard (UN GHS No Cat.). Note that some test methods such as OECD TG 437 on BCOP, TG 438 on ICE and TG 491 on STE may be used to initiate the top-down and the bottom-up approaches at the same time, because they are able to provide both UN GHS Cat. 1 and No Cat. predictions, so that the two tiers of the strategy could be covered with one single *in vitro* assay, provided the test chemical fits the applicability domain and does not fall within the limitations of the test method for each tier (see Table 2). However, a test chemical that is neither predicted as UN GHS Cat. 1 nor as UN GHS No Cat. in the bottom-up or top-down approach would require further testing with optimised *in vitro* methods not yet adopted by the OECD (Module 5) as described e.g. in paragraphs 24 to 26. If results obtained with these optimised *in vitro* methods not yet adopted by the OECD may be used to identify UN GHS Cat. 1 test chemicals, other outcomes can only be used in a new WoE evaluation to be conducted with the newly generated *in vitro* data together with the existing information (see Figure 1). *In vivo* testing is to be used only as a last resort if still required e.g. by regulators to establish a definitive classification (UN GHS Cat. 1, Cat. 2 (Cat. 2A or Cat. 2B if applicable) or No Cat.).

22. The currently adopted *in vitro* test methods (OECD TGs 437, 438, 460, 491 and 492) are applicable to both substances and mixtures. Indeed, OECD TGs 437 (BCOP), 438 (ICE), 491 (STE) and 492 (RhCE) have undergone evaluation studies conducted on both substances and mixtures (OECD 2013a, 2013b, 2015a, 2015b). Examples of mixtures tested include agrochemicals, detergent and cleaning products, anti-microbial cleaning products, cosmetics and personal care products, surfactant-based mixtures, petroleum products and other mixtures (OECD 2013c, 2013d, 2015a, 2015b). The only exception is the test method falling within OECD TG 460 which has undergone a validation study mainly based on substances, but is nevertheless considered to be applicable to the testing of mixtures (OECD 2012b). In cases where evidence can be demonstrated on the non-applicability of the Test Guideline to a specific category of mixture, based on the chemistry and/or physico-chemical property, the Test Guideline should not be used for that specific category. While agrochemical formulations have successfully been tested using the EpiOcular EIT (OECD TG 492) for the identification of UN GHS No Cat., the BCOP (OECD TG 437) was found to be under-predictive for identification of UN GHS Cat 1 agrochemical mixtures (Kolle *et al*., 2015). This could be due to the fact that the majority of the tested agrochemical mixtures (n=19 out of 21) in this study were classified *in vivo* based on persistence of effects only, which is a type of effect known not to be identified *per se* by the currently OECD adopted *in vitro* methods aiming at the identification of UN GHS Cat. 1 (see chapter 4.3).

23. The applicability domain and performance of the OECD adopted individual test methods are described in their respective Modules in chapter 4 and are summarized in Table 2. When using adopted *in vitro* test methods, it is critical to ensure using the most appropriate OECD TG for the specific purpose and chemical to be tested. In particular, the applicability domain plays an important role in the choice of the test method to be used. For example, test methods having the highest possible accuracy for the chemical

class tested should be preferentially used. Similar care should be taken in case optimised *in vitro* test methods not yet adopted by the OECD are used and information on applicability domain is available on these test methods. In addition, it is important to take into account the mechanistic insights provided by *in vitro* test methods, and how those cover the mechanisms taking place in the *in vivo* test method (see paragraph 24 and chapter 4.2.2). Finally, when using two or more test methods (see paragraphs 24 and 25), the conditional independence of these test methods should be considered. This can help to decide which test methods to be included in the Top-Down / Bottom-Up approaches and optimise the overall performance of the approach chosen.

24. One of the problems associated with the originally proposed two-tier Top-Down/Bottom-Up testing strategy (Scott *et al*., 2010) is that a default UN GHS Cat. 2 classification after only testing in two test methods would generate a significant number of false negative (Cat. 1 underclassified as Cat. 2) and false positive (No Cat. overclassified as Cat. 2) results (see Table 2). Currently accepted methods for identifying UN GHS Cat. 1, like BCOP and ICE, underpredict 14-48% of the *in vivo* Cat. 1 chemicals, mostly those inducing persistent effects without occurrence of initial high level injuries (classified *in vivo* based only on persistence of effects). Therefore, since the single *in vitro* test methods aiming at the identification of UN GHS Cat. 1 may not cover all mechanisms of action resulting in serious eye damage (e.g. persistence of effects) and can produce a certain amount of false negatives (see chapter 4.3), the use of additional *in vitro* test methods suitable for identifying UN GHS Cat. 1 based, if possible, on different mechanism of actions, may be considered in case a negative result is obtained with a first *in vitro* test method used for this purpose. Moreover, it is clear that due to the very high sensitivity required by regulatory authorities for accepting the use of *in vitro* test methods to identify chemicals not requiring hazard classification and labelling for serious eye damage/eye irritation (UN GHS No Cat.), their specificity will never go beyond 60-80% (the highest the specificity, the more limited the applicability). RhCE test methods, ICE and STE are those showing the best accuracy for identifying UN GHS No Cat. chemicals and their specificity is only 63-81% with already a few false negatives being obtained (sensitivity around 95%). In such a scenario, several methods capable of identifying UN GHS No Cat. chemicals with very high sensitivity will need to be combined to increase the overall specificity of the testing strategy to acceptable values. Therefore, since the currently OECD adopted *in vitro* test methods aiming at the identification of UN GHS No Cat. produce a significant amount of false positives (see chapter 4.3), the use of additional OECD adopted *in vitro* test methods for identifying UN GHS No Cat. may be considered in case a positive result is obtained with a first *in vitro* test method used for this purpose. In addition to the OECD adopted *in vitro* test method, the use of optimised non-OECD adopted *in vitro* test methods and/or endpoints, as described in chapter 4.5, may be used to identify UN GHS Cat. 1 test chemicals, or to be considered as complementary information in a WoE evaluation for the identification of other eye hazard categories.

25. Some examples on the use of the proposed testing strategy approach have been reported. In particular for antimicrobial and cleaning products, the US Environmental Protection Agency (EPA) recommends the use of a testing approach for determining the appropriate eye hazard classification and labelling. The strategy, which represents a replacement of the *in vivo* data requirement, utilizes a decision tree involving the use of the BCOP, EpiOcular time-to-toxicity (ET_{50}) and Cytosensor Microphysiometer test methods (US EPA, 2015). Other potential ways of combining *in vitro* tests methods in testing strategies based on the concept of the Bottom-up and Top-down approaches have been investigated by Kolle *et al*. (2011) and Hayashi *et al*. (2012a, 2012b). Both studies showed that combinations of methods in Defined Approaches (DAs) can lead to better predictions as compared to each individual test method on its own. Kolle *et al*. (2011) combined EpiOcular™ EIT and BCOP in a two-tier Bottom-up/Top-Down test strategy and Hayashi *et al*. (2012b) combined EpiOcular™ EIT, BCOP, STE and HET-CAM in a two-stage Bottom-Up tiered approach. Based on the data presented in these two publications, Schaeffer and co-workers (2014) showed that specificity for identifying UN GHS No Cat. chemicals can increase substantially by combining in a test strategy several methods able to identify UN GHS No Cat. test

chemicals (including both OECD adopted and non-adopted test methods). This occurs as a result of multiple methods complementing each other by correctly identifying different sets of UN GHS No Cat. chemicals. Interestingly the authors show that the increase in specificity of the test strategy as compared to the individual methods is not accompanied by a significant decrease in sensitivity due to the very high sensitivity already displayed by all of these methods on their own. Furthermore, the accuracy for the identification of UN GHS Cat. 2 by default at the end of the strategy would be significantly improved (Schaeffer *et al*., 2014).

26. Indeed, it is generally recognized that when using the Top-Down and Bottom-Up approaches, the main difficulty lies in predicting the middle category of irritancy (e.g. UN GHS Cat. 2, Cat. 2A or Cat. 2B). The optional use of additional *in vitro* test method(s) may be helpful in improving the prediction of UN GHS Cat. 2. This could be due to an increased accuracy of a default Cat. 2 prediction by decreasing the number of false positives when identifying No Cat. and by decreasing the number of false negatives when identifying Cat. 1. Nevertheless further work and data are needed to reach an acceptable level of predictivity for UN GHS Cat. 2 chemicals. For example, conduct of statistical modelling (taking into consideration the conditional independence of the test methods as described in paragraph 23) may allow to define the desirable performances of the *in vitro* test methods that may, when combined in e.g., Defined Approaches to testing and assessment, and used within the appropriate applicability domain and regulatory context, be used to derive a default UN GHS Cat. 2 prediction if neither a UN GHS Cat. 1 nor a UN GHS No Cat. prediction can be made.

27. The *in vivo* rabbit eye test (OECD TG 405) should be conducted only as a last resort after all the existing information in Part 1 of the IATA has been considered, and after the *in vitro* testing in Part 3 has been conducted and evaluated in an additional WoE evaluation together with the existing data. The *in vivo* animal test, if e.g. required by regulators, should be considered after *in vitro* testing only when:

i) the test chemical is not directly identified as UN GHS Cat. 1 or as UN GHS No Cat. by the *in vitro* test methods and WoE assessment cannot conclude with high enough confidence if the test chemical is Cat. 1, Cat. 2 (or Cat. 2A or Cat. 2B, if applicable), or No Cat. Depending on country-specific regulatory requirements, test methods not yet adopted by the OECD should also be considered both prospectively and in the WoE evaluation.

ii) the test chemical cannot be tested with the *in vitro* test methods due to the limitations of the test methods or when falling outside of the applicability domain of the test method.

4. DESCRIPTION OF THE ELEMENTS OF THE IATA FOR SERIOUS EYE DAMAGE AND EYE IRRITATION

28. The individual sources of information to be used in Modules 1 to 8 (Table 1) and the elements of the weight of evidence evaluation of the collected information to be conducted in Module 9, within the IATA for the hazard identification of serious eye damage and eye irritation potential of test chemicals (or the absence thereof), have been characterised and are described below.

4.1. Module 1: Existing human data on serious eye damage and eye irritation

29. Existing human data include historical data that should be taken into account when evaluating intrinsic hazards of test chemicals. New testing in humans for hazard identification purposes is not acceptable for ethical reasons. Existing data can be obtained from single or repeated exposure(s) from case reports, poison information centres, medical clinics, occupational experience, epidemiological studies and volunteer studies. Note however, that the availability of the epidemiological studies for this endpoint is likely to be rare and the quality often questionable. The quality and relevance for hazard assessment of the existing human data should be critically reviewed. For example, in occupational studies with mixed exposure it is important that the test chemical causing serious eye damage or eye irritation is accurately identified. There may also be a significant level of uncertainty in human data due to poor reporting and lack of specific information on exposure. However, well-documented existing human data from various sources can provide useful information on serious eye damage and eye irritation hazard potential of a test chemical, sometimes for a range of exposure levels. For example, the MAGAM[5] study, first conducted by a retrospective collection of data from poison control centres (in Germany, Austria and Switzerland) between 1998 and 2007 (Stürer *et al*., 2010), led to the MAGAM II prospective study conducted in 2013-2015, in which the criteria for data collection were defined prior to the start of the study to ensure high quality of the collected data from the poison centres. MAGAM II represents a multicentre study aimed at collecting and evaluating data on human eye exposures to detergents and maintenance products from a number of poison control centres[6], which includes, among other, information on severity of effects, duration and outcome.

30. Good quality and relevant human data can be used to determine serious eye damage or eye irritation potential of a test chemical and have precedence over other data. However, absence of reported ocular incidents in humans is no evidence in itself for no classification. The usefulness of the human data on adverse ocular effects will depend on the extent to which the effect, and its magnitude, can be reliably attributed to the test chemical of interest. Examples of how existing human data can be used in hazard classification for ocular effects have been reported (MAGAM II study; ECETOC, 2002). In humans, an ophthalmic examination by a physician would reveal a decay of vision. If it is not transient but persistent it implies classification in Category 1. If the discrimination between Category 1 and Category 2 is not obvious, then Category 1 might be chosen; however, other types of information may be generated e.g. by performing *in vitro* testing, to support the final hazard classification conclusion.

[5] MAGAM is the acronym for Multinational Analysis of data from Poison control centres on corrosive Eye lesions of Automatic dishwashing detergent and other detergent and cleaning products.

[6] Located in Germany, Austria, Italy, Denmark, Czech Republic and Slovakia.

Module 1 – Existing human data on serious eye damage and eye irritation	
General description	
Regulatory use (UN GHS classification)	Human data from accident (e.g. from hospitals) or poison control centre databases can provide evidence for UN GHS Cat. 1 and Cat. 2 classification. However, absence of incidents is not in itself evidence for no classification as exposures are generally unknown or uncertain.
Validation & regulatory acceptance status	Existing human data include historical data that should be taken into account when evaluating intrinsic hazards of test chemicals. New testing in humans for hazard identification purposes is not acceptable for ethical reasons.
Potential role in the IATA	Good quality and relevant human data would be expected to have precedence over other data when the adverse ocular effect and its magnitude can be reliably attributed to the test chemical of interest, however most often such information is not available so that human data on eye hazard effects are generally rather used in a WoE approach. Furthermore, absence of incidence in humans does not necessarily overrule *in vitro* data or existing animal data of good quality that are positive. Finally, if the discrimination between Category 1 and Category 2 is not obvious other types of information may be generated e.g. by performing *in vitro* testing, to support the final hazard classification conclusion.
Description	Ophthalmic examination by a physician revealing a decay of vision, which if not transient but persistent, implies classification in Category 1. If the discrimination between Category 1 and Category 2 is not obvious, then Category 1 might be chosen.
Scientific basis incl. MoA	All MoA are potentially covered.
Protocol available	No standard protocol is available. However efforts have been undertaken to standardize collection of data from poison centres (e.g., MAGAM II study). Existing human data might be derived (e.g., in occupational, consumer, transport, or emergency response scenarios) from single or repeated exposure(s) from case reports, poison information centres, medical clinics, occupational experience, epidemiological studies and volunteer studies. Note however, that the availability of the epidemiological studies for this endpoint is likely to be rare and the quality often questionable.
Strengths and weaknesses	<u>Strengths</u> - Relevant data as obtained directly from the species of interest (humans). - Examples available on how existing human data can be used (MAGAM II study; ECETOC, 2002). <u>Weaknesses</u> - Not standardised. - Mostly based on accidental/uncontrolled exposure often in combination with co-exposure, leading to a high level of uncertainty. - Sufficient data to evaluate the actual exposure (duration and dose) might not be always available.

Module 1 – Existing human data on serious eye damage and eye irritation	
	- Data might be incomplete, insufficient or inaccurate. - Data on the reversibility of the effect might not be always available. - Data on additional, potentially confounding factors (e.g., purity, health status of the affected person, additional exposures) might not be available. - No UN GHS criteria for C&L based on human data are available.
Identification of UN GHS Cat. 1 and Cat. 2	
Applicability domain and limitations	Applicability domain - All test chemicals for which a clear and direct effect on the eye can be concluded from the available data (note that the exposure scenario and chemical identity (needed for concluding on a direct effect) are often not clearly defined in data obtained from accidental exposure). Limitations - Rarely available and, if available, not often with the necessary quality to be used on its own for C&L decisions, so that it is most often used in a WoE evaluation with other existing data to make C&L decisions.
Predictive capacity	The usefulness of human data will depend on the amount and quality of the available information. It is often associated with a high level of uncertainty due to lack of critical information such as chemical identity and purity, exposure scenario (dose and duration), health status of the persons exposed and/or the reported symptoms.
Reliability	Difficult to assess due to uncontrolled exposures (dose and timings) and reporting, although efforts exist to improve quality of data collection from poison centres (e.g., MAGAM II study).

4.2. Module 2: *In vivo* animal data according to OECD TG 405 on serious eye damage and eye irritation

4.2.1. Description and use of the in vivo rabbit eye test method (OECD TG 405) within the IATA

31. The OECD TG 405 (OECD, 2012) on *in vivo* Acute Eye Irritation/Corrosion testing recommends the use of rabbits as preferred species. It was originally adopted in 1981, and revised in 2002 to include i) a supplement on a sequential testing and evaluation strategy for eye hazard identification, ii) use of dermal irritation/corrosion test data to predict eye corrosion prior to considering the conduct of an *in vivo* animal test and iii) the possibility to rinse solid materials from the eyes 1 hour after treatment (instead of the previous 24 hours). In 2012 the TG was further revised to include the possibility to use topical anaesthetics, systemic analgesics, and humane endpoints during *in vivo* animal testing to avoid most or all pain and distress without affecting the outcome of the test.

32. *In vivo* animal testing should not be considered until all available data relevant to the eye hazard potential (or absence thereof) of a test chemical have been evaluated in a WoE analysis according to the present IATA, and the necessary prospective *in vitro* testing conducted as described in chapter 3 (see also Figure 1). This includes conducting a study on the skin corrosion potential of the test chemical before the *in vivo* animal test on serious eye damage and eye irritation. In cases where the *in vivo* animal test is

required, it is recommended that it is performed in a sequential manner using initially one animal. If the results of this initial test with one animal indicate the test chemical to induce serious eye damage, further testing should not be performed. If serious eye damage is not observed in the initial test, the irritant or negative response should be confirmed using up to two additional animals. However, if an irritant effect was observed in the initial test the confirmatory test should be conducted in one animal at a time, rather than exposing the two additional animals simultaneously. It may not be necessary to test a total of three animals if classification of the test chemical can be achieved using only two animals. Finally, due consideration should be made to the intrinsic characteristics of the *in vivo* rabbit eye test method as described in chapter 4.2.2.

Module 2 – *In vivo* animal data on serious eye damage and eye irritation according to OECD TG 405	
General description	
Regulatory use (UN GHS classification)	Classification decision on serious eye damage (UN GHS Cat. 1), eye irritation (UN GHS Cat. 2, Cat. 2A and Cat. 2B), and no need for classification (UN GHS No Cat.).
Validation & regulatory status	The animal test method adopted in OECD TG 405 was never formally validated but has been the historical regulatory test method for testing serious eye damage and eye irritation hazard potential of test chemicals.
Potential role in the IATA	In case *in vivo* animal test data of adequate quality are available, these should carry a certain intrinsic weight in the context of a WoE analysis, taking into consideration the critical appraisal of the intrinsic characteristics (e.g., uncertainty, variability, drivers of classification) of the *in vivo* rabbit test method as described in chapter 4.2.2. The *in vivo* animal test should be conducted only as a last resort after i) considering results from the *in vivo* and/or *in vitro* skin corrosion test method, ii) considering and evaluating all available information relevant to the serious eye damage and eye irritation hazard potential of the test chemical in a WoE analysis (Parts 1 and 2 of the IATA as described in Table 1 and Figure 1), and iii) considering the results obtained with prospective *in vitro* testing (Modules 3 and 5 of Part 3 of the IATA as described in Table 1 and Figure 1).
Description	The test chemical is applied in a single dose (0.1 mL for liquids or an amount corresponding to a volume of 0.1 mL or a weight of not more than 100 mg for solids, pastes and particulate substances) to the conjuctival sac of one of the eyes of the experimental animal (albino rabbit is the preferred animal species) whereas the untreated eye serves as control. Degree of serious eye damage and eye irritation is assessed by scoring lesions to cornea (opacity), iris and conjunctiva (redness and oedema) at specific time intervals and the duration of the study should be sufficient to evaluate the reversibility or irreversibility of the effects. The UN GHS classification is based on the mean tissue scores obtained (as recorded per animal) at 24, 48 and 72 hours after exposure, and on the reversibility or irreversibility of effects observed for up to 21 days. Other effects in the eye and possible adverse systemic effects are also assessed to provide a complete evaluation of the effects.
Scientific basis incl. MoA	The test method allows assessing: - Serious eye damage, i.e. the production of tissue damage in the eye, or serious physical decay of vision, which is not fully reversible within 21 days of application, and

Module 2 – *In vivo* animal data on serious eye damage and eye irritation according to OECD TG 405	
	- Eye irritation, i.e. the production of changes in the eye, which are fully reversible within 21 days of application. The mechanisms by which such effects are produced and detected in the *in vivo* animal test method are multiple and depend on the type of chemicals tested. Regarding the cornea, these may include as a first step cell disruption, denaturation and swelling of collagen. This is followed in a second step by the production and release of intermediates that initiate the process of inflammation, causing the oedema in corneal stroma and invasion of leukocytes. In a third step, regeneration of epithelium may gradually occur resulting in decreased corneal opacity. Finally in some cases as a fourth step, destruction of cornea and stromal ulceration may occur 2 to 3 weeks after injury, mediated by hydrolytic enzymes coupled with inadequate collagen synthesis. When the cornea has reepithelialised or when the corneal stroma becomes totally vascularised, corneal ulceration ceases (Berta, 1992, Pfister, 1983; McCulley, 1987; Lemp, 1974). Other mechanisms of injury detected by the test method include i) inflammation of the conjunctivae in which the dilation of blood vessels can cause redness, and the increased effusion of water can cause oedema/chemosis, and ii) secretion of mucous leading to an increase in discharge. Iritis can also occur either as a direct effect or as a secondary reaction due to the corneal injury. Once iris is inflamed, infiltration of fluids can follow which can affect visual acuity accompanied by symptoms of itching, burning and stinging. Finally, other possible mechanisms of injury covered by the *in vivo* animal test method include: i) loss of corneal innervations, ii) tear film abnormalities due to injury to the lacrimal glands, iii) intense pain, lacrimation, and blepharospasm due to direct stimulation of free nerve endings located in the epithelium of the cornea and conjuctival lining, iv) neurogenic inflammation. Irreversible effects may occur when the damage extends to and beyond the corneal endothelium causing corneal perforation that may cause permanent loss of vision. Other persistent effects include discolouration of the cornea by a dye chemical, adhesion, pannus, and interference with the function of the iris or any other effects that impair sight which do not reverse within the test period.
Protocol available	OECD TG 405 (2012)[7] based on the scoring system developed by Draize and co-workers (1944).
Strengths and weaknesses	<u>Strengths</u> - The *in vivo* animal test method reflects all possible modes of action of serious eye damage and eye irritation reactions present in rabbit eyes. - It formed the basis for the GHS classification system, and can therefore identify the entire spectrum of eye effects i.e., UN GHS No Cat., Cat. 2 (and the UN GHS Cat. 2A and 2B), and Cat. 1. - Reversibility and/or persistence of effects can be directly observed. <u>Weaknesses</u>

[7] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 2 – *In vivo* animal data on serious eye damage and eye irritation according to OECD TG 405	
	- Not formally validated. - The possibility of concluding Cat. 1 on the basis of a single eye exposure, which, depending on the type of effect(s) observed, can be associated with a very high uncertainty. - Reproducibility compromised by e.g.: • Subjectivity in the allocation of the ocular tissue scores; • Unclear duration and amount of exposure of the test chemical in the rabbit eyes which can vary depending on the properties of the test chemical (solid, paste or liquid) as well as the blinking and tear reflex from the animal (Prinsen, 2006); • Differences in animal behaviour (e.g., lacrimation, blinking, etc) which can lead to differences in reactions even before scoring of effects takes place (Prinsen, 2006);; • Absence (or presence) of post-treatment care. - For certain test chemicals (e.g., solid, sticky), blinking can result in mechanical damage, contributing to a higher degree of irritation (Prinsen, 2006). - Enclosure of test materials in the conjunctival cul-de-sac in combination with mechanical damage can lead to exacerbation of effects and secondary inflammation not directly caused by the test chemical (Prinsen, 2006). - The animal type of exposure does not reflect human accidental exposure scenarios (Wilhelmus, 2001). - There are differences in physiology and sensitivity to test chemicals between rabbit and human eyes. - Poor correlation was found between rabbit and human mean time to clear (Freeberg *et al*., 1986b) - The testing can be very painful to the rabbits.
Identification of UN GHS Cat. 1, Cat. 2 (A and B) and No Cat.	
Applicability domain and limitations	The test method is applicable to substances, mixtures and aerosols.
Predictive capacity	Differences in physiology and sensitivity exist between rabbit and human eyes, and the *in vivo* rabbit test has been shown to be in general more sensitive to hazard chemicals than the eyes of humans (Roggeband *et al*., 2000; Gershbein and McDonald, 1977; Wilhelmus, 2001; ILSI, 1996). More recently this has been shown to be particularly the case for test chemicals inducing serious eye damage (Ishii *et al*., 2013).
Reliability	Taking into account the animal within-test variability only, at least 11% of chemicals classified *in vivo* as UN GHS Cat. 1 could be equally identified as Cat. 2 by the *in vivo* rabbit eye test itself, and about 12% of the Cat. 2 chemicals could be equally identified as non-classified chemicals (Adriaens *et al*., 2014). If variability between repeat studies were taken into account, the observed concordance of UN GHS classifications when considering a unified Cat 2 classification was found to be of 65.2 % (15/23) (Barroso *et al*., 2017).. If Cat 2A and Cat 2B are considered as different classifications, the observed concordance of UN GHS classifications was found to be 56.5 % (13/23). Finally concordance of

Module 2 – *In vivo* animal data on serious eye damage and eye irritation according to OECD TG 405	
	the same main driver of classification (see chapter 4.2.2) was found to occur for 39.1 % (9/23) of the chemicals (Barroso *et al*., 2017). Furthermore, evaluation of public data from ECHA online dossiers on 9,782 *in vivo* rabbit eye studies on 3,420 unique substances, showed that the most reproducible outcomes were for the negative results (94% reproducible) and for chemicals inducing serious eye damage (73% reproducible), whereas there was a 10% chance of a non-irritant evaluation be given after a prior severe-irritant result based on the UN GHS classification criteria (Luechtefeld *et al*., 2016).

4.2.2. Considerations on the intrinsic characteristics of the in vivo rabbit eye test method

33. In a recent study by Adriaens *et al*. (2014), co-sponsored by the European Commission and Cosmetics Europe, statistical resampling of *in vivo* rabbit test data (according to OECD TG 405) on 2134 chemicals demonstrated an overall probability of at least 11% that chemicals classified as UN GHS Cat. 1 by the *in vivo* rabbit eye test could be equally identified as UN GHS Cat. 2 and of about 12% for UN GHS Cat. 2 chemicals to be equally identified as UN GHS No Cat. simply due to the test method's inherent within-test variability. On the other hand, the chances for UN GHS No Cat. and UN GHS Cat. 2 test chemicals to be predicted in a higher UN GHS Category - was found to be negligible (< 1%). Altogether, these observations suggest that the classification criteria of the *in vivo* rabbit eye test are highly sensitive on their own (Adriaens *et al*., 2014). Taking into account the variability between repeat studies, an overall concordance of 65.2 % (15/23) was found for the UN GHS Cat. 1, a unified Cat. 2 and No Cat. classifications (Barroso *et al*., 2017). If Cat 2A and Cat 2B were considered as different classifications, an overall concordance of 56.5 % (13/23) was found for the UN GHS Cat. 1, Cat. 2A, Cat. 2B and No Cat. classifications (Barroso *et al*., 2017). An evaluation of public data from ECHA online dossiers on 9,782 *in vivo* rabbit eye studies on 3,420 unique substances, further showed that the most reproducible outcomes were for negative results (94% reproducible) and chemicals inducing serious eye damage (73% reproducible), whereas there was a 10% chance of a non-irritant evaluation be given after a prior serious eye damage result based on the UN GHS classification criteria (Luechtefeld *et al*., 2016). Considering these results, it is probably not achievable to develop *in vitro* test methods with no false negatives.

34. The results of the study by Adriaens and colleagues (Adriaens *et al*., 2014) also indicate that the persistence and severity of corneal opacity play an equally important role in the classification of a chemical as UN GHS Cat. 1, whereas corneal opacity and conjunctival redness are the most important tissue effects that determine the classification of UN GHS Cat. 2 eye irritants. In a study co-sponsored by the European Commission and Cosmetics Europe, a further evaluation was performed to establish which of the *in vivo* rabbit eye test drivers of classification are most important from a regulatory point of view for driving UN GHS classification (Barroso *et al*., 2017). For this purpose a *in vivo* rabbit eye test Reference Database was compiled containing 681 independent *in vivo* studies on 634 individual chemicals representing a wide range of chemical classes. The analyses confirmed the previous results from Adriaens *et al*. (2014) by showing that corneal opacity is the most important tissue effect driving Cat. 1 classification (including corneal opacity mean ≥ 3 (days 1-3, severity) and corneal opacity persistence on day 21 in the absence of severity), whereas Cat. 2 classification was found to be mostly driven by corneal opacity mean ≥ 1 and conjunctival redness mean ≥ 2. Based on the evidence presented in the manuscript, the authors identified a number of key criteria that should be taken into consideration when selecting reference chemicals for the development, evaluation and/or validation of alternative methods and/or strategies for serious eye damage/eye irritation testing. Such understanding is critical for properly assessing their predictive capacity and limitations. Furthermore, a critical revision of the UN GHS decision criteria for the classification of

chemicals from the *in vivo* rabbit eye test data was proposed by Adriaens *et al*. (2014) and Barroso *et al*. (2017) based on the results of their analyses of historical *in vivo* data.

4.3. Module 3: *In vitro* data from OECD adopted test methods on serious eye damage and eye irritation

35. The present chapter provides a description of the *in vitro* information sources and their use within the IATA for serious eye damage and eye irritation. A number of *in vitro* test methods have been adopted since 2009 to identify i) test chemicals inducing serious eye damage (UN GHS Cat. 1), and/or ii) test chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Cat.). An overview of the regulatory use, applicability, limitations and performance of the OECD adopted *in vitro* test methods for eye hazard identification is given in Table 2. When using classification systems other than the UN GHS, the appropriate regulatory authorities should be consulted.

36. As compared to the *in vivo* rabbit eye test (OECD TG 405, 2012), the currently available *in vitro* information sources do not directly assess effects on the iris although it should be noted that effects on the iris are of lesser importance for classification of test chemicals according to UN GHS (Adriaens *et al*., 2014; Barroso *et al*., 2017). Furthermore, the neurogenic components that drive tear film production are usually not present in the *in vitro* test methods. As a consequence, when compared with an *in vivo* rabbit eye study, application of a test chemical in the absence of this protective barrier might be expected to cause an increase in false positive outcomes. Finally, the adopted *in vitro* test methods do not allow for an assessment of the potential for systemic toxicity associated with ocular exposure. 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 that have been already characterised for the presence/absence of systemic toxicity effects.

Table 2: Regulatory use, applicability, limitations and performance of the OECD adopted *in vitro* test methods for eye hazard identification.

	BCOP (OECD TG 437)	**ICE (OECD TG 438)**	**STE (OECD TG 491)**	**RhCE (OECD TG 492)**	**FL (OECD TG 460)**
Identification of UN GHS Category 1					
Applicability	Substances and mixtures	Substances and mixtures	Substances, multi-constituent substances and mixtures that are dissolved or uniformly suspended for at least 5 minutes	Not applicable	Water soluble substances and mixtures
Limitations	- Alcohols and ketones risk overprediction	- Alcohols risk overprediction	No other specific limitation reported	Not applicable	Strong acids and bases, cell fixatives, highly volatile test chemicals, coloured and viscous test chemicals, solid chemicals suspended in liquid that have the tendency to precipitate
Accuracy*	79% (150/191)	86% (120/140)	83% (104/125)	Not applicable	77% (117/151)
False positive rate* (1-specificity)	25% (32/126)	6% (7/113)	1% (1/86)	Not applicable	7% (7/103)
False negative rate* (1-sensitivity)	14% (9/65)	48% (13/27)	51% (20/39)	Not applicable	56% (27/48)
Identification of UN GHS No Category					
Applicability	Substances and mixtures	Substances and mixtures	Substances, multi-constituent substances and mixtures that are dissolved or uniformly suspended for at least 5 minutes	Substances and mixtures. Test chemicals interfering with MTT measurement (by i.e., colour interference or reduction of MTT) require the use of appropriate controls or HPLC-UPLC analysis if colour incompatibility with MTT higher than 60% is reported.	Not applicable
Limitations	Due to high false positive rates, BCOP	- Anti-fouling organic solvent-containing paints	- Highly volatile substances with vapour pressure > 6 kPa (at 25°C)	.	Not applicable

	should not be the first choice method to initiate a Bottom-up approach	may be under-predicted - For solid materials leading to a GHS No Cat. outcome, a second testing run is recommended	- Solid chemicals (substances and mixtures) other than surfactants and mixtures of surfactants only - Mixtures containing substances with vapour pressure > 6kPa may risk underpredictions		
Accuracy*	69% (135/196)	82% (125/152)	90% (92/102)	80% (n=112)	Not applicable
False positive rate* (1-specificity)	69% (61/89)	33% (26/79)	19% (9/48)	37% (n=55)	Not applicable
False negative rate* (1-sensitivity)	0% (0/107)	1% (1/73)	2% (1/54)	4% (n=57)	Not applicable

* As reported in the respective Test Guidelines.

BCOP: Bovine Corneal Opacity and Permeability; FL: Fluorescein Leakage; ICE: Isolated Chicken Eye; RhCE: Reconstructed human Cornea-like Epithelium; STE: Short Time Exposure.

4.3.1. Bovine Corneal Opacity and Permeability (BCOP) test method (OECD TG 437)

37. The OECD TG 437 on the BCOP test method was originally adopted in 2009 and updated in 2013 (OECD, 2013a; OECD, 2013c). The BCOP test method underwent two retrospective validation studies by the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), in conjunction with the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM), in 2006 and 2010 (ICCVAM, 2006; ICCVAM, 2010b). In the two evaluations, the BCOP was endorsed as a scientifically valid test method for use as a screening test to identify UN GHS Cat. 1 substances and mixtures (ICCVAM, 2006, 2010b; ESAC, 2007). Furthermore the second evaluation study and a further retrospective evaluation of the *in vitro* and *in vivo* dataset used in the validation study concluded that the BCOP test method can also be used to identify UN GHS No Category substances and mixtures (ICCVAM 2010b; OECD, 2013c). From these evaluations it was concluded that the BCOP test method can correctly identify test chemicals (both substances and mixtures) inducing serious eye damage (UN GHS Cat. 1) as well as those not requiring classification for eye hazard (UN GHS No Category), and it was therefore endorsed as scientifically valid for both purposes.

Module 3 – *In vitro* data: Bovine Corneal Opacity and Permeability (BCOP) test method (OECD TG 437)	
General description	
Regulatory use (UN GHS classification)	Identification of i) test chemicals inducing serious eye damage (UN GHS Cat. 1), and ii) test chemicals not requiring classification for eye hazard (UN GHS No Cat.).
Validation & regulatory acceptance status	Validated and adopted as OECD TG 437[8].
Potential role in the IATA	While the BCOP test method is not considered valid as a full-replacement replacement for the *in vivo* rabbit eye test, it can be used for regulatory classification and labelling (Figure 1) to identify, without further testing: - test chemicals inducing serious eye damage (UN GHS Cat. 1); and - test chemicals that do not require classification for eye hazard (UN GHS No Cat.). BCOP can be used to initiate a Top-Down approach. However, since it has a high overprediction rate for test chemicals that do not require classification for eye hazard (69%), it should not be the first choice to initiate a Bottom-Up approach. Other validated and accepted *in vitro* methods with similar high sensitivity but higher specificity should be used instead as first tier of a Bottom-Up approach (e.g., OECD TG 492). A test chemical that is neither predicted as UN GHS Cat. 1 nor as UN GHS No Cat. with the BCOP test method would require additional WoE evaluation with other existing information and if still needed additional testing (*in vitro and/or in vivo*) as a last resort to establish a definitive classification (see Figure 1).

[8] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 3 – *In vitro* data: Bovine Corneal Opacity and Permeability (BCOP) test method (OECD TG 437)	
Description	The BCOP test method is an organotypic *ex vivo* assay that makes use of isolated corneas from the eyes of freshly slaughtered cattle placed on corneal holders. Test chemicals are applied to the epithelial surface of the cornea by addition to the anterior chamber of the corneal holder. Damage by the test chemical is assessed by quantitative measurements of: - Corneal opacity changes, measured as the amount of light transmission through the cornea with the help of an opacitometer[9]; and - Permeability, measured as the amount of sodium fluorescein dye that passes from the medium in the anterior chamber of the corneal holder, across the full thickness of the cornea, to the medium in the posterior chamber, detected with the help of a visible light spectrophotometer. Both measurements are used to calculate an *In vitro* Irritancy Score (IVIS). An IVIS score higher than (>) 55 leads to a UN GHS Category 1 prediction; an IVIS score smaller than or equal to (≤)3 leads to UN GHS No Category prediction. If in contrast 3 < IVIS ≤ 55, no prediction can be made on the UN GHS classification.
Scientific basis incl. MoA	The BCOP test method addresses corneal effects, which are one of the major drivers of classification *in vivo* when considering the UN GHS classification (Adriaens *et al*., 2014; Barroso *et al*., 2017). Furthermore, it addresses the following Modes of Action for eye irritation: (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), and (iii) coagulation (precipitation/denaturation of macromolecules, particularly protein, characteristic of acid, alkali, or organic solvent exposure). If histopathological information is available, it may also address (iv) actions on macromolecules (chemicals that react with cellular constituents/organelles that may or may not lead to overt lysis or coagulation, e.g., alkylation, oxidative attack on macromolecules such as essential proteins or nucleic acids) (OECD, 2013c).
Protocol available	OECD Guidance Document 160 (OECD, 2011).
General reproducibility	Evaluation of the BCOP reliability showed a median coefficient of variation (CV) for IVIS for replicate corneas (n=3) within individual experiments ranging from 11.8% to 14.2% in one study, and median CV values for IVIS for replicate corneas (n=4) within individual experiments of 35%, respectively, in a second study (ICCVAM, 2006). The between experiment mean CV values of IVIS for 16 chemicals tested two or more times in three laboratories ranged from 12.6% to 14.8%, while the median CV values ranged from 6.7% to 12.4% (ICCVAM, 2006).
Strengths and weaknesses	<u>Strengths</u> - Officially validated test method. - Quantitative and objective measurements of opacity and permeability. - Controlled exposure conditions, including test chemical concentration and exposure duration. - Histological evaluation of the exposed eyes may provide additional information about e.g., the depth and type of injury (Furukawa *et al*., 2015; Maurer *et al*., 2002; OECD, 2011)

[9] A new project has been submitted to the OECD for the use of an alternative Laser Light Based Opacitometer.

Module 3 – *In vitro* data: Bovine Corneal Opacity and Permeability (BCOP) test method (OECD TG 437)	
	Weaknesses - The BCOP test method is not recommended for the identification of test chemicals that should be classified as irritant to eyes (UN GHS Cat. 2 or Cat. 2A) or test chemicals that should be classified as mildly irritant to eyes (UN GHS Cat. 2B) due to the considerable number of UN GHS Cat. 1 chemicals underclassified as UN GHS Cat. 2, 2A or 2B and UN GHS No Cat. chemicals overclassified as UN GHS Cat. 2, 2A or 2B. For this purpose, further testing with another suitable method may be required. - The reversibility of tissue lesions cannot be evaluated *per se* in the BCOP test method, although use of histological evaluations could aid predictions as to e.g., whether damage is irreversible (Furukawa *et al*., 2015; Maurer *et al*., 2002). - The BCOP was found to be under-predictive for identification of UN GHS Cat 1 agrochemical formulations (Kolle *et al*., 2015). - Gases and aerosols have not been assessed yet in a validation study.
Identification of UN GHS Category 1	
Applicability domain and limitations	Applicability - The BCOP test method can be used for the testing of substances and mixtures (OECD, 2013a, 2013c, US EPA, 2015). Limitations - Positive results obtained with alcohols and ketones should be interpreted cautiously due to potential overprediction. However, since not all alcohols and ketones are overpredicted by the BCOP test method and some are correctly predicted as UN GHS Cat. 1, these two organic functional groups are not considered to be out of the applicability domain of the test method. - Solids and chemicals inducing persistent, non severe effects may risk underprediction (OECD, 2013a; Barroso *et al*., 2017). However, none of the false negatives resulted in IVIS ≤ 3 (criterion triggering UN GHS No Cat. prediction). Moreover, BCOP false negatives in this context are not critical since all test chemicals that produce an 3 < IVIS ≤ 55 would be subsequently tested and evaluated following the sequential testing strategy as described in chapter 3 and Figure 1. Finally, given the fact that some solid chemicals are correctly predicted by the BCOP test method as UN GHS Cat. 1, this physical state is also not considered to be out of the applicability domain of the test method. - Increased corneal permeability in the absence of corneal opacity, or in the presence of low grade corneal opacity, e.g. as observed following exposure of the bovine corneas to some types of substances (such as some surfactants and detergent products), should be carefully considered, possibly along with histopathological data, as this might indicate potential for eye hazard effects (OECD, 2011; ICCVAM, 2006)
Predictive capacity	When used for identification of UN GHS Cat. 1 test chemicals, the BCOP test method showed an overall accuracy of 79% (150/191), a false positive rate of 25% (32/126), and a false negative rate of 14% (9/65) when compared to results obtained with the *in vivo* rabbit eye test method (OECD TG 405) classified according to the UN GHS classification system (OECD, 2013a).

Module 3 – *In vitro* data: Bovine Corneal Opacity and Permeability (BCOP) test method (OECD TG 437)	
Reliability	When distinguishing UN GHS Cat. 1 chemicals from the other UN GHS eye hazard categories, 72% (91/127) of the test chemicals were found to have 100% of agreement of classification between laboratories (ICCVAM, 2010b).
Identification UN GHS No Category	
Applicability domain and limitations	Applicability - The BCOP test method can be used for the testing of substances and mixtures (OECD, 2013a, 2013c, Kolle *et al*., 2016). Limitations - Since the BCOP test method can only identify correctly 31% of the test chemicals that do not require classification for eye irritation or serious eye damage, this test method should not be the first choice to initiate a Bottom-Up approach. Other validated and accepted *in vitro* methods with similar high sensitivity but higher specificity should be used instead as first tier of a Bottom-Up approach (e.g., RhCE test methods falling within OECD TG 492). Nevertheless, although the false positive rate obtained with BCOP is considerably high (69%), it is not considered critical since all test chemicals that produce an 3 < IVIS ≤ 55 would be subsequently tested and evaluated following the sequential testing strategy as described in chapter 3 and Figure 1.
Predictive capacity	When used for the identification of UN GHS No Cat. test chemicals, the BCOP test method showed an overall accuracy of 69% (135/196), a false positive rate of 69% (61/89), and a false negative rate of 0% (0/107), when compared to the *in vivo* rabbit eye test method (OECD TG 405) data classified according to the UN GHS classification system (OECD, 2013a).
Reliability	When distinguishing UN GHS No Cat. chemicals from chemicals classified for eye hazard (UN GHS Cat. 1 and 2), 80% (103/128) of the test chemicals were found to have 100% agreement of classification between laboratories (ICCVAM, 2010b).

4.3.2. Isolated Chicken Eye (ICE) test method (OECD TG 438)

38. The OECD TG 438 on the ICE test method was originally adopted in 2009 and updated in 2013 (OECD, 2013b; OECD, 2013d). The ICE test method underwent two retrospective validation studies by the US ICCVAM in conjunction with EURL ECVAM and JaCVAM, in 2006 and 2010 (ICCVAM, 2006; ICCVAM, 2010b). In the two evaluations, the ICE was endorsed as a scientifically valid test method for use as a screening test to identify UN GHS Cat. 1 substances and mixtures (ICCVAM, 2006, 2010b; ESAC, 2007). A further retrospective evaluation of the *in vitro* and *in vivo* dataset used in the validation study concluded that the ICE test method can also be used to identify UN GHS No Category substances and mixtures (OECD, 2013d). From these evaluations it was concluded that the ICE test method can correctly identify test chemicals (both substances and mixtures) inducing serious eye damage (UN GHS Cat. 1) as well as those not requiring classification for eye hazard (UN GHS No Category).

Module 3 – *In vitro* data: Isolated Chicken Eye (ICE) test method (OECD TG 438)	
General description	
Regulatory use (UN GHS classification)	Identification of i) test chemicals inducing serious eye damage (UN GHS Cat. 1), and ii) test chemicals not requiring classification for eye hazard (UN GHS No Cat.).
Validation & regulatory acceptance status	Validated and adopted as OECD TG 438[10].
Potential role in the IATA	While the ICE test method is not considered valid as a full-replacement for the *in vivo* rabbit eye test, it can be used to initiate either the Top-Down or the Bottom-Up approach for regulatory classification and labelling (Figure 1) to identify, without further testing: - test chemicals inducing serious eye damage (UN GHS Cat. 1); and - test chemicals that do not require classification for eye hazard (UN GHS No Cat.). A test chemical that is neither predicted as UN GHS Cat. 1 nor as UN GHS No Cat. with the ICE test method would require additional WoE evaluation with other existing information and if still needed additional testing (*in vitro and/or in vivo*) as a last resort to establish a definitive classification (see Figure 1).
Description	The ICE test method is an organotypic *ex vivo* assay based on the short-term maintenance of chicken eyes *in vitro*. Toxic effects to the cornea are measured by (i) a quantitative measurement of increased corneal thickness (swelling), (ii) a qualitative assessment of corneal opacity, (iii) a qualitative assessment of damage to epithelium based on application of fluorescein to the eye (fluorescein retention), and (iv) a qualitative evaluation of macroscopic morphological damage to the surface. If histopathological information is available, it may also address depth of injury and predict reversibility of effects (OECD, 2011; Maurer *et al*., 2002; Cazelle *et al*., 2014) as well as possible actions on macromolecules (chemical effects on cellular constituents/organelles that may or may not lead to overt lysis or coagulation due to e.g., alkylation, oxidative attack on macromolecules such as essential proteins or nucleic acids) (Scott *et al*., 2010).” The corneal swelling, opacity and damage assessments following exposure to a test chemical are assessed individually and assigned a qualitative categorization, that are then combined together to derive an *in vitro* eye hazard classification, either as UN GHS Cat. 1 or as UN GHS No Cat. However, no decision on classification can be made for test chemicals not predicted to be UN GHS Cat. 1 or UN GHS No Cat. with the ICE test method.
Scientific basis incl. MoA	The ICE test method addresses corneal effects, which are one of the major drivers of classification *in vivo* when considering the UN GHS classification (Adriaens *et al*., 2014; Barroso *et al*., 2017). Furthermore, it addresses the following Modes of Action for eye irritation: (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), and (iii) coagulation

[10] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 3 – *In vitro* data: Isolated Chicken Eye (ICE) test method (OECD TG 438)	
	(precipitation/denaturation of macromolecules, particularly protein, characteristic of acid, alkali, or organic solvent exposure). If histopathological information is available, it may also address (iv) actions on macromolecules (chemicals that react with cellular constituents/organelles that may or may not lead to overt lysis or coagulation, e.g., alkylation, oxidative attack on macromolecules such as essential proteins or nucleic acids) (Scott *et al*., 2010).
Protocol available	OECD Guidance Document 160 (OECD, 2017).
General reproducibility	Evaluation of the ICE reliability showed coefficient of variation (CV) values for the corneal thickness measurement, when results were compared within experiments, varying from 1.8% to 6.3% (OECD, 2013d). The other endpoints evaluated produced larger ranges of CV values due to the relatively small values that were produced by test chemicals not requiring classification. Regarding the between-laboratory reproducibility, the EC/HO international validation study on alternatives to the in *vivo* rabbit eye test showed inter-laboratory correlations of 82.9, 84.9 and 84.4% (OECD, 2013d).
Strengths and weaknesses	<u>Strengths</u> - Officially validated test method. - Measurements are performed both quantitatively and qualitatively with the help of a slit-lamp. - Controlled exposure conditions, including test chemical concentration and exposure duration. - Histological evaluation of the exposed eyes may provide additional information about e.g., the depth and type of injury and reversibility of effects (Maurer *et al*., 2002; OECD, 2011; Cazelle *et al*., 2014) <u>Weaknesses</u> - The ICE test method is not recommended for the identification of test chemicals that should be classified as irritant to eyes (UN GHS Cat. 2 or Cat. 2A) or test chemicals that should be classified as mildly irritant to eyes (UN GHS Cat. 2B) due to the considerable number of UN GHS Cat. 1 chemicals underclassified as UN GHS Cat. 2, 2A or 2B and UN GHS No Cat. chemicals overclassified as UN GHS Cat. 2, 2A or 2B. For this purpose, further testing with another suitable method may be required. - The reversibility of tissue lesions cannot be evaluated *per se* in the ICE test method. However, histological evaluation could aid predictions as to e.g., whether damage is irreversible (Cazelle *et al*., 2014; Maurer *et al*., 2002). - Gases and aerosols have not been assessed yet in a validation study.
Identification of UN GHS Category 1	
Applicability domain and limitations	<u>Applicability</u> - The ICE test method can be used for the testing of substances and mixtures (OECD 2013b, 2013d; Cazelle *et al*., 2014, 2015). - It is applicable to solids, liquids, emulsions and gels. Liquids may be aqueous or non-aqueous and solids may be soluble or insoluble in water. <u>Limitations</u> - Positive results obtained with alcohols should be interpreted cautiously due to potential overprediction. However, since not all alcohols are overpredicted by the ICE test method and some are correctly predicted as UN GHS Cat. 1, this

Module 3 – *In vitro* data: Isolated Chicken Eye (ICE) test method (OECD TG 438)	
	organic functional groups is not considered to be out of the applicability domain of the test method. - Solids, surfactants and chemicals inducing persistent, non severe effects may risk underprediction (OECD, 2013b; Barroso *et al*., 2017). However, false negative rates in this context (UN GHS Cat. 1 identified as not being UN GHS Cat. 1) are not critical since all test chemicals that come out negative would be subsequently tested and evaluated following the sequential testing strategy as described in chapter 3 and Figure 1. Furthermore use of histopathology may help to decrease the under-prediction of non-extreme pH detergents (Cazelle *et al*., 2014).
Predictive capacity	When used for identification of UN GHS Cat. 1 test chemicals, the ICE test method showed an overall accuracy of 86% (120/140), a false positive rate of 6% (7/113) and a false negative rate of 48% (13/27) when compared *to in vivo* rabbit eye test method (OECD TG 405) classified according to the UN GHS classification system (OECD, 2013b).
Reliability	When distinguishing UN GHS Cat. 1 from the other UN GHS eye hazard categories, a between-laboratories reproducibility of 75% (44/59) was observed (ICCVAM, 2006).
Identification UN GHS No Category	
Applicability domain and limitations	<u>Applicability</u> - The ICE test method can be used for the testing of substances and mixtures (OECD 2013b, 2013d). - It is applicable to solids, liquids, emulsions and gels. Liquids may be aqueous or non-aqueous and solids may be soluble or insoluble in water. <u>Limitations</u> - Anti-fouling organic solvent-containing paints may be underpredicted (OECD, 2013d). - In the case of solid materials leading to a GHS No Cat. outcome, a second run of three eyes is recommended to confirm or discard the negative outcome.
Predictive capacity	When used for the identification of UN GHS No Cat. test chemicals, the ICE test method showed an overall accuracy of 82% (125/152), a false positive rate of 33% (26/79), and a false negative rate of 1% (1/73), when compared to *in vivo* rabbit eye test method (OECD TG 405) classified according to the UN GHS (OECD, 2013b). When anti-fouling organic solvent containing paints are excluded from the database, the accuracy of the ICE test method was found to be 83% (123/149), the false positive rate 33% (26/78), and the false negative rate of 0% (0/71) for the UN GHS classification system (OECD, 2013b).
Reliability	When distinguishing UN GHS No Cat. from chemicals classified for eye hazard (UN GHS Cat. 1 and 2), 75% (44/59) of the tested chemicals were found to have 100% agreement of classification between laboratories (ICCVAM, 2010b).

4.3.3. Short Time Exposure (STE) test method (OECD TG 491)

39. The OECD TG 491 on the STE test method was adopted in 2015 (OECD, 2015a). The STE test method underwent two prospective validation studies, one conducted by the Validation Committee of the Japanese Society for Alternative to Animal Experiments (JSAAE) (Sakaguchi *et al*., 2011) and another by JaCVAM (Kojima *et al*., 2013). A peer review was conducted by the US ICCVAM (ICCVAM, 2013), and from these evaluations it was concluded that the STE test method can correctly identify test chemicals (both substances and mixtures) inducing serious eye damage (UN GHS Cat. 1) as well as chemicals (excluding highly volatile substances and all solid chemicals other than surfactants) not requiring classification for eye hazard (UN GHS No Category).

Module 3 – *In vitro* data: Short Time Exposure (STE) test method (OECD TG 491)	
General description	
Regulatory use (UN GHS classification)	Identification of i) test chemicals inducing serious eye damage (UN GHS Cat. 1), and ii) test chemicals (excluding highly volatile substances and all solid chemicals other than surfactants) not requiring classification for eye hazard (UN GHS No Cat.).
Validation & regulatory acceptance status	Validated and adopted as OECD TG 491[11].
Potential role in the IATA	While the STE test method is not considered valid as a full-replacement for the *in vivo* rabbit eye test, it can be used to initiate either the Top-Down or the Bottom-Up approach for regulatory classification and labelling (Figure 1) to identify, without further testing: - test chemicals inducing serious eye damage (UN GHS Cat. 1); and - limited types of test chemicals (excluding highly volatile substances and solid substances and mixtures other than surfactants) that do not require classification for eye hazard (UN GHS No Cat.). A test chemical that is neither predicted as UN GHS Cat. 1 nor as UN GHS No Cat. with the STE test method would require additional WoE evaluation with other existing information and if still needed additional testing (*in vitro* and/or *in vivo*) as a last resort to establish a definitive classification (see Figure 1).
Description	The STE test method is a cytotoxicity-based *in vitro* assay that is performed on a confluent monolayer of Statens Seruminstitut Rabbit Cornea (SIRC) cells, cultured on a 96-well polycarbonate microplate. Each test chemical is tested at both 5% and 0.05% concentrations. After five-minute exposure to the test chemical, cell viability is assessed by the quantitative measurement, after extraction from cells, of the blue formazan salt produced by the living cells by enzymatic conversion of the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide), also known as Thiazolyl Blue Tetrazolium Bromide (Mosmann, 1983). The relative cell viability of the treated SIRC cells (compared to the solvent

[11] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 3 – *In vitro* data: Short Time Exposure (STE) test method (OECD TG 491)	
	control) is used to estimate the potential eye hazard of the test chemical. A test chemical is classified as UN GHS Cat. 1 when both the 5% and 0.05% concentrations result in a relative cell viability smaller than or equal to (≤) 70%. Conversely, a test chemical is predicted as UN GHS No Cat. when both 5% and 0.05% concentrations result in a relative cell viability higher than (>) 70%.
Scientific basis incl. MoA	It has been reported that 80% of a solution dropped into the eye of a rabbit is excreted through the conjunctival sac within three to four minutes, while greater than 80% of a solution dropped into the human eye is excreted within one to two minutes. The STE test method attempts to approximate these exposure times using the five-minute exposure to the test chemical. Decrease in cell viability is then used to predict potential adverse effects leading to ocular damage. The cytotoxic effects of test chemicals on corneal epithelial cells is an important mode of action leading to corneal epithelium damage and eye irritation.
Protocol available	The Test Guideline is based on a protocol developed by Kao Corporation (Takahashi *et al*., 2008).
General reproducibility	Evaluation of the STE reliability showed CV values for within-laboratory variability for test chemicals classified as UN GHS No Cat. spanning from 0.3% to 23.5% in four studies evaluated. Test chemicals classified *in vitro* tended to have greater CV values, as expected, because the cell viability for these test chemicals was often quite low. The mean viability for the positive control, 0.01% sodium lauryl sulfate, was 41.7% (N = 71) with a CV of 24.7%. A between-laboratory reproducibility of 83-100% was observed (ICCVAM, 2013).
Strengths and weaknesses	<u>Strengths</u> - Officially validated test method. - Quantitative measurements of cell viability. - Controlled exposure conditions, including test chemical concentration and exposure duration. <u>Weaknesses</u> - The STE test method is not recommended for the identification of test chemicals that should be classified as irritant to eyes (UN GHS Cat. 2 or Cat. 2A) or test chemicals that should be classified as mildly irritant to eyes (UN GHS Cat. 2B) due to the considerable number of UN GHS Cat. 1 chemicals underclassified as UN GHS Cat. 2, 2A or 2B and UN GHS No Cat. chemicals overclassified as UN GHS Cat. 2, 2A or 2B. For this purpose, further testing with another suitable method may be required. - The reversibility of tissue lesions cannot be evaluated *per se* in the STE test method. - Gases and aerosols have not been assessed yet in a validation study.
Identification of UN GHS Category 1	
Applicability domain and limitations	<u>Applicability</u> - Test chemicals (substances and mixtures) that are dissolved or uniformly suspended for at least 5 minutes in physiological saline, 5% dimethyl sulfoxide (DMSO) in saline, or mineral oil. <u>Limitations</u> - The high false negative rate observed (51%), is not critical in the present context, since all test chemicals that induce a cell viability of ≤ 70% at a 5%

Module 3 – *In vitro* data: Short Time Exposure (STE) test method (OECD TG 491)	
	concentration and > 70% at 0.05% concentration would be subsequently would be subsequently tested and evaluated following the sequential testing strategy as described in chapter 3 and Figure 1.
Predictive capacity	When used for identification of UN GHS Cat. 1 test chemicals, the STE test method showed an overall accuracy of 83% (104/125), a false positive rate of 1% (1/86), and a false negative rate of 51% (20/39) as compared to the *in vivo* rabbit eye test method (OECD TG 405) classified according to the UN GHS (OECD, 2015a).
Identification UN GHS No Category	
Applicability domain and limitations	Applicability - Test chemicals (substances and mixtures) that are dissolved or uniformly suspended for at least 5 minutes in physiological saline, 5% dimethyl sulfoxide (DMSO) in saline, or mineral oil. Limitations - Highly volatile substances with a vapour pressure over 6 kPa (at 25°C) are excluded from the applicability domain of the STE test method for the identification of UN GHS No Cat. due to the high false negative rate. Results obtained with mixtures containing substances with vapour pressure higher than 6kPa should be interpreted cautiously due to potential underprediction, and should be justified on a case-by-case basis. - Solid chemicals (substances and mixtures) other than surfactants and mixtures composed only of surfactants are also excluded from the applicability domain of the STE test method for the identification of UN GHS No Cat. due to high false negative rates observed.
Predictive capacity	When used for the identification of UN GHS No Cat. test chemicals, the STE test method showed an overall accuracy of 85% (110/130), a false negative rate of 12% (9/73), and a false positive rate of 19% (11/57) as compared to the *in vivo* rabbit eye test method (OECD TG 405) classified according to the UN GHS (OECD, 2015a). If highly volatile substances and solid chemicals (substances and mixtures) other than surfactants are excluded from the dataset, the overall accuracy improves to 90% (92/102), the false negative rate to 2% (1/54), and the false positive to 19% (9/48). In addition in-house data on 40 mixtures, showed an accuracy of 88% (35/40), a false positive rate of 50% (5/10), and a false negative rate of 0% (0/30) for predicting UN GHS No Cat. when compared to the *in vivo* rabbit eye test (Saito *et al*., 2015).

4.3.4. Reconstructed human Cornea-like Epithelium Eye Irritation Test (RhCE EIT) (OECD TG 492)

40. The OECD TG 492 on the RhCE Test Method was adopted in 2015 and revised in 2017 (OECD, 2017). The *in vitro* test methods currently covered by this Test Guideline are the EpiOcular™ Eye Irritation Test (EIT) which makes use of the commercially available EpiOcular™ OCL-200 RhCE tissue construct, and the SkinEthic™ Eye Irritation Test . The two EIT test methods underwent a prospective validation study conducted by EURL ECVAM and Cosmetics Europe (Barroso *et al*., 2014) and by industry (Alépée *et al*., 2016a, 2016b) respectively, and the outcome was peer-reviewed by the EURL ECVAM Scientific Advisory Committee (ESAC, 2014, 2016a). From

these evaluations it was concluded that the both EIT test methods can correctly identify test chemicals (substances and mixtures) not requiring classification for eye hazard (UN GHS No Category). In contrast to the *in vitro* methods described earlier (BCOP, ICE and STE), the RhCE EIT is not applicable for the identification of test chemicals inducing serious eye damage (UN GHS Cat. 1).

Module 3 – *In vitro* data: Reconstructed human Cornea-like Epithelium - Eye Irritation Test (OECD TG 492)	
General description	
Regulatory use (UN GHS classification)	Identification of test chemicals not requiring classification for eye hazard (UN GHS No Cat.).
Validation & regulatory acceptance status	Validated and adopted as OECD TG 492[12].
Potential role in the IATA	While the EpiOcular™ EIT is not considered valid as a full-replacement for the *in vivo* rabbit eye test, it can be used within the Top-Down and Bottom-Up approaches and in particular to initiate the Bottom-Up approach for regulatory classification and labelling (Figure 1) to identify, without further testing: - test chemicals that do not require classification for eye hazard (UN GHS No Cat). The EpiOcular™ EIT is not intended to differentiate between UN GHS Cat. 1 (serious eye damage) and UN GHS Cat. 2 (eye irritation). This differentiation will need to be addressed by another tier of a test strategy (Figure 1). A test chemical that is not predicted as not requiring classification for eye hazard (UN GHS No Cat.) with EpiOcular™ EIT will thus require additional *in vitro* testing and/or additional WoE evaluation with other existing information before progressing with further testing within the IATA in order to establish a definitive classification (see Figure 1).
Description	Three-dimensional RhCE tissues are reconstructed from primary human cells, which have been cultured for several days to form a stratified, highly differentiated squamous epithelium morphologically similar to that found in the human cornea. The EpiOcular™ RhCE tissue construct consists of at least 3 viable layers of cells and a non-keratinized surface, showing a cornea-like structure analogous to that found *in vivo*. The test chemical is applied topically to a minimum of two RhCE tissue constructs. Following the exposure and post-treatment incubation periods, tissue viability is assessed by the enzymatic conversion in viable cells of the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide) into a blue MTT formazan salt which is extracted from the tissues and quantitatively measured (Mosmann, 1983). Test chemicals not requiring classification and labelling according to UN GHS are identified as those having a tissue viability higher than (>) 60%.
Scientific basis incl. MoA	Reconstructed human cornea-like epithelium (RhCE) closely mimics the histological, morphological, biochemical and physiological properties of the human

[12] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 3 – *In vitro* data: Reconstructed human Cornea-like Epithelium - Eye Irritation Test (OECD TG 492)	
	corneal epithelium. The use of viability of the RhCE tissues after topical exposure to a test chemical to discriminate UN GHS No Cat. chemicals from those requiring classification and labelling (UN GHS Cat. 1 and 2) is based on the assumption that all chemicals inducing serious eye damage or eye irritation will induce cytotoxicity in the corneal epithelium and/or conjunctiva. Indeed, it has been shown that cytotoxicity plays an important mechanistic role in determining the overall serious eye damage and eye irritation response of a chemical regardless of the physicochemical processes underlying tissue damage (Jester *et al*., 1998; Maurer *et al*., 2002).
Protocol available	DB-ALM protocol n°. 164 on the EpiOcular™ EIT test method (DB-ALM, 2015).
Strengths and weaknesses	<u>Strengths</u> - Officially validated test method. - Human-based 3D tissue model. - Quantitative measurements of cell viability. - Controlled exposure conditions, including test chemical concentration and exposure duration. <u>Weaknesses</u> - The EIT RhCE test method does not allow discrimination between eye irritation/reversible effects on the eye (UN GHS Cat. 2) and serious eye damage/irreversible effects on the eye (UN GHS Cat. 1), nor between eye irritants (UN GHS Cat. 2A) and mild eye irritants (UN GHS Cat. 2B). For these purposes, further testing with other suitable test methods is required. - Gases and aerosols have not been assessed yet in a validation study. - The reversibility of tissue lesions cannot be evaluated *per se* in the EIT RhCE test method.
Identification UN GHS No Category	
Applicability domain and limitations	<u>Applicability</u> - The RhCE test method can be used for the testing of substances and mixtures (OECD, 2015b; Kolle *et al*., 2015, 2016). - It is applicable to solids, liquids, semi-solids and waxes. Liquids may be aqueous or non-aqueous and solids may be soluble or insoluble in water. <u>Limitations</u> - test chemicals presenting non-specific interactions with MTT (i.e., absorbing light in the same range as MTT formazan (naturally or after treatment) or able to directly reduce the vital dye MTT (to MTT formazan)) that are ≥ 60% should be taken with caution when OD is used to measure the extracted MTT formazan. However, use of HPLC/UPLC-spectrophotometry as an alternative procedure to measure MTT formazan allows circumventing this, and is especially useful for those test chemicals strongly absorbing in the same wavelength as MTT formazan which are not compatible with the standard optical density (OD) measurement (Alépée *et al*., 2015).
Predictive capacity	When used for the identification of UN GHS No Cat. test chemicals, the EpiOcular™ EIT test method showed an overall accuracy of 80% (based on 112 chemicals), a false negative rate of 4% (based on 57 chemicals), and a false positive

Module 3 – *In vitro* data: Reconstructed human Cornea-like Epithelium - Eye Irritation Test (OECD TG 492)	
	rate of 37% (based on 55 chemicals) when compared to *in vivo* rabbit eye test data (OECD TG 405) classified according to the UN GHS (OECD, 2015b). In addition a study on agrochemical formulations using the RhCE test method according to OECD TG 492, showed an overall accuracy of 82% (based on 97 formulations), a false negative rate of 9% (based on 54 formulations) and a false positive rate of 28% (based on 43 formulations) for predicting UN GHS No Cat. when compared to the *in vivo* rabbit eye test (Kolle *et al*., 2015).
Reliability	The concordance of predictions obtained with the EpiOcular™ EIT RhCE test method was found to be in the order of 95% within laboratories and 93% between laboratories.

4.3.5. Fluorescein Leakage (FL) test method (OECD TG 460)

41. The OECD TG 460 on the FL test method was adopted in 2012 (OECD, 2012b, 2012c). The FL test method has been evaluated in a retrospective validation study coordinated by EURL ECVAM in collaboration with US ICCVAM and JaCVAM (EURL ECVAM, 2008a, b), followed by peer review by the ESAC (ESAC, 2009b). From these evaluations it was concluded that the test method can correctly identify water-soluble test chemicals (both substances and mixtures) inducing serious eye damage (UN GHS Cat. 1). In contrast to the *in vitro* methods described earlier (BCOP, ICE, STE, RhCE), the FL assay is not applicable for the identification of test chemicals not requiring classification for eye hazard (UN GHS No Category).

Module 3 – *In vitro* data: Fluorescein Leakage (FL) test method (OECD TG 460)	
General description	
Regulatory use (UN GHS classification)	Identification of test chemicals inducing serious eye damage (UN GHS Cat. 1).
Validation & regulatory acceptance status	Validated and adopted as OECD TG 460[13].
Potential role in the IATA	While the FL test method is not considered valid as a full-replacement for the *in vivo* rabbit eye test, it can be used within the Top-Down and Bottom-Up approaches and in particular to initiate the Top-Down approach for regulatory classification and labelling (Figure 1) to identify, without further testing: - limited types of test chemicals (water soluble substances and mixtures), inducing serious eye damage (UN GHS Cat. 1). A test chemical that is not predicted to be UN GHS Cat. 1 with the FL test method will require additional *in vitro* testing and/or additional WoE evaluation with other existing information before progressing with further testing within the IATA in

[13] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 3 – *In vitro* data: Fluorescein Leakage (FL) test method (OECD TG 460)	
	order to establish a definitive classification (see Figure 1).
Description	The FL test method is a cell-function based *in vitro* assay that is performed on a confluent monolayer of Madin-Darby Canine Kidney (MDCK) CB997 tubular epithelial cells cultured on permeable inserts. The toxic effects of a test chemical are measured after a short exposure time (1 minute) by an increase in permeability of sodium fluorescein through the epithelial monolayer of MDCK cells. The amount of fluorescein leakage that occurs is proportional to the chemical-induced damage to the tight junctions, desmosomal junctions and cell membranes, and is used to estimate the ocular toxicity potential of a test chemical. The concentration of test chemical (mg/mL) causing 20% FL relative to the value recorded for the untreated confluent monolayer and inserts without cells (FL_{20}), is used to predict UN GHS Cat. 1 classification (i.e., $FL_{20} \leq 100$ mg/ml: UN GHS Cat. 1).
Scientific basis incl. MoA	The potential for a test chemical to induce serious eye damage is assessed by its ability to induce damage to an impermeable confluent epithelial monolayer. The MDCK cell line model the non-proliferating state of the *in vivo* corneal epithelium and forms tight junctions and desmosomal junctions similar to those found on the apical side of conjunctival and corneal epithelia. The short exposure period allows water-based substances and mixtures to be tested neat, if they can be easily removed after the exposure period, which allows more direct comparisons of the results with the chemical effects in humans. The integrity of trans-epithelial permeability is a major function of an epithelium such as that found in the conjunctiva and the cornea. Trans-epithelial permeability is controlled by various tight junctions. Tight and desmosomal junctions *in vivo* prevent solutes and foreign materials penetrating the corneal epithelium. Loss of trans-epithelial impermeability, due to damaged tight junctions and desmosomal junctions, is one of the early events in chemical-induced ocular irritation. Increasing permeability of the corneal epithelium *in vivo* has been shown to correlate with the level of inflammation and surface damage observed as eye irritation develops (OECD, 2012).
Protocol available	DB-ALM Protocol n° 71 on the FL test method (DB-ALM, 2013).
General reproducibility	Based on the data acquired in the validation study for 60 chemicals according to INVITTOX protocol 71, 43/60 materials (71.7%) had 100% agreement among all 4 participating laboratories. When concordance between 3 of the 4 laboratories was investigated, 59/60 materials (98.3%) had 100% agreement among 3 of the 4 laboratories. Moreover, data from INVITTOX protocol 120 were used as weight of evidence to further assess the Reproducibility of the FL test method. A good agreement of classification was obtained with 7/9 materials (77.8%) having 100% agreement among 3 laboratories, and 26/29 materials (89.7%) having 100% agreement among 2 laboratories (OECD, 2012c).
Strengths and weaknesses	<u>Strengths</u> - Officially validated test method. - Quantitative measurements. - Controlled exposure conditions, including test chemical concentration and exposure duration. - The FL test method may also assess recovery. Preliminary analyses indicated that recovery data (up to 72 h following exposure to the test chemical) could potentially increase the predictive capacity of the FL test method, although

Module 3 – *In vitro* data: Fluorescein Leakage (FL) test method (OECD TG 460)	
	further evaluation is needed and would benefit from additional data preferably acquired by further testing (OECD, 2012). Weaknesses - The FL test method is not recommended for the identification of test chemicals that should be classified as mild/moderate irritants (UN GHS Cat. 2 or UN GHS Cat. 2A and 2B), or of test chemicals which should not be classified for ocular irritation (UN GHS No Cat.), as demonstrated by the validation study (EURL ECVAM, 2008). - Gases and aerosols have not been assessed yet in a validation study.
Identification of UN GHS Category 1	
Applicability domain and limitations	Applicability - The test method is applicable to water soluble test chemicals (substances and mixtures) and/or where the toxic effect is not affected by dilution. Limitations - Strong acids and bases, cell fixatives and highly volatile test chemicals are excluded from the applicability domain as these chemicals have mechanisms that are not measured by the FL test method, e.g. extensive coagulation, saponification or specific reactive chemistries. - Coloured and viscous test chemicals are difficult to remove from the monolayer following the short exposure period but predictivity of the test method could be improved if a higher number of washing steps was used. - The final concentration to cells of solid test chemicals suspended in liquid that have the propensity to precipitate can be difficult to determine.
Predictive capacity	When used for identification of UN GHS Cat. 1 test chemicals, data obtained with the FL test method showed an accuracy of 77% (117/151), a false positive rate of 7% (7/103) and a false negative rate of 56% (27/48) when compared to *in vivo* rabbit eye test method (OECD TG 405) classified according to the UN GHS classification system (OECD, 2012b).

4.4. Module 4: Other existing animal data from non-OECD adopted test methods on serious eye damage and eye irritation

42. Existing data from modified OECD TG 405 or *in vivo* animal test methods adopted by specific countries and/or regulatory authorities similar but not fully compliant with OECD TG 405, shall be considered. Although not fully following the recommendations from the OECD TG 405, existing data obtained from these *in vivo* animal studies may be useful in giving indication on the potential eye hazard effects of a test chemical. Examples of such *in vivo* animal test methods include the original Draize test method (Draize *et al.*, 1944), the US FHSA method 16CFR 1500.42 (US CPSC 2015b) and eventual modifications to TG 405. An evaluation shall be made on the degree of similarity and differences of these test methods as compared to the OECD TG 405, and the results used in WoE assessment in Module 9 to support classification and labelling decisions.

43. In addition to the above test methods, another non-OECD adopted *in vivo* animal test is the Low Volume Eye Test (LVET) which involves the application of $1/10^{th}$ of the amount applied in OECD TG 405 (e.g., 10 µL instead of 100 µL for liquids) directly onto the cornea (instead of into the conjunctival sac) and uses the same scale and the data interpretation as those used in OECD TG 405. Such amount is based on anatomical and physiological considerations indicating that the tear volume

in both rabbit and human eyes is approximately the same (~7-8 µL), and that after blinking, the volume capacity in the human eye is ~10 µL (A.I.S.E. 2006). Furthermore, the use of direct cornea exposure mimics human exposure scenarios that can be reasonably expected from e.g. accidental ocular exposure to household detergents and cleaning products. Indeed, the LVET has been mainly used for detergent and cleaning products (Freeberg *et al*., 1984; Freeberg *et al*., 1986a,b; Cormier *et al*., 1995; Roggeband *et al*., 2000). It was found to still overpredict the effects in man, but to a lesser extent as compared to the classical *in vivo* rabbit eye test described in OECD TG 405 (Freeberg *et al*., 1984, 1986b; Roggeband *et al*., 2000). Following a retrospective validation study and independent peer review, the LVET was not recommended for prospective use, i.e. to generate new data (ESAC, 2009a; ICCVAM, 2010a). Furthermore, although the LVET was considered to have a tendency to classify in lower hazard categories when compared to OECD TG 405 (ICCVAM, 2010a), it was acknowledged that retrospective LVET data may still be useful on a case-by-case basis (e.g. in a WoE approach) to identify potential ocular irritants for the limited use domain of detergent and cleaning products and their main ingredients (i.e., surfactants used in these products) (ESAC, 2009a; ICCVAM, 2010a).

4.5. Module 5: Other data from non-OECD adopted alternative test methods on serious eye damage and eye irritation

44. In addition to the OECD adopted *in vitro* test methods, a number of promising alternative test methods and complementary endpoints exist that may provide with complementary and/or useful information for predicting eye hazard effects. These encompass: (i) histopathology as an additional *in vitro* endpoint recommended by the OECD GD 160 (2017); (ii) test methods that underwent validation studies according to e.g. the OECD GD 34 (2005); and (iii) promising optimized alternative methods for predicting e.g., irreversible effects and UN GHS Cat. 2 classification. Table 3 provides with an overview of these test methods including a description of the endpoints assessed, their proposed application and their validation and regulatory status. Furthermore other non-OECD adopted alternative methods on serious eye damage and eye irritation may also include test methods derived or adapted from OECD adopted *in vitro* test methods that make use of e.g., i) the same endpoint but measured with a different technology, ii) a new endpoint (in)directly related to the endpoint addressed in the OECD adopted test method(s), and iii) an adapted methodology(s) using the adopted model.

4.5.1. OECD Guidance Document 160 on the use of histopathology as an additional endpoint

45. In 2011, the OECD adopted the Guidance Document n. 160 which provides standard procedures for the collection, fixing and processing of tissues for histological evaluation as an additional endpoint to the BCOP and ICE test methods for eye hazard testing (OECD, 2017). The Guidance Document suggests that histopathological evaluation may be useful for (i) assessing histological damage of chemical classes or formulations that are not well characterized in these test methods; (ii) assisting with determination of a mode of action where it cannot be easily predicted; (iii) assisting with determination of the likelihood of delayed effects; (iv) evaluation of the depth of injury, which has been proposed as a measure of reversibility or irreversibility of ocular effects (Maurer *et al*., 2002); (v) further characterization of the severity or scope of the damage as needed (Harbell *et al*., 2006; Maurer *et al*., 2002); and (vi) assisting with discrimination of cases where the response falls along the borderline between two categories based on the standard test method decision criteria. GD 160 mainly addresses the use of histopathology as an additional endpoint to the BCOP and ICE (TG 437 and TG 438) based on the experiences gained so far with these test methods, however, it is conceivable that such endpoint may also be applicable to other tissue models such as the more recently adopted RhCE (TG 492) and the non-OECD adopted IRE test method (see chapter 4.5.2).

Table 3: Overview of non-OECD adopted test methods useful in supporting eye hazard identification. Note that this list is likely to be non-exhaustive. Furthermore it is recommended to check latest status of those methods under discussions at the OECD level[14].

Test method	Endpoint(s) assessed	Proposed application by test developer	Validation & regulatory status	Comments
OECD Guidance Document 160 on the use of histopathology as an additional endpoint to the BCOP and ICE test methods				
Histopathology as an additional endpoint	- Assisting in determining mode of action, likelihood of delayed effects, depth of injury, and borderline effects in standard ICE and BCOP - Further characterization of chemical classes / formulations not well characterized in BCOP and ICE - Further characterization of the severity or damage - May be useful to other tissue-based methods such as RhCE and IRE	ICE: Identification of **UN GHS Cat. 1** *detergents and cleaning products*.	OECD GD 160	-
RhCE *in vitro* test methods that underwent validation				
LabCyte CORNEA-MODEL24 EIT	RhCE me-too assay falling within TG 492	Identification of **UN GHS No Cat.** test chemicals	Validated based on performance standards & under peer review	Under discussions at the OECD level
Vitrigel EIT method	Barrier function of human corneal epithelium cells cultured in a collagen vitrigel membrane chamber	Identification of **UN GHS No Cat.** test chemicals	Underwent validation & under peer-review	Under discussions at the OECD level

Test method	Endpoint(s) assessed	Proposed application by test developer	Validation & regulatory status	Comments
EpiOcular time-to-toxicity (ET_{50}) assay	Time of exposure to reduce tissue viability of 50% as compared to the control tissues, using a reconstructed human corneal-like epithelial model	Moderate to mild irritants	Underwent validation & peer review for specific applicability domain Accepted by certain regulatory authorities (US EPA, 2015)	Accepted for testing antimicrobial and cleaning products, when used in combination with BCOP and Cytosensor Microphysiometer
Macromolecular *in vitro* test methods that underwent validation				
Ocular Irritection®	Denaturation of a macromolecular matrix composed of proteins, glycoproteins, lipids and low molecular weight components	Identification of: **UN GHS Cat. 1** *test chemicals falling within the applicability domain of the test method* **UN GHS No Cat.** *test chemicals falling within the applicability domain of the test method, excluding test chemicals having the functional groups acrylate, carboxamide and cycloalkene*	Underwent validation & peer review Accepted by certain regulatory authorities for the identification of serious eye damage (ECHA, 2015)	Under discussions at the OECD level
Cell-based *in vitro* test methods that underwent validation				
Cytosensor Microphysiometer	Metabolic rate of L929 fibroblasts	Identification of: **UN GHS Cat. 1** *water-soluble test chemicals* **UN GHS No Cat.** *water-soluble surfactants and surfactant-containing test chemicals*	Considered scientific valid following peer review Accepted by certain regulatory authorities (ECHA 2015; USA EPA, 2015)	The original version of the apparatus is no longer commercially available at the time of redaction of this GD. A new OECD TG will be considered when new apparatuses showing similar performances as the original apparatus are available

Test method	Endpoint(s) assessed	Proposed application by test developer	Validation & regulatory status	Comments
Neutral Red Release (NRR)	Cytotoxicity, measured as release of neutral red dye in monolayer fibroblast cell cultures	Identification of **UN GHS No Cat.** *water soluble test chemicals* Proof-of-concept study with a modified prediction model showed the NRR to be useful also for identification of **UN GHS Cat. 1** agrochemical formulations (Settivari *et al*., 2016)	Underwent validation & peer review	Further work was recommended before a statement on the scientific validity of the NRR could be made
Red Blood Cell (RBC) test	Haemolysis of red blood cells, oxyhaemoglobin denaturation (from e.g. calf blood from slaughterhouse, human blood, rabbit blood)	Identification of serious eye damage and no need for classification	Underwent validation & peer review	Further work was recommended before a statement on the scientific validity of the RBC could be made
Organotypic *in vitro* test methods that underwent validation				
Isolated Rabbit Eye (IRE)	Corneal effects of enucleated rabbit eyes (obtained e.g. from the food chain of from euthanized laboratory rabbits used for other purposes than ocular procedures, providing that no abnormalities are detected in the eyes prior to use)	Identification of **UN GHS Cat. 1** test chemicals	Underwent validation & peer review Accepted by certain regulatory authorities for the identification of serious eye damage (ECHA, 2015)	Further work was recommended before a statement on the scientific validity of the IRE could be made

Test method	Endpoint(s) assessed	Proposed application by test developer	Validation & regulatory status	Comments
Hen's Egg Test on the Chorio-Allantoic Membrane (HET-CAM)	Haemorrhage, lysis and coagulation of blood vessels of the chorioallantoic membrane (CAM) of fertilized chicken eggs	Identification of: **UN GHS Cat. 1** test chemicals (based on coagulation) **UN GHS No Cat.** test chemicals (based on coagulation, haemorrhage and lysis)	Underwent validation & peer review Accepted by certain regulatory authorities for the identification of serious eye damage (ECHA, 2015)	Further work was recommended before a statement on the scientific validity of the HET-CAM can be made Depending upon the regulatory context, this assay may be considered an animal test
Chorio-Allantoic Membrane Vascular Assay (CAMVA)	Haemorrhage, hyperaemia and constriction of blood vessels of the chorioallantoic membrane (CAM) of fertilized chicken eggs	Moderate to mild irritants	Underwent validation	Further work required to evaluate the scientific validity of the test method
Other promising assays				
Porcine Ocular Cornea Opacity/Reversibility Assay (PorCORA)	Reversibility of cornea injury in air-interface *ex vivo* porcine corneas cultured for 21 days	Proposed for identification of Serious Eye Damage based on persistence/reversibility of effects as well as severity of effects and a better discrimination between Cat. 1 and Cat. 2	Optimised	-
***Ex Vivo* Eye Irritation Test (EVEIT)**	Reversibility of epithelial and stromal damage of isolated corneas from rabbit eyes (obtained from the food chain) cultured at air-liquid interface for 72 hours	Proposed for the assessment of both severity and persistence/reversibility of ocular lesions, allowing to discriminate between all UN GHS categories (No Cat., Cat. 2A/B and Cat. 1)"	Optimised	-

Test method	Endpoint(s) assessed	Proposed application by test developer	Validation & regulatory status	Comments
3D hemi-cornea	Cytotoxicity and/or depth of injury of a multilayered human-based epithelium and stroma with embedded keratocytes cultured in a collagenous matrix	Proposed for identification of all UN GHS categories (Cat. 1, Cat. 2A/2B and No Cat.)	Optimised	-
Slug mucosal irritation (SMI) assay	Mucus produced from the mucosal surface of slugs	Proposed for identification all UN GHS categories (No Cat., Cat. 2A/B and Cat. 1)	Optimised	Depending upon the regulatory context, this assay may be considered an animal test

46. In particular, the use of histopathology as an additional endpoint to the ICE test method was found to decrease the rate of false negatives (as observed with the ICE test method when used as a stand-alone) for the identification of UN GHS Cat. 1 for the limited applicability domain of non-extreme pH detergent and cleaning products (Cazelle *et al*., 2014, 2015). Interestingly, these mixtures were mostly classified *in vivo* UN GHS Cat. 1 due to persistence of effects, i.e. mild ocular effects that persisted over the 21 day observation period in the tested rabbits. The authors developed a decision criteria for identification of UN GHS Cat. 1 based on semi-quantitative histopathological observations (Prinsen *et al*., 2011) in which epithelial vacuolation (mid and lower layers) and epithelial erosion (at least moderate level) were found to be the most typical histopathological effects induced by UN GHS Cat. 1 non-extreme pH formulations that were classified *in vivo* mainly due to persistence of effects (Cazelle *et al*., 2014). Use of such criteria for non-extreme pH detergent and cleaning formulations that were identified by the standard ICE test method as 'no prediction can be made', allowed to decrease the rate of Cat. 1 false negatives observed with the ICE test method alone whilst maintaining a good accuracy and an acceptable specificity (Cazelle *et al*., 2014).

47. Furthermore, recent studies on the BCOP test method suggest that histopathology might be useful in predicting *in vivo* ocular irritation, particularly for test chemicals with $3 < \text{IVIS} \leq 25$ that would be classified as mild irritants (Cat. 2B) according to the UN GHS (Furukawa *et al*., 2015). The authors showed that corneal epithelial lesions caused by Cat. 2B test chemicals were localized on the border between the corneal epithelium and stroma.

4.5.2. In vitro test methods that underwent validation studies

48. Methods that underwent validation studies according to e.g. the OECD GD 34 (2005), encompass reconstructed human tissue models, organotypic test methods, cell based assays and a macromolecular test method. These test methods as well as additional test methods may become available for addressing eye hazards, therefore it is advised to always check the latest status of these test methods on the OECD website[15].

4.5.2.1. Reconstructed human tissue models

49. The ***LabCyte CORNEA-MODEL*** is a RhCE model that underwent a performance-based validation study according to the OECD GD 216 (2015c). Furthermore, results obtained on 61 test chemicals showed good predictive capacity of the test method (Katoh *et al*., 2013). It has been proposed as a me-too assay to the RhCE test method falling within the OECD TG 492, and at the time of the redaction of this document, it is currently under peer-review.

50. The ***Vitrigel-eye irritancy test method*** is a RhCE based assay which assesses the effects of test chemical on the barrier function of human corneal epithelium cells cultured in a collagen vitrigel membrane. Prediction of UN GHS No Cat. is based on a time-dependent profile of transepithelial electrical resistance assessed for 3 min after exposure to the test chemicals. A total of 118 chemicals have been tested, and when test chemicals having a $pH \leq 5$ are removed from the applicability domain, the assay showed performances in line with the adopted test methods for the prediction of UN GHS No Cat. test chemicals (Yamaguchi *et al*., 2016). The assay underwent a formal validation study and at the time of the redaction of this document, is currently under peer-review.

51. Finally, the ***EpiOcular time-to-toxicity (ET_{50}) assay*** is a RhCE assay in which the eye hazard effects are evaluated by the time necessary to reduce tissue viability to e.g. 50% (in contrast to the decrease in cell viability with a fixed exposure time recommended in OECD TG 492). The assay underwent validation studies focusing on surfactant ingredients and a limited number of formulations (Blazka *et al*., 2000, 2003). It further underwent a peer-review when used as a part of a testing strategy together with the BCOP and Cytosensor Microphysiometer test methods, to evaluate anti-

[15] http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788

microbial cleaning products (ICCVAM, 2010c). Such test strategy was accepted by the US EPA for determining the appropriate eye hazard classification for antimicrobial cleaning products (US EPA, 2015). Further work evaluated the usefulness of this assay, when combined with the NRR to evaluate the eye hazard potential of agrochemical formulations (Settivari *et al*., 2016)

4.5.2.2. Macromolecular assays

52. The ***Ocular Irritection® (OI) assay*** is based on a macromolecular reagent produced from a biological extract that is composed of proteins, glycoproteins, lipids and low molecular weight components that self-associate to form a complex matrix. Eye hazard is assessed based on the premise that irritant test chemicals will lead to protein denaturation and disaggregation of the macromolecular matrix. The changes in protein structure result in changes in turbidity which are measured at an OD of 405 nm. The assay underwent a prospective and a retrospective validation study (Eskes *et al*., 2014), in which the test method is proposed to identify test chemicals falling within its applicability domain (both substances and mixtures) inducing serious eye damage (UN GHS Cat. 1) as well as those not requiring classification for eye hazard (UN GHS No Cat.). The test method showed good within-laboratory variability including transferability to a naïve laboratory, and between-laboratory based on concordance of classifications. When used for the identification of UN GHS Cat. 1 versus other categories, and for the identification UN GHS No Cat. versus classified materials, excluding the functional groups *acrylate*, *carboxamide* and *cycloalkene*, the test method showed accuracy, false negative and false positive rates which were in line with currently adopted test methods for that purpose (Eskes *et al*., 2014). The outcome of the validation study was subsequently evaluated by EURL ECVAM and peer reviewed by the ESAC (ESAC, 2016b), in which a few technical issues were identified, which are currently under discussion at the OECD level. Furthermore, the OI assay is accepted by certain countries for the prediction of serious eye damage (UN GHS Cat. 1) (ECHA, 2015).

4.5.2.3. Cell-based assays

53. A draft OECD Test Guideline has been proposed on the ***Cytosensor Microphysiometer (CM)*** test method (OECD, 2012d). The CM has been evaluated in a retrospective validation study coordinated by EURL ECVAM in collaboration with US ICCVAM and JaCVAM (EURL ECVAM, 2008b), followed by peer review by the ESAC (ESAC, 2009b). From these evaluations it was concluded that the test method can correctly identify water-soluble test chemicals (both substances and mixtures) inducing serious eye damage (UN GHS Cat. 1) as well as water-soluble surfactants and surfactant-containing test chemicals not requiring classification for eye hazard (UN GHS No Cat.). The assay is performed on a sub-confluent monolayer of adherent mouse L929 fibroblasts cultured in a sensor chamber using a pH-meter to detect changes in acidity (Harbell *et al*., 1997). The rate of change in acidity (per unit time) measured during the assay serves as a read-out to determine the metabolic rate of the population of cells. If a test chemical causes cytotoxicity to this population of cells, it is assumed that the metabolic rate will fall. The concentration of a test chemical that leads to a 50 % decline (MRD_{50}) in the basal metabolic rate of the population is the parameter used to indicate cytotoxic effects. Identification of water-soluble test chemicals inducing serious eye damage (UN GHS Cat. 1) is triggered by an $MRD_{50} \leq 2$ mg/ml whereas UN GHS No Cat. water-soluble surfactants and surfactant containing mixtures are identified by an $MRD_{50} \geq 10$ mg/ml (OECD, 2012d). The CM may also address questions of cell metabolism and recovery. However, the assay requires the use of a Cytosensor Microphysiometer instrument, and at the time of redaction of this GD the original version of this apparatus is no longer commercially available so that the implementation of the assay with newly acquired original apparatus is not possible. Nevertheless, similar me-too apparatus are being commercialised but these have not been validated yet. Adoption of an OECD TG on the CM will be considered when new apparatuses are available that show similar performances to the original version. However, the Cytosensor Microphysiometer is accepted by certain regulatory authorities (ECHA, 2015; US EPA, 2015).

54. The ***Neutral Red Release (NRR)*** is based on near-confluent monolayer cell cultures, and assesses the eye hazard effects of test chemicals by exposure to serial dilutions of test chemicals for 1 to 5 minutes. The concentration of test chemical producing a 50% release of pre-loaded neutral red dye is obtained by extrapolation from the dose–response curve and used to predict eye hazard. The NRR test method has been evaluated in a retrospective validation study coordinated by EURL ECVAM in collaboration with US ICCVAM and JaCVAM (EURL ECVAM, 2008b), followed by peer review by the ESAC (ESAC, 2009b). However, further work was recommended before a statement on the scientific validity of the NRR could be made including to test additional number and classes of chemicals, and to obtain more data on between-laboratory variability (ESAC, 2009b). Although not formally endorsed as scientifically valid, the NRR test method was considered promising by the Validation Management Group for the identification of UN GHS No Cat., water-soluble test chemicals (EURL ECVAM, 2008b; ESAC 2009c). Furthermore, a recent proof-of-concept study making use of a modified prediction model, suggested the test method to be useful also for the identification of UN GHS Cat. 1 agrochemical formulations (Settivari *et al*., 2016).

55. The ***Red Blood Cell (RBC)*** haemolysis test is based on the potential of a test chemical to disrupt cell membranes as assessed by measuring photometrically the leakage of haemoglobin from freshly-isolated red blood cells incubated with the test chemical under standard conditions (Muir *et al*., 1983; Pape *et al*., 1987, Pape & Hope 1990; Pape *et al*., 1999; Lewis *et al*., 1993). The denaturation (i.e. change in protein configuration) of oxyhaemoglobin is used as second toxicological endpoint. Mammalian erythrocytes might be obtained through e.g. slaughterhouse material. The RBC test method has been evaluated in a retrospective validation study coordinated by EURL ECVAM in collaboration with US ICCVAM and JaCVAM (EURL ECVAM, 2008b), followed by peer review by the ESAC (ESAC, 2009b). The evidence then available was considered insufficient to support a recommendation on the RBC's scientific validity for regulatory use. In particular a more consistent dataset was deemed necessary to improve confidence on the RBC's applicability domain (ESAC, 2009b, 2009c).

4.5.2.4. Organotypic assays

56. The ***Isolated Rabbit Eye (IRE)*** is based on the same principles as the ICE test method, but instead of chicken eyes it uses enucleated rabbit eyes (obtained from the food chain or from euthanized laboratory rabbits, providing that the animals have not previously been used for ocular procedures, and that no abnormalities are detected in the eyes prior to use in the IRE by e.g. slit-lamp examination), instead of chicken eyes. The effects of test chemicals are assessed by evaluating the corneal thickness (swelling), corneal opacity, area of corneal involvement, fluorescein penetration and morphological changes to the corneal epithelium. Similar to BCOP and ICE, histopathology may be used as an additional endpoint. Furthermore, confocal microscopy may be used to determine the extent and depth of ocular injury. The IRE test method underwent retrospective validation by ICCVAM (ICCVAM 2006, 2010b) in which further work was recommended before a statement on its scientific validity could be made. The main reason was the fact that several endpoints and protocols for the IRE were applied and evaluated, each with insufficient data provided to make a sound conclusion (ESAC, 2007). Despite this, the IRE continues to be used (Guo *et al*., 2010) and is accepted by certain countries for the prediction of serious eye damage (UN GHS Cat. 1) (ECHA, 2015). However, depending upon the regulatory context, this assay may or may not be considered as an animal test.

57. The ***Hen's Egg Test on the Chorio-Allantoic Membrane (HET-CAM)*** is an assay that allows evaluating vascular effects. It makes use of the chorioallantoic membrane (CAM) of fertilized chicken eggs, a vascular foetal membrane composed of the fused chorion and allantois. The acute effects induced by a test chemical on the small blood vessels and proteins of this soft tissue membrane can be used as indicator of ocular effects induced by the test chemical (ICCVAM, 2010b). This characteristic makes the HET-CAM particularly suited to predict conjunctival injury and effects to the vascular system. The test chemical is applied directly to the CAM of fertilized hen eggs, and acute effects such as haemorrhage, lysis of blood vessels and coagulation are assessed. The test method is

accepted by certain countries for the identification of serious eye damage (UN GHS Cat. 1) (ECHA, 2015) although further work was recommended before a statement on its scientific validity could be made (ICCVAM 2006, 2010b). One potential reason for such outcome is the existence of a variety of protocols and prediction models used for the same test method. A workshop organized in 2012 by the German Federal Institute for Risk Assessment (BfR), the European Partnership for Alternative Approaches to Animal Testing (EPAA) and Services and Consultation on Alternative Methods (SeCAM) have made recommendations on the most suitable endpoints and protocols to be used either for the identification of UN GHS Cat. 1 or for the identification on UN GHS No Cat. Briefly, for the identification of serious eye damage (UN GHS Cat. 1), coagulation was the recommended endpoint based either on the mean time to develop coagulation or on the severity of coagulation observed at a single time after exposure (Spielmann *et al*., 1991; Steiling *et al*., 1999). For the identification of test chemicals not requiring classification (UN GHS No Cat.), the evaluation of coagulation, haemorrhage and lysis at different fixed time points (0.5, 2 and 5 min) was recommended (Luepke, 1985), based on the IS(a) prediction model (ICCVAM, 2010b). The necessity of re-considering the validation status of the method was also raised during this workshop in order to re-analyze the HET-CAM predictive capacity (for the identification of both UN GHS Cat. 1 and UN GHS No Cat.) taking into account the new data generated since 2009. In particular a new validation study has been initiated in 2015 by the Brazilian Centre for the Validation of Alternative Methods (BraCVAM) to complete such dataset. Due to the fact that the HET-CAM method uses live chick embryos, depending upon the regulatory context, this assay may or may not be considered as an animal test. The HET-CAM is accepted by certain countries for the prediction of serious eye damage (UN GHS Cat. 1) (ECHA, 2015).

58. **The Chorioallantoic Membrane Vascular Assay (CAMVA)** is another assay that also assesses the potential hazard effects of test chemicals to the blood vessels of the CAM. In preparing for the test, a small opening is cut into the shell of the egg four days after fertilisation and a small amount of albumen is removed, to permit optimal growth of the CAM. On day 10, the test substance is applied directly onto a small area of the CAM, and after exposure for 30 minutes, the eggs are examined for any vascular change to the CAM, such as haemorrhaging or hyperaemia (capillary injection) or the occurrence of vessels devoid of blood flow (ghost vessels). The concentration of a test material eliciting such damaging effects in 50 % of the treated eggs is used to predict eye hazard. The CAMVA has been included in a number of validation studies (for review see Eskes *et al*., 1995), and has mostly been applied to the assessment of materials in the mild-to-moderate irritation range (Cerven and Moreno, 1998). However, the CAMVA has not been assessed in parallel by more than two or three laboratories; thus, larger-scale validation or a retrospective validation based on the existing data is required, in order to further evaluate the scientific validity of the test method (Brantom *et al*., 1997; Bagley *et al*., 1992, 1999).

4.5.3. Promising optimized in vitro test methods

59. A number of assays have been developed to address mechanisms of action not covered by the currently accepted test method. Perhaps the most important mechanism is the discrimination of reversible vs. irreversible effects. As described in chapter 4.2.2, persistence of effects appears as a major driver for UN GHS Cat. 1 classification that may not be directly predicted by the currently adopted *ex vivo* and *in vitro* test methods. Two test methods have been developed specifically to address this mechanism of action, the Porcine Cornea Opacity/Reversibility Assay and the *Ex Vivo* Eye Irritation Test. The ***Porcine Cornea Opacity/Reversibility Assay (PorCORA)*** assay makes use of an air-interface culture system to sustain *ex vivo* porcine corneas in culture for 21 days (similar to the *in vivo* observation period described in TG 405), and determines whether cornea injury once inflicted will reverse (Piehl *et al*., 2010). Corneal injury reversibility is measured using Sodium Fluorescein stain to detect compromised epithelial barrier function. The test method was shown to identify test chemicals causing both reversible and irreversible serious eye damage in the *in vivo* rabbit eye test based on 32 tested UN GHS Cat. 1 test chemicals(Piehl *et al*., 2011). The ***Ex Vivo Eye Irritation Test (EVEIT)*** in contrast is based on isolated corneas from rabbit eyes (slaughtered for food process), cultured in an air-liquid interface in conditions that allow maintenance of the normal physiological and biochemical functions of the entire rabbit cornea *in vitro* for 72 hours after sample application.

Effects on cornea and reversibility of epithelial and stromal damage are assessed at 24h, 48h and 72h after test chemical application, by measuring corneal opacity (by macroscopic imaging in combination with fluorescein staining) as well as depth of damage, corneal thickness and structural changes assessed by the non-invasive Optical Coherence Tomography technique. A prediction model has been developed to identify the three UN GHS Categories of eye hazard (Spöler *et al*., 2015).

60. In addition, two test methods have been suggested for the discrimination of the entire range of ocular hazards, including the UN GHS Cat. 2 classification i.e., the 3D hemi-cornea model, and the slug mucosal irritation (SMI) assay. The ***3D hemi-cornea model*** is a new model comprised of a multilayered RhCE and a stroma with embedded human corneal keratocytes in a collagenous matrix for which two different test approaches are proposed. The first approach quantifies the cytotoxicity within the epithelium and the stroma separately and uses both values obtained, based on pre-defined thresholds for each compartment, to predict the potential eye hazard (Bartok *et al*., 2015). The second approach quantifies the cytotoxicity by microscopically assessing the depth of injury within the hemi-corneal tissue (Zorn-Kruppa *et al*., 2014). Preliminary results showed the capacity of the two approaches to differentiate UN GHS Cat. 1 from UN GHS Cat. 2 test chemicals based on 30 chemicals tested with the first approach, 25 chemicals tested with the second approach, and 14 chemicals tested in both approaches covering the entire range of eye hazards (Bartok *et al*., 2015; Zorn-Kruppa *et al*., 2014; Tandon *et al*., 2015). The same studies support the initial approach to differentiate between all 3 GHS categories, although the selectivity of both methods still must be improved before they can be used as stand-alone methods. The successful method transfer has been demonstrated in a ring trial with both approaches (Mewes *et al*., 2017).The ***slug mucosal irritation assay*** in contrast predicts the eye hazard of test chemicals based on the protein release from the mucosal surface of slugs (*Arion lusitanicus*). A pre-validation study was conducted with four participating laboratories and the testing of 20 chemicals covering the entire range of eye hazards. All UN GHS No Cat. were predicted correctly by the four laboratories. Furthermore, identification of both UN GHS Cat. 2 and UN GHS Cat. 1 showed good predictivity (Adriaens *et al*., 2005, 2008). The assay was also shown to be promising to predict ocular discomfort caused by shampoos (Lenoir *et al*., 2011). However, depending upon the regulatory context, this assay may or may not be considered as an animal test.

4.6. Module 6: Existing human, *in vivo* and *in vitro* data on skin corrosion

61. Existing human, *in vivo* and *in vitro* data generated on skin corrosion should be taken into account, such as those derived from an Integrated Approach on Testing and Assessment for Skin Corrosion and Irritation (OECD, 2014a). If sufficient and adequate quality data exists to assign Skin Corrosive Cat. 1, 1A, 1B or 1C, the risk of serious damage to eyes is considered implicit (UN GHS Cat. 1).

4.7. Module 7: Physico-chemical properties (existing, measured or estimated)

62. Test chemicals having pH $\leq$ 2.0 or pH $\geq$ 11.5 are predicted to be corrosive to skin or cause serious eye damage (UN GHS Cat. 1). However, where extreme pH is the only basis for classification as serious eye damage, it may also be important to take into consideration the acid/alkaline reserve i.e., a measure of the buffering capacity of a test chemical, especially for classification of mixtures containing acidic or alkaline substances (Young *et al*., 1988).

63. The determination of pH should be performed following OECD TG 122 (2013e). This Test Guideline also describes procedures to determine acid reserve or alkali reserve for test chemicals that are acidic (pH < 4) or alkaline (pH > 10) by titration with standard sodium hydroxide or sulphuric acid solution using electrometric endpoint detection.

64. However, the pH or pH in combination with buffering capacity should not be used alone to exonerate from serious eye damage classification. Indeed, when the pH or pH in combination with

acid/alkaline reserve suggest that the test chemical might not induce serious eye damage, further *in vitro* testing should be considered.

65. Other physico-chemical properties such as melting point, molecular weight, octanol-water partition coefficient, surface tension, vapour pressure, aqueous solubility and lipid solubility, may also be used to identify chemicals not likely to cause such adverse health effects (Gerner *et al*., 2005; Tsakovska *et al*., 2005). Such physico-chemical parameters may be measured or estimated using non-testing methods (see module 7), e.g., (Q)SARs, and may be used to help orient chemicals to a Top-Down or Bottom-Up approach in Part 3 of the IATA (Figure 1).

Module 7 – Data on physico-chemical properties: Extreme pH	
General description	
Regulatory use (UN GHS classification)	Prediction of serious eye damage (UN GHS Cat. 1)
Validation & regulatory status	Not formally validated but accepted as part of IATA.
Potential role in the IATA	Useful to identify test chemicals with potential to induce serious eye damage. However, the pH or pH in combination with buffering capacity should not be used alone to exonerate from serious eye damage classification. Indeed, when the pH or pH in combination with acid/alkaline reserve suggest that the test chemical might not induce serious eye damage, further *in vitro* testing should be considered.
Description	pH measurement (considering buffering capacity, if relevant).
Scientific basis incl. MoA	Test chemicals exhibiting extreme pH (either pH ≤ 2.0 or pH ≥ 11.5), with high buffering capacity when relevant, are likely to produce necrosis to the eyes.
Protocol available	OECD TG 122 (2013e)[16] describes the procedure to determine pH, acidity and alkalinity of aqueous solutions or aqueous dispersions having a pH ≤ 14.
Strengths and weaknesses	<u>Strengths</u> - Simplicity. - Low cost. <u>Weaknesses</u> - No information available on the test method reliability (reproducibility). - Predicts serious eye damage induced by pH effects but not by other mechanisms. - There are known cases of test chemicals with extreme pH that do not induce serious eye damage and therefore, use of pH information alone for deciding on Cat. 1 classification may lead to overclassification.
Identification of UN GHS Category 1	

[16] Note that the latest version of an adopted test guideline should always be used when generating new data. For more information on the latest version available, please consult: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Module 7 – Data on physico-chemical properties: Extreme pH	
Applicability domain and limitations	Applicability - Although OECD TG 122 allows pre-treatment with acetone to avoid plugging of the electrodes, some test chemical properties, such as low water solubility or rapid hydrolysis, might impair pH measurements. Limitations - For extreme pH mixtures having low or no buffering capacity suggesting the mixture may not be corrosive despite the low or high pH value, the non-corrosive classification still needs to be confirmed by other data (preferably by data from an appropriate validated *in vitro* test method).

4.8. Module 8: Non-testing data on serious eye damage and eye irritation

4.8.1. (Q)SAR, expert systems, grouping and read-across (substances)

66. Non-testing methods can be used if they provide adequate, relevant and reliable data for serious eye damage and eye irritation for the substance of interest. For substances, the non-testing methods can be divided into two different categories:

- Read-across using grouping of substances[17], and
- Qualitative and quantitative Structure-Activity-Relationships ((Q)SAR) as well as expert and other prediction systems that often incorporate multiple SARs, QSARs, expert rules and/or data.

67. With the introduction of the OECD (Q)SAR Toolbox[18] in combination with the eChemPortal[19], useful tools are provided for:

- Finding existing data on the substance under question (target),
- Identifying analogues for potential read-across and grouping and finding existing data on these analogues,
- Applying a number of SARs and other profilers for serious eye damage and irritation to the target structure,
- Grouping and deriving simple (Q)SAR or trend relationships.

68. Guidance on how to apply (Q)SARs for regulatory use and on how to assess the validity and suitability of (Q)SAR models and adequacy of their predictions is provided in the OECD GD 69 (OECD, 2007) and is also available from the corresponding section of the OECD website[20]. Other useful guidance documents have also been published to aid in determining how and when to

17 Category approach and analogue approach can be used for the grouping of substances, whilst the term read-across is reserved for a technique to fill data gaps, i.e. to transfer knowledge from one or more substances called source(s) to another substance with data gap, named target substance.

18 http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm

19 http://www.echemportal.org

20 http://www.oecd.org/env/ehs/risk-assessment/guidancedocumentsandreportsrelatedtoqsars.htm

apply QSAR models[21,22]. Together, these resources can help inform a determination of whether a (Q)SAR result might be used to replace a test result. Furthermore, examples of how to build and report grouping of substances and read-across are also available[23].

69. The mechanism of serious eye damage/eye irritation involves toxicodynamic and toxicokinetic parameters. Some (Q)SAR models predict serious eye damage and eye irritation based on toxicodynamic properties only (e.g. acidity or basicity, electrophilicity, other reactivity, surfactant activity, membrane destruction). These models have to be checked whether they also take into account, or have to be used in combination with models covering toxicokinetic parameters such as potential of a substance to cross relevant outer membranes of the eye (cornea) and to be active in the living tissue underneath. Conversely, some (Q)SAR models predict (the absence of) serious eye damage and/or eye irritation solely from e.g. physico-chemical properties considered to illustrate the toxicokinetic behaviour of a substance and have to be checked whether they also take into account, or to be used in combination with models relying on toxicodynamic properties. Ideally, such models would also take into account the potential for metabolism, autoxidation, or hydrolysis of the parent compound and how that might impact any effects on the eye.

70. For example, the BfR rule-base implemented in Toxtree and the OECD QSAR Toolbox contains both physico-chemical exclusion rules and structure-based inclusion rules (structural alerts). Evaluations of these rules for the prediction/exclusion of eye irritation have been carried out in accordance with the OECD principles for (Q)SAR validation (Tsakovska *et al.*, 2005, on structural alerts; Tsakovska *et al.*, 2007, on physico-chemical exclusion rules). However, inclusion and exclusion rules were evaluated separately, and not used in combination in these works.

71. When applied, these two sets of rules may sometimes provide contradictory information, i.e. a structural alert might indicate serious eye damage and/or eye irritation potential, while at the same time, based on physico-chemical properties, absence of effect is predicted. In such cases, it is recommended to consider additional information (e.g. on the behaviour of chemically similar substances). In other cases, applicability of one (or more) of the physico-chemical exclusion rules might indicate absence of serious eye damage and/or eye irritation potential of the target substance, while no structural alert for serious eye damage and/or eye irritation is triggered. Given that the absence of any known structural alert is not equivalent to the absence of a potential effect, in such a situation the substance should still be examined for potentially reactive substructures (and examining the behaviour of chemical analogues would still be beneficial).

72. While these considerations apply to the use of the BfR rule-base for direct classification/non-classification, less certainty might be required for e.g. a decision on further *in vitro* testing i.e., where the exclusion rules suggest the absence of an effect, a Bottom-Up approach could be followed (see Figure 1).

4.8.1.1. SARs, grouping and read-across for serious eye damage and eye irritation

73. Read-across, SARs and Grouping/Category formation are treated together because they represent approaches based on the same basic concept. Note that, depending on the legal framework and Member Country, specific requirement may be associated to the read-across and grouping approaches. For example, under the EU REACH Regulation, read-across needs to be justified, documented, and supported by reliable data on the source(s), i.e. one or more substances (ECHA, 2015).

---

21 https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

22 https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

23 https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across

74. Toxicological data gaps for a chemical can be filled by prediction based on similar chemicals for which test data are available. While this has historically been accomplished based on structure and physico-chemical properties, mechanistic (biological) similarity is increasingly being used to add confidence to this process. Efforts are ongoing to develop consensus on applying these principles to facilitate their effective use in regulatory context (Ball *et al*., 2016; Zhu *et al*., 2016). Structural alerts are substructures in the substance that are considered to reflect chemical or biochemical reactivity underlining the toxicological effect. The occurrence of a structural alert for a substance suggests the presence of an effect, and structural analogues that have exhibited serious eye damage or irritation potential can be used to predict serious eye damage or eye irritation effects of the substance of interest, or be used to tailor further testing and assessment, as indicated in the Figure 1. Structural alerts for serious eye damage/eye irritation have been described in the literature, e.g. in Gerner *et al*. (2005).

75. The similarity of two substances can also be based for example on a common functional group, common precursors or common break-down products (analogue approach). Grouping requires that toxicological properties of the target substance may be predicted from the data of the source substances, basically by interpolation. Predictions based on read-across may therefore be possible for chemically similar substances if it can be shown that their similarity reflects reactive substructures able to react with ocular tissue, even if that substructure has so far not been coded into a structural alert in any of the available literature or software models. Indeed, knowledge on structural alerts for serious eye damage and irritation is always evolving (in particular where new classes of substances are introduced into the market).

76. While not typically useful for regulatory decisions in isolation, negative data can be useful in certain cases. In these cases it is helpful to consider both the structural and mechanism of action similarity along with applicability domain.

4.8.1.2. QSARs and expert systems for serious eye damage and eye irritation

77. An overview of the available (Q)SARs for serious eye damage/eye irritation is provided in Table 4, and more details can be retrieved in published reviews (ECHA, 2015 – appendix R.7.2-3; Gallegos Saliner *et al*., 2006, 2008). Furthermore, in recent studies QSAR models based on multiple artificial neural network molecular descriptor selection functionalities were developed, to maximize the applicability domain of the battery for the assessment of both eye irritation and serious eye damage potential (Verma and Matthews, 2015a, 2015b). The same authors developed an *in silico* expert system based upon exclusion rules of physico-chemical properties to facilitate the rapid screening and prioritization of test chemicals (Verma and Matthews, 2015c). Predictions from multiple QSAR models in a weight of evidence also allows enhancing the confidence in the prediction.

78. Expert systems are computer programs that guide hazard assessment by predicting toxicity endpoints of certain substance structures based on the available information. They can be based on automated rule-induction systems (e.g., TopKat and MultiCASE), or on a knowledge-based system (e.g. Derek).

79. The freely downloadable OECD QSAR Toolbox software contains two profilers relevant for serious eye damage/eye irritation based on the BfR rule-base. This rulebase is based on the combined use of two predictive approaches: a) physicochemical exclusion rules to identify chemicals with no eye irritation/serious eye damage potential; and b) structural inclusion rules (SARs) to identify chemicals with eye irritation/serious eye damage potential (Gerner *et al*., 2005). The use of a combination of profilers and data for analogues could allow for the prediction of serious eye damage/eye irritation for new substances through a read-across or category approach.

80. Not all of the models were developed for the purpose of UN GHS classification, so that it is important to assess in each case whether the endpoint or effect being predicted corresponds to the regulatory endpoint of interest.

4.8.2. Bridging principles and theory of additivity (mixtures)

81. Non-testing methods for mixtures can be divided into (UN, 2015):

- Bridging, when data are not available for the complete mixture, and
- Theory of additivity, when data are available for the ingredients of the mixture.

82. Bridging principles are used when the mixture itself has not been tested for serious eye damage and irritation, but there are sufficient data on both the individual ingredients and similar tested mixtures to adequately characterise the hazards of the mixtures. The following bridging principles may be used: dilution, batching, concentration of mixtures of the highest serious eye damage/eye irritation category, interpolation within one hazard category, substantially similar mixtures, and aerosols (see chapter 3.3.3.2 of UN, 2015).

83. The theory of additivity is used when data are available for all or only some of the ingredients, but not on the mixture as a whole. It assumes that each ingredient inducing serious eye damage or eye irritation contributes to the overall serious eye damage and/or irritation properties of the mixture in proportion to its potency and concentration. When applying such theory, considerations on the quality of the data of the ingredients is critical (e.g., data reported in Safety Data Sheets may be based on *in vivo*, *in vitro* or no test data). The mixture is classified as inducing serious eye damage and/or irritation when the sum of the concentrations of the relevant ingredients exceeds a pre-set cut-off value / concentration limit (see chapter 3.3.3.3 of UN, 2015).

Table 4: Overview of available (Q)SARs for serious eye damage and eye irritation. Note that this list is likely to be non-exhaustive and does not imply endorsement by OECD of any of the listed models for a particular prediction.

Source	Applicability domain
Literature Models	
Verma and Matthews (2015a, 2015b)	Based on ADMET Predictor program
Solimeo *et al.* (2012)	N.A.
Gerner *et al*. (2005)	Based on physico-chemical values
Abraham *et al.* (2003)	Pure bulk liquids
Barratt (1995, 1997)	Neutral organic chemicals
Computerised Models – Freely available	
BfR rule base: included in the OECD QSAR Toolbox and Toxmatch, Toxtree, ToxPredict and Ambit*	EU New Chemicals (NONS) database, organic chemicals with no significant hydrolysis potential and purity > 95 %

Source	Applicability domain
PaDEL-DDPredictor (Liew and Yap, 2013)	Calculated by the model based on the range of descriptors
US FDA (Verma and Matthews, 2015c)	Based on physico-chemical properties
Computerised Models - Commercial	
ACD/Percepta	Organic chemicals
Derek Nexus	Organic chemicals and some metals
HazardExpert	Organic chemicals
Molcode	Organic chemicals
MultiCASE	Organic chemicals
TopKat	Organic chemicals
Review papers	
Gallegos Saliner *et al*. (2006, 2008)	N.A.
Patlewicz *et al*. (2003)	N.A.

N.A. – Not Applicable. A detailed description of the above models is given in Appendix R.7.2-3 of the ECHA IR/CSA guidance 7a (ECHA, 2015, p. 252-257).* Underwent independent assessment.

4.8.3. Overview of non-testing data on serious eye damage and eye irritation

Module 8 – Non-testing data on serious eye damage and irritation	
Regulatory use (UN GHS classification)	*Substances* Usually used as supporting information in a weight of evidence approach. *Mixtures* To be used for classification decision on serious eye damage (UN GHS Cat. 1), eye irritation (UN GHS Cat. 2 and UN GHS Cat. 2A and 2B), and no need for classification (UN GHS No Cat.).
Validation & regulatory status	*Substances* Validation and regulatory acceptance based on case-by-case. *Mixtures* Regulatory adopted approach.
Potential role in the IATA	*Substances* Non-testing methods are usually used as supporting information in a WoE approach, e.g., to support observations from available data from other *in vivo* test methods (Module 4) and to support *in vitro* results on serious eye damage and/or eye irritation (Modules 3 and 5). If further testing is required, information generated with this Module may be used for deciding how to address Part 3 i.e., to initiate a Top-Down or a Bottom-Up approach (Figure 1). *Mixtures* To be used when data are not available on the complete mixture or when data are available for all or some ingredients of the mixture. Furthermore, when validated *in vitro* test methods for serious eye damage and eye irritation are available, these may

Module 8 – Non-testing data on serious eye damage and irritation	
	be used to generate data to classify the mixture instead of or in conjunction with the non-testing methods.
Description	*Substances* - Analogue approaches (read-across, SARs, and grouping). - (Q)SARs. - Expert and other prediction systems that often include several (Q)SARs, expert rules and data. *Mixtures* - Bridging principles - Theory of additivity
Scientific basis incl. MoA	*Substances:* Mainly correlative approaches based on the general assumption that substances with comparable structural properties have comparable serious eye damage and/or eye irritation properties. However this might change once the Adverse Outcome Pathway (AOP) project (OECD, 2013g) has made further progress or more (Q)SARs might become available built on mechanistically based high-throughput *in vitro* data. *Mixtures:* Bridging principles are used when there are sufficient data on both the individual ingredients and similar tested mixtures to adequately characterise the hazards of the mixtures. The following bridging principles may be used: based on dilution, batching, concentration of the highest corrosion/irritation category, interpolation within one hazard, substantially similar mixtures, and aerosols. The theory of additivity is used when data are available on the ingredients, but not on the mixture as a whole. It assumes that each ingredient inducing serious eye damage and/or eye irritation contributes to the overall serious eye damage and/or irritation properties of the mixture in proportion to its potency and concentration. The mixture is classified as inducing serious eye damage or eye irritation when the sum of the concentrations of the relevant ingredients exceeds a cut-off value / concentration limit (see chapter 3.3.3.3 of UN, 2015).
Strengths and weaknesses	<u>Strengths</u> *Substances and mixtures* - Ease of application. - Low cost. <u>Weaknesses</u> *Substances* Results may be less relevant compared to experimental data, depending on the substance as well as the non-testing method and its underlying (model development/validation) dataset. *Mixtures* An impact assessment carried out by A.I.S.E. (Cazelle *et al*., 2014) showed that the use of the UN GHS theory of additivity for classification of detergent and cleaning products can result in the over-labelling of many products currently not requiring classification according to consistent animal, *in vitro* and human data. Similar findings were reported for agrochemical formulations (Corvaro *et al.*, submitted)

Module 8 – Non-testing data on serious eye damage and irritation	
Applicability domain and limitations	<u>Applicability</u> *Substances* Model-specific and needs to be defined in a (Q)SAR Model Reporting Format (QMRF). Also (Q)SAR Prediction Reporting Format (QPRF) are used to describe whether a prediction for a specific substance should be regarded as within the Applicability Domain or not. Application of these non-testing approaches is rather straight-forward for mono-constituent substances, whereas for multi-constituent substances, this only holds true if the composition of the substance is known (i.e. percentage of each of the discrete organic constituents) because then predictions can be performed on each constituent and the effect of the multi-constituent substance predicted by employing a dose addition approach. For Substances of Unknown and Variable Composition and Biologicals (UVCB), by definition, not all of the constituents are known with respect to their identity and/or their relative concentrations. (Q)SAR models and grouping approaches have, however, been employed on multi-constituent substances and UVCBs with partly unknown composition details for other endpoints than serious eye damage and irritation by accepting some uncertainty and assuming that all constituents of the considered UVCBs are represented by a few known constituents/groups of constituents, on which QSAR models or grouping approaches then could be employed. *Mixtures* The bridging principle is applicable to mixtures having data on both their individual ingredients and similar tested mixtures. The theory of additivity is applicable to mixtures that have data available for all or for some ingredients. <u>Limitations</u> *Substances* - Limited applicability to the UN GHS classification scheme. - Applicability limited to the applicability domain of the model. *Mixtures* Need to have sufficient data on similar tested mixtures as well as the ingredients of the mixture.
Predictive capacity	*Substances* Model-, domain- and context-specific. *Mixtures* - Only limited data available. An impact assessment carried out by A.I.S.E. showed that the use of the UN GHS theory of additivity for classification of detergent and cleaning products can result in the over-labelling of many products currently not requiring classification according to consistent animal, *in vitro* and human experience data. Furthermore, a retrospective analysis of 225 agrochemical formulations indicated that, while overpredictive across categories, the use of the UN GHS theory of additivity for classification of agrochemical formulation can provide value for the identification of UN GHS No Cat. consistent with the classification based on *in vivo* animal test (Corvaro *et al*., submitted).

4.9. Module 9: Phases and elements of Weight of Evidence (WoE) approaches

84. A weight of evidence determination means that all available and scientifically justified information bearing on the determination of hazard is considered together. In case of serious eye damage and eye irritation this includes structural information, information on physico-chemical parameters (e.g., pH, acid/alkaline reserve), information from category approaches (e.g., grouping) or read-across, (Q)SAR results, the results of suitable *in vitro* tests, relevant animal data, and human data. The quality and consistency of the data should be taken into account when weighing each piece of available information. Information such as study design, mechanism or mode of action, dose-effect relationships and biological relevance may be considered. Both positive and negative results can be assembled together in a single weight of evidence determination. Evaluation must be performed on a case-by-case basis and with expert judgement. In case of inconsistent data, the quality and relevance has to be carefully assessed in order to derive a conclusion. No formula can be presented for this analysis; a detailed explanation of the expert judgment used to overrule e.g. a single positive finding should accompany the derived conclusion.

85. A WoE approach may involve an assessment of the relative values/weights of different pieces of the available information that has been retrieved and gathered in previous steps (for an example see Hulzebos and Gerner, 2010). These weights/values can be a*ssign*ed either in a more objective way by applying a formalised procedure (e.g., based on Bayesian logic, as in Rorije *et al*., 2013), by using meta-analyses (either weighted or unweighted) or by using expert judgement. Examples of tools to evaluate the quality include the Klimisch scores for experimental data (Klimisch *et al*., 1997) and Hill's criteria for evaluation of epidemiological data (Hill, 1965), as well as the JRC's ToxRTool for scoring *in vivo* and *in vitro* data (Schneider *et al*., 2009). Under the GHS (UN, 2015), in sub-chapter 3.3.2.2 a weight of evidence approach based on expert judgement is also recommended.

86. The weight given to the available evidence will be influenced by factors such as the quality of the data, consistency of results/data, nature and severity of effects, relevance of the information for the given regulatory endpoint. For each study/data, the relevance, reliability and adequacy for the purpose have to be considered. All available information that can contribute to the determination of classification for an endpoint is considered together. In the following paragraphs a suggestion of the steps and elements of WoE is given.

4.9.1. Place/role of WoE in the IATA

87. WoE should be carried out before any new prospective *in vitro* or *in vivo* testing is performed. A combination of physico-chemical information, (Q)SAR, read-across, grouping information and/or existing *in vivo*, *in vitro* and/or human data might be considered sufficient to conclude on serious eye damage and eye irritation effects.

4.9.2. Coverage of relevant sources of information

88. The IATA specifies several types of existing information that can be used, with the condition that these are of sufficient quality. Structural information, physico-chemical properties, data on structurally-related chemicals obtained by read-across or grouping approaches, (Q)SAR modelling data, existing human and relevant laboratory animal data as well as *in vitro* data are listed. In the WoE analysis, the availability of specified types of data should be checked. The sources of those data obviously vary, ranging from clinical study reports, scientific publications, data from poison information centres, guideline tests, up to worker surveillance data of the chemical companies.

4.9.3. Assessment of data quality

89. The quality of the data that is obtained for a WoE needs to be assessed, since the quality will contribute to the value/weight of each data element. In case the quality of a certain study is deemed to be inappropriate, it is recommendable not to consider those data in the WoE, but focus on other pieces of information which are of sufficient quality. Quality might be inappropriate e.g., due to negative outcome in the validation of the methodology, "non-adherence" to the relevant test guideline/method, lack of adequate controls, deficiencies in data reporting etc. Furthermore, quality may need to be evaluated based on expert judgement in case of e.g incomplete or unavailable validation of a test method.

90. The quality of the study, the method, the reporting of the results, and the conclusions that are drawn, must be evaluated carefully. Reasons why existing study data may vary in quality include the use of outdated test guidelines, the failure to characterise the test chemical properly (in terms of purity, physical characteristics, etc.) and the use of crude techniques/procedures that have since become refined, moreover, other reasons could be poor reporting of information and poor quality assurance.

91. For many existing test chemicals, at least some of the available information could have been generated prior to the requirements of Good Laboratory Practice (GLP) and the standardisation of testing methods. While such information may still be usable, both the data and the methodology used must be evaluated in order to determine their reliability. Such an evaluation would ideally require an evidence-based evaluation i.e., a systematic and consistent evaluation following pre-defined, transparent and independently reviewed criteria before making decisions. These should always include justifications for the use of particular data sets on the basis of the criteria-based evaluation. For some test chemicals in contrast, information may be available from tests conducted according to OECD Test Guidelines (or other standards like CEN, ISO, ASTM, OSPAR methods, national standard methods), and in compliance with the principles of GLP or equivalent standards.

4.9.5. Adequacy of information

92. Adequacy defines the usefulness of information for the purpose of hazard and risk assessment, in other words whether the available information allows clear decision-making about whether the test chemical induces (or not) serious eye damage and eye irritation and an adequate classification can be derived. The evaluation of adequacy of test results and documentation for the intended purpose is particularly important for test chemicals where there may be (a number of) results available, but where some or all of them have not been carried out according to current standards. Where there is more than one study, the greatest weight is attached to the studies that are the most relevant and reliable. For each endpoint, robust summaries need to be prepared for the key studies. Sound scientific judgement is an important principle in considering the adequacy of information and determining the key study.

4.9.6. Non-testing data

(Q)SAR data

93. It is important to distinguish between the proposed validity of the (Q)SAR model *per se*, and the reliability and adequacy of an individual (Q)SAR estimate (i.e., the application of the (Q)SAR model to a specific substance), and the appropriateness of the documentation (e.g., QMRF) associated with models and their predictions.

94. Guidance on how to characterise (Q)SAR models according to the OECD (Q)SAR validation principles is provided in the OECD GD 69 (OECD, 2007). Other useful guidance has also been published to aid in determining how to use and report on QSAR models[24].

95. The information in the QMRF and QPRF should be used when assessing whether a prediction is adequate for the purpose of classification and labelling and/or risk assessment. The assessment will also need to take into account the regulatory context. This means that the assessments of (Q)SAR validity (typically proposed in scientific publications) and (Q)SAR estimate reliability need to be supplemented with an assessment of the relevance of the prediction for the regulatory purposes, which includes an assessment of *completeness*, i.e., whether the information is sufficient to make the regulatory decision, and if not, what additional (experimental) information is needed. The decision will be taken on a case-by-case basis.

96. (Q)SAR predictions may be gathered from databases (in which the predictions have already been generated and documented) or generated *de novo* through the available models.

Data obtained by grouping approaches

97. Conclusions about the likely properties of a substance can also be based on the knowledge of the properties of one or more similar chemicals, by applying *grouping methods*.

98. The corresponding OECD guidance document No. 194 provides information on the use of grouping of chemicals and read-across approaches (OECD, 2014b).

99. As with (Q)SARs, grouping approaches can be used to indicate either the presence or the absence of an effect.

4.9.7. Existing human data

100. The strength of the epidemiological evidence for specific health effects depends, among other things, on the type of analyses and on the magnitude and specificity of the response. Human data other than from epidemiological studies can be obtained from e.g., case reports, clinical studies, occupational disease registries or other occupational surveillance schemes and from poison centre information. In principle all types of toxic effects can be reported in such studies. Confidence in the findings is increased when comparable results are obtained in several independent studies on populations exposed to the same agent under different conditions. Other characteristics that support causal associations are the presence of a dose-response relationship, a consistent correlation in time and (biological) plausibility, i.e., aspects covered by epidemiological criteria such as those described by Hill (1965), Fedak *et al*. (2015) and Lucas & McMichael (2005).

101. A comprehensive guidance of both the evaluation and use of epidemiological evidence for risk assessment purposes is provided by Kryzanowski *et al.* (WHO, 2000).

102. High quality human data may be considered as one of the strongest basis for classification and labelling decision making (subject to the ethical considerations relevant for the respective regulatory programme). However, when contradictory human and animal (OECD TG 405) data are available and WoE analysis including all other existing data and (Q)SAR profiling is not conclusive towards one or the other result, confirmatory *in vitro* testing should be performed.

103. It is emphasised that testing with human volunteers for hazard identification is strongly discouraged for ethical reasons, but data from accidental human exposures, while not necessarily of the highest quality, can be used to support WoE conclusions.

[24] https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

4.9.8. Evaluation of consistency of the data

104. The consistency of the existing data coming from various sources is crucial and should therefore be thoroughly evaluated in a WoE analysis. Consistent data which come from several studies/sources may be considered sufficient for regulatory purposes. In case the data elements are of comparable weight but give inconsistent evidence (e.g., (Q)SAR is positive and available limited human data is negative), usually WoE analysis will not be conclusive and prospective *in vitro* and/or *in vivo* testing will have to be conducted (Part 3 of the IATA). In case the weights of the individual pieces of evidence differ considerably, a WoE conclusion may be drawn according to the evidence carrying the highest weight. If high quality human (Module 1), *in vitro* (Module 3) and/or *in vivo* (Module 2) data are available, these should carry the highest weight in the WoE assessment.

4.9.10. Assessment of the coverage of relevant parameters and observations

105. In a standard *in vivo* test guideline the required parameters/observations have been specified and often build the basis for decision making (e.g., classification and labelling for serious eye damage and eye irritation is mainly derived from the *in vivo* rabbit eye test). However, when taking together (in an integrating phase), it is not always possible to extract information equivalent to those parameters from non-testing data. Therefore, an important element of WoE is to consider to what extent the parameters and observations were addressed by each data element of the WoE.

4.9.11. Conclusions of WoE

106. In the final analysis of the WoE, each data element will be characterised for its quality, relevance, coverage (e.g., serious eye damage and eye irritation) and associated uncertainty. The assessor would either decide to include or exclude the existing information based on these. When consistency is seen among ”qualified” data elements, WoE may reach a conclusion that the relevant endpoint or information requirement has been sufficiently covered and further testing is not necessary. When on the other hand, insufficient information remains after the ”non-qualified” data have been rejected/put aside and/or when the remaining information is inconsistent or contradictory, WoE would reach to a conclusion that the relevant endpoint or information requirement has not been sufficiently covered and further testing is necessary, depending on the specific legal/regulatory framework, and inform on which test to conduct to fill the data gap.

107. The WoE assessment needs to be transparently explained and documented to enable a logical flow leading to the decision/conclusion. An example for a simple approach to the documentation of the WoE is presented in Annex 1. Furthermore examples of evaluations are given for detergents and agrochemical mixtures in annex 2.

5. REFERENCES

Abraham M.H., Hassanisadi M., Jalali-Heravi M., Ghafourian T., Cain W.S., Cometto-Muniz J.E. (2003) Draize rabbit eye test compatibility with eye irritation thresholds in humans: a quantitative structure-activity relationship analysis. Toxicol. Sci. 76, 384-91.

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.

Adriaens E., Bytheway H., De Wever B., Eschrich D., Guest R., Hansen E., Vanparys P., Schoeters G., Warren N., Weltens R., Whittingham A., Remon J.P. (2008). Successful prevalidation of the slug mucosal irritation test to assess the eye irritation potency of chemicals. Toxicology *In Vitro* 22, 1285-1296.

Adriaens E., Dhondt M.M., Remon J.P. (2005). Refinement of the Slug Mucosal Irritation test as an alternative screening test for eye irritation. Toxicology *In Vitro* 19, 79-89.

A.I.S.E. (2006). Low Volume Eye Test (LVET): A.I.S.E. Submission to ECVAM, January 2006.

Alépée N., Barroso J., De Smedt A., De Wever B., Hibatallah J., Klaric M., Mewes K.R., Millet M., Pfannenbecker U., Tailhardat M., Templier M., McNamee P. (2015) Use of HPLC/UPLCSpectrophotometry for Detection of Formazan in *In vitro* Reconstructed Human Tissue (RhT)-Based Test Methods Employing the MTT-Reduction Assay to Expand their Applicability to Strongly Coloured Test Chemicals. Toxicology *In Vitro* 29, 741-761.

Alépée N., Leblanc V., Adriaens E., Grandidier M.H., Lelièvre D., Meloni M., Nardelli L., Roper C.S., Santirocco E., Toner F., Van Rompay A., Vinall J., Cotovio J. (2016a). Multi-laboratory validation of SkinEthic HCE test method for testing serious eye damage/eye irritation using liquid chemicals. Toxicology *In Vitro 31, 43-53.*

Alépée N., Adriaens E., Grandidier M.H., Meloni M., Nardelli L., Vinall C.J., Toner F., Roper C.S., Van Rompay A.R., Leblanc V., Cotovio J. (2016b). Multi-laboratory evaluation of SkinEthic HCE test method for testing serious eye damage/eye irritation using solid chemicals and overall performance of the test method with regard to solid and liquid chemicals testing. Toxicology *In Vitro* 34, 55-70.

Bagley D.M., Bruner L.H., Silva O. de, Cottin M., O'Brien K.A.F., Uttley M., Walker A.P. (1992). An evaluation of five potential alternatives *in vitro* to the rabbit eye irritation test *in vivo*. Toxicology *In Vitro* 6, 136-149.

Bagley D.M., Cervin D., Harbell J.W. (1999). Assessment of the chorioallantoic membrane vascular assay (CAMVA) in the COLIPA *in vitro* eye irritation validation study. Toxicology *In Vitro* 13, 285-293.

Ball N., Cronin M.T.D., Shen J., Adenuga M.D., Blackburn K., Booth E.D., Bouhifd M., Donley E., Egnash L., Freeman J.J., Hastings C., Juberg D.R., Kleensang A., Kleinstreuer N., Kroese E.D., Lee A.C., Luechtefeld T., Maertens A., Marty S., Naciff J.M., Palmer J., Pamies D., Penman M., Richarz A.-N., Russo D.P., Stuard S.B., Patlewicz G., van Ravenzwaay B., Wu S., Zhu H., Hartung T. (2016). Toward good read-across practice (GRAP) guidance. ALTEX 33,149-166.

Barratt M.D. (1995) The role of structure-activity relationships and expert systems in alternative strategies for the determination of skin sensitisation, skin corrosivity and eye irritation. Altern. Lab. Anim. 23, 111-22.

Barratt M.D. (1997) QSARs for the Eye Irritation Potential of Neutral Organic Chemicals. Toxicology *In Vitro* 11, 1-8.

Barroso J., Alépée N., Cole T., Eskes C., Freeman S.J., Liška R., McNamee P., Pfannenbecker U., Reus A.A., Rubingh C.M., Schaeffer M.W., Zuang V. (2014). The EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Cornea-like Epithelium (RhCE)-based test methods for identifying chemicals not requiring classification and labelling for serious eye damage/eye irritation: Validation Study Report. EUR 28125 EN; doi:10.2787/41680. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC100280.

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.

Bartok M., Gabel D., Zorn-Kruppa M., Engelke M. (2015). Development of an *in vitro* ocular test system for the prediction of all three GHS categories. Toxicology *In Vitro* 29, 72-80.

Berta A. (1992). Chapter 17: Enzymatic changes in tears in caustic injuries. In Enzymology of the tears. Boca Raton, CRC Press, p. 285-295.

Blazka M.E., Harbell J.W., Klauzner M., Raabe H., Kubilus J., Hsia F., Minerath B., Kotler M., Bagley D.M. (2000). Colgate-Palmolive's program to validate the EpiOcularTM human tissue construct model. The Toxicologist 54, 188.

Blazka M., Harbell J.W., Klausner M., Merrill J.C., Kubilus J., Kloss C., Bagley D.M. (2003). Evaluating the ocular irritation potential of 54 test articles using the EpiOcular human tissue construct model (OCL-200). The Toxicologist 72, 221.

Brantom P.G., Bruner L.H., Chamberlain M., Desilva O., Dupuis J., Earl L.K., Lovell D.P., Pape W.J.W., Uttley M., Bagley D.M., Baker F.W., Brachter M., Courtellemont P., Declercq L., Freeman S., Steiling W., Walker A.P., Carr G.J., Dami N., Thomas G., Harbell J., Jones P.A., Pfannenbecker U., Southee J.A., Tcheng M., Argembeaux H., Castelli D., Clothier R., Esdaile D.J., Itigaki H., Jung K., Kasai Y., Kojima H., Kristen U., Larnicol M., Lewis R.W., Marenus K., Moreno O., Peterson A., Rasmussen E.S., Robles C., Stern M. (1997). A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. Toxicology *In Vitro* 11, 141-179.

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.

Cerven D., Moreno O. (1998). Bovine corneal opacity and permeability test validation as an alternative to the Draize eye irritation assay. In Adv Anim Altern Saf Efficacy Test (ed. H. Salem & S.A. Katz), pp. 261-267. Washington, DC, USA: Taylor & Francis.

Clippinger A.J., Hill E., Curren R., Bishop P. (2016). Bridging the gap between regulatory acceptance and industry use of non-animal methods. ALTEX 33, 453-458.

Cormier E.M., Hunter J.E., Billhimer W., May J., Farage M.A. (1995) Use of clinical and consumer eye irritation data to evaluate the low-volume eye test. J. Toxicol. Cut. Ocul. Toxicol. 14, 197-205.

Cormier E.M., Parker R.D., Henson C., Cruze L.W., Merritt A.K., Bruce R.D., Osborne R. (1996). Determination of the intra- and inter-laboratory reproducibility of the Low Volume Eye Test and its statistical relationship to the Draize tes. Reg. Tox. Pharmac. 23, 156-161.

Corvaro M., Gehen S., Macleod F., Metha J, (submitted). A retrospective analysis of *in vivo* skin and eye irritation and skin sensitisation studies with agrochemical formulations: setting the scene for development of alternative strategies.

DB-ALM (2013). Protocol n° 71: Fluorescein Leakage Test. European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), Ispra, Italy: Available at: http://ecvam-dbalm.jrc.ec.europa.eu

DB-ALM (2015). Protocol n° 164: EpiOcular™ Eye Irritation Test – Summary. European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), Ispra, Italy: Available at: http://ecvam-dbalm.jrc.ec.europa.eu

Draize J.H., Woodard G., Calvery H.O. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. Journal of Pharmacology and Experimental Therapeutics 82, 377-390.

ECETOC (2002) ECETOC Monograph No.32, Use of Human Data in Hazard Classification for Irritation and Sensitisation. Available at: http://www.ecetoc.org/publications.

ECHA (2015). Skin corrosion / irritation, serious eye damage / eye irritation and respiratory tract corrosion / irritation. In: Guidance on Information Requirements and Chemical Safety Assessment - Chapter R.7a: endpoint specific guidance, Version 4.1, October 2015, p. 168-235. Available at: http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf.

ESAC (2007). Statement on the conclusion of the ICCVAM retrospective study on organotypic *in vitro* assays as screening tests to identify potential ocular corrosives and severe eye irritants. Available at: https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/scientific-advice-stakeholders-networks/ecvam-scientific-advisory-committee-esac/statements-opinions.

ESAC (2009a). Statement on the use of existing Low Volume Eye Test (LVET) data for Weight-of-Evidence decisions on classification and labelling of cleaning products and their main ingredients. Available at: https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/scientific-advice-stakeholders-networks/ecvam-scientific-advisory-committee-esac/statements-opinions.

ESAC (2009b). Statement on the scientific validity of cytotoxicity/cell-function based *in vitro* assays for eye irritation testing. Available at: https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/scientific-advice-stakeholders-networks/ecvam-scientific-advisory-committee-esac/statements-opinions.

ESAC (2009c). ESAC Peer Review Panel consensus report the retrospective validation of cytotoxicity/cell function based *in vitro* assays. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-eye-irritation/1_Final%20draft%20PRP%20CBA.pdf.

ESAC (2014). ESAC Opinion on the EURL ECVAM Eye Irritation Validation Study (EIVS) on EpiOcular™ EIT and SkinEthic™ HCE and a related Cosmetics Europe study on HPLC/UPLC-

spectrophotometry as an alternative endpoint detection system for MTT-formazan. ESAC Opinion No. 2014-03 of 17 November 2014; EUR 28173 EN; doi:10.2787/043697. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC103702.

ESAC (2016a). ESAC Opinion on the SkinEthic™ Human Corneal Epithelium (HCE) Eye Irritation Test (EIT). ESAC Opinion No. 2016-02 of 24 June 2016; EUR 28175 EN; doi:10.2787/390390. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC103704.

ESAC (2016b). ESAC Opinion on the Ocular Irritection® test method for prediction of serious eye damage/ eye irritation potential of chemicals. ESAC Opinion No. 2016-01 of 24 June 2016; EUR 28174 EN; doi:10.2787/991443. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC103703.

Eskes C., Hoffmann S., Facchini D., Ulmer R., Wang A., Flego M., Vassallo M., Bufo M., van Vliet E., d'Abrosca F., Wilt N. (2014). Validation Study on the Ocular Irritection® Assay for Eye Irritation Testing. Toxicology *In Vitro* 28, 1046-1065.

EURL ECVAM (2008a). Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing. Available under Validation & Regulatory Acceptance, Section Topical Toxicity - Eye Irritation at: https://eurl-ecvam.jrc.ec.europa.eu.

EURL ECVAM (2008b). Cytotoxicity/cell function based *in vitro* assays: Recommendations from the Validation Management Group. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-eye-irritation/VMG%20CBA%20recommendations%20-%20final.pdf.

Fedak K.M., Bernal A., Capshaw Z.A., Gross S. (2015). Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerging Themes in Epidemiology 12, 14. doi: 10.1186/s12982-015-0037-4.

Freeberg F.E., Griffith J.F., Bruce R.D., Bay P.H.S. (1984) Correlation of animal test methods with human experience for household products. J. Toxicol. Cut. Ocul. Toxicol. 1, 53-64.

Freeberg F.E., Hooker D.T., Griffith J.F. (1986a) Correlation of animal eye test data with human experience for household products: an update. J. Toxicol. Cut. Ocul. Toxicol. 5, 115-23.

Freeberg F.E., Nixon G.A., Reer P.J., Weaver J.E., Bruce R.D., Griffith J.F., Sanders III L.W. (1986b) Human and rabbit eye responses to chemical insult. Fund Appl. Toxicol. 7, 626-34.

Furukawa M., Sakakibara T., Itoh K., Kawamura K., Sasaki S., Matsuura M. (2015). Histopathological evaluation of the ocular-irritation potential of shampoos, make-up removers and cleansing foams in the bovine corneal opacity and permeability assay. J. Toxicol. Pathol. 28, 243-248.

Gallegos Saliner A., Patlewicz G., Worth A.P. (2006). Review of literature-based models for skin and eye irritation and corrosion. JRC report EUR 22320 EN. European Chemicals Bureau, Ispra, Italy.

Gallegos Saliner A., Patlewicz G., Worth A.P. (2008) A Review of (Q)SAR Models for Skin and Eye Irritation and Corrosion. QSAR Comb. Sci. 27, 49-59.

Gerner I., Liebsch M., Spielmann H. (2005). Assessment of the Eye Irritating Properties of Chemicals by Applying Alternatives to the Draize Rabbit Eye Test: The Use of QSARs and *In vitro* Tests for the Classification of Eye Irritation. Alternatives to Laboratory Animals 33, 215-237.

Gershbein L.L., McDonald J.E. (1977). Evaluation of the corneal irritancy of test shampoos and detergents in various animal species. Food and Cosmetics Toxicology 15, 131-134.

Griffith J.F., Nixon G.A., Bruce R.D., Reer P.J., Bannan E.A. (1980). Dose-response studies with chemical irritants in the albino rabbit eye as basis for selecting optimum testing conditions for predicting hard to the human eye. Toxicology and Applied Pharmacology 55, 501-513.

Guo X., Yang X.F., Yang Y., Hans R., Cai J.H., Xue J.Y., Tan X.H., Xie X.P., Xiong X.K., Huang J.M. (2010). Prediction of ocular irritancy of 26 chemicals and 26 cosmetic products with isolated rabbit eye (IRE) test. Biomed. Environ. Sci. 25, 359-366.

Harbell J.W., Koontz S.W., Lewis R.W., Lovell D., Acosta D. (1997) IRAG Working Group 4. Cell cytotoxicity assays. Food and Chemical Toxicology 35, 79-126.

Harbell, J.W., Mun G., Curren, R.D. (2006), Application of histological evaluation to enhance the bovine opacity and permeability (BCOP) assay. Toxicologist 90: 326. Available at: http://www.iivs.org/pages/publication_view.php?doc_id=135.

Hayashi K., Mori T., Abo T., Koike M., Takahashi Y., Sakaguchi H., Nishiyama N. (2012a). 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. J. Toxicol. Sci. 37, 269-280.

Hayashi K., Mori T., Abo T., Ooshima K., Hayashi T., Komano T., Takahashi Y., Sakaguchi H., Takatsu A., Nishiyama N. (2012b). Two-stage bottom-up tiered approach combining several alternatives for identification of eye irritation potential of chemicals including insoluble or volatile substances. Toxicology *In Vitro* 26, 1199-1208.

Hill A.B. (1965). The environment and disease: association or causation? Proc. R. Soc. Med. 58, 295−300.

Hulzebos E., Gerner I. (2010). Weight factors in an Integrated Testing Strategy using adjusted OECD principles for (Q)SARs and extended Klimisch codes to decide on skin irritation classification. Regulatory Toxicology and Pharmacology 58,131-144.

ICCVAM (2006). Test Method Evaluation Report: *In vitro* Ocular Toxicity Test Methods for Identifying Ocular Severe Irritants and Corrosives. NIH Publication No.: 07-4517. Available: http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/ocular/in-vitro/tmer/index.html.

ICCVAM (2010a). ICCVAM Test Method Evaluation Report: Recommendations to Discontinue Use of the Low Volume Eye Test for Ocular Safety Testing. NIH Publication No. 10-7515. Available at: https://ntp.niehs.nih.gov/iccvam/docs/ocutox_docs/lvet/front.pdf.

ICCVAM (2010b). Test Method Evaluation Report: Current Validation Status of *In vitro* Test Methods Proposed for Identifying Eye Injury Hazard Potential of Chemicals and Products. National Institutes of Health Publication Number 10-7553A. National Toxicology Program, North Carolina, USA. Available at: http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/ocular/in-vitro-test-methods/index.html.

ICCVAM (2010c). ICCVAM Test Method Evaluation Report: Current validation status of a proposed *in vitro* testing strategy for U.S. Environmental Protection Agency ocular hazard classification and labeling of antimicrobial cleaning products. NIH Publication No. 10-7513. Available at: http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/ocular/amcp/index.html.

ICCVAM (2013). Short Time Exposure (STE) Test Method Summary Review Document, NIH. Available at: http://www.ntp.niehs.nih.gov/iccvam/docs/ocutox_docs/STE-SRD-NICEATM-508.pdf.

ILSI Health and Environmental Sciences Institute (1996). Replacing the Draize eye irritation test : scientific background and research needs. J. Toxicol. Cut. & Ocular Toxicol. 15, 211-234.

Ishii S., Ishii K., Nakadate M., Yamasaki K. (2013). Correlation study in skin and eye irritation between rabbits and humans based on published literatures. Food Chem. Toxicol. 55, 596-601.

Jester J.V., Li, H.F., Petroll, W.M., Parker, R.D., Cavanagh, H.D., Carr, G.J., Smith, B., Maurer, J.K. (1998). Area and Depth of Surfactant Induced Corneal Injury Correlates with Cell Death. Invest. Ophthalmol. Vis. Sci. 39, 922–936.

Katoh M., Hamajima F., Ogasawara T., Hata K. (2013). Establishment of a *new in vitro* test method for evaluation of eye irritancy using a reconstructed human corneal epithelial model, LabCyte CORNEA-MODEL. Toxicology *In Vitro* 27, 2184-2192.

Klimisch H.J., Andreae M., Tillmann U. (1997) A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology 25, 1-5.

Kojima H, Hayashi K., Sakaguchi H., Omori T., Otoizumi T., Sozu T., Kuwahara H., Hayashi T., Sakaguchi M., Toyoda A., Goto H., Watanabe S., Ahiko K., Nakamura T., Morimoto T.. (2013). Second-Phase Validation of Short Time Exposure Tests for Assessment of Eye Irritation Potency of Chemicals. Toxicology *In Vitro* 27, 1855-1869.

Kolle S.N., Kandárová H., Wareing B., van Ravenzwaay B., Landsiedel R. (2011). In-house validation of the EpiOcular(TM) eye irritation test and its combination with the bovine corneal opacity and permeability test for the assessment of ocular irritation. Altern. Lab. Anim. 39, 365-387.

Kolle S.N., Moreno, M.C.R., Mayer, W., van Cott, A., van Ravenzwaay, B., Landsiedel, R. (2015). The EpiOcular™ Eye Irritation Test is the Method of Choice for *In vitro* Eye Irritation Testing of Agrochemical Formulations: Correlation Analysis of EpiOcular™ Eye Irritation Test and BCOP Test Data to UN GHS, US EPA and Brazil ANIVSA Classifications. ATLA 43, 1-18.

Kolle S.N., Sauer U.G., Moreno M.C., Teubner W., Wohlleben W., Landsiedel R. (2016). Eye irritation testing of nanomaterials using the EpiOcular™ eye irritation test and the bovine corneal opacity and permeability assay. Part. Fibre Toxicol. 13, 18. doi: 10.1186/s12989-016-0128-6.

Lemp M.A. (1974). Cornea and sclera. Arch Ophthalmol 92, 158-170.

Lenoir J., Claerhout I., Kestelyn R., Komp A., Remon J.P., Adriaens E. (2011). The slug mucosal irritation (SMI) assay: development of a screening tool for the evaluation of ocular discomfort caused by shampoos. Toxicol *In Vitro* 25, 1919-1925.

Lewis R.W., McCall J.C., Botham P.A. (1993). A comparison of two cytotoxicity tests for predicting the ocular irritancy of surfactants. Toxicology *In Vitro* 7, 155-158.

Liew C.Y., Yap C.W. (2013). QSAR and predictors of eye and skin effects. Mol. Inform. 32, 281-290.

Lucas R.M., McMichael A.J. (2005). Association or causation: evaluating links between “environment and disease”. Bull World Health Organ. 83, 792-795.

Luechtefeld T., Maertens A., Russo D.P., Rovida C., Zhu H., Hartung T. (2016). Analysis of Draize eye irritation testing and its prediction by mining publicly available 2008-2014 REACH data. ALTEX. 33, 123-134.

Luepke N.P. (1985). Hen's egg chorioallantoic membrane test for irritation potential. Food and Chemical Toxicology 23, 287-291.

Marzulli F.N., Ruggles D.I. (1973). Rabbit eye irritation test: collaborative study. J. Ass. Off. Analyt. Chem. 56, 905-914.

Maurer J.K., Parker R.D., 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.

McCulley J.P. (1987). Chemical injuries. In Cornea, scientific foundations and clinical practice (Smolin G, Thoft RA Ed.). Boston, Little, Brown p. 527-542.

Mewes K.R., Engelke M., Zorn-Kruppa M., Bartok M., Tandon R., Brandner J.M., Petersohn D. (*in press*), *In vitro* eye irritation testing using the open source reconstructed hemicornea - a ring trial. ALTEX, 2017 Jan. 14. Epub ahead of print. doi: 10.14573/altex.

Mosmann T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. J. Immunol. Methods 65, 55-63.

Muir C.K., Flower C., Abb N.J. van (1983). A novel approach to the search for *in vitro* alternatives to *in vivo* eye irritancy testing. Toxicology Letters 18, 1-5.

OECD (2005). Series on Testing and Assessment No.34. Guidance Document the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2007). Series on Testing and Assessment No. 69. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2017). Series on Testing and Assessment No. 160. 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 (Second Edition). OECD, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2012a). Guideline for Testing of Chemicals No. 405. Acute Eye Irritation/Corrosion. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2012b). Guidelines for Testing of Chemicals No. 460. Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants. Organisation for Economic Co-operation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2012c). Series on Testing and Assessment No. 180: Streamlined summary document related to the Fluorescein Leakage (FL) test method for identification of ocular corrosives and severe irritants (TG 460). Organisation for Economic Co-operation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2012d). Draft Test Guideline on the Cytosensor Microphysiometer Test Method: An *in vitro* Method for Identifying Ocular Corrosive and Severe Irritant Chemicals as well as Chemicals not Classified as Ocular Irritants. Available at: http://www.oecd.org/env/testguidelines.

OECD (2013a). Guidelines for Testing of Chemicals No. 437. Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage. Organisation

for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2013b). Guideline for Testing of Chemicals No. 438. Isolated Chicken Eye Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2013c). Series on Testing and Assessment No. 189: Streamlined summary document supporting OECD Test Guideline 437 on the Bovine Corneal Opacity and Permeability for eye irritation/corrosion. Organisation for Economic Co-operation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2013d). Series on Testing and Assessment No. 188: Streamlined summary document supporting OECD Test Guideline 438 on the Isolated Chicken Eye for eye irritation / corrosion. Organisation for Economic Co-operation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2013e). Guideline for the Testing of Chemicals No. 122. Determination of pH, Acidity and Alkalinity. OECD, Paris Available at: http://www.oecd.org/env/testguidelines.

OECD (2014a). Series on Testing and Assessment No. 203: Guidance Document on Integrated Approaches to Testing and Assessment of Skin Irritation/Corrosion. Organisation for Economic Co-operation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2014b). Series on Testing and Assessment No. 194. Guidance on Grouping of Chemicals, second edition. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD 2014c. Series on Testing and Assessment No 211. Guidance Document for Describing Non-guideline *In Vitro* Test Methods. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2015a). Guideline for the Testing of Chemicals No. 491. Short Time Exposure *In vitro* Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2015b). Guideline for the Testing of Chemicals No. 492. Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD (2015c). Series on Testing and Assessment No 216. Peformance Standandards for the Assessment of Proposed Similar or Modified *In vitro* Reconstructed human Cornea-like Epithelium (RhCE) test methods for eye hazard. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

OECD, 2016. Series on Testing and Assessment n. 255. Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be used within Integrated Approaches to Testing and Assessment (IATA). Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/testguidelines.

Pape, W.J.W., Pfannenbecker, U. & Hoppe, U. (1987). Validation of the red blood cell test system as *in vitro* assay for the rapid screening of irritation potential of surfactants. Molecular Toxicology 1, 525-536.

Pape W.J.W., Hoppe U. (1990). Standardization of an *in vitro* red blood cell test for evaluating the acute cytotoxic potential of tensides. Arzneimittelforschung 40, 498-502.

Pape W.J.W., Pfannenbecker U., Argembeaux H., Bracher M., Esdaile D.J., Hagino S., Kasai Y., Lewis R.W, (1999). COLIPA validation project on in vitro eye irritation tests for cosmetic ingredients and finished products (phase I): the red blood cell test for the estimation of acute eye irritation potentials. Present status. Toxicology *In Vitro* 13, 343–354.

Patlewicz G., Rodford R., Walker J.D. (2003). Quantitative structure-activity relationships for predicting skin and eye irritation. Environ. Toxicol. Chem. 22, 1862-1869.

Pfister R.R. (1983). Chemical injuries of the eye. Ophtamology 90, 1246-1253.

Piehl M., Gilotti A., Donovan A., DeGeorge G., Cerven D. (2010). Novel cultured porcine corneal irritancy assay with reversibility endpoint. Toxicology *In Vitro* 24, 231-239.

Piehl M., Carathers M., Soda R., Cerven D., DeGeorge G. (2011). Porcine Corneal Ocular Reversibility Assay (PorCORA) predicts ocular damage and recovery for global regulatory agency hazard categories. Toxicology *In Vitro* 25, 1912-1918.

Prinsen M.K. (2006). The Draize eye test and *in vitro* alternatives, a left-handed marriage? Toxicology *In Vitro* 20, 78-81.

Prinsen M.K., Schipper M.E.I., Wijnands M.V.W. (2011). Histopathology in the Isolated Chicken Eye Test and comparison of different stainings of the cornea. Toxicology *In Vitro* 25, 1475–1479.

Roggeband R., York M., Pericoi M., Braun W. (2000). Eye irritation responses in rabbit and man after single applications of equal volumes of undiluted model liquid detergent products. Food and Chemical Toxicology 38, 727-734.

Rorije E., Aldenberg T., Buist H., Kroese D., Schüürmann G. (2013). The OSIRIS weight of evidence approach: ITS for skin sensitization. *Regul. Toxicol. Pharmacol.* 67, 146-156.

Saito K., Miyazawa M., Nukada Y., Ei K., Abo T., Sakaguchi H. (2015). Predictive performance of the Short Time Exposure test for identifying eye irritation potential of chemical mixtures. Toxicology *In Vitro* 29, 617-20.

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.

Schaeffer M., Barroso J., Worth A., Zuang V. (2014). Chapter 4. Serious eye damage and eye irritation. In: Alternative methods for regulatory toxicology – a state-of-the-art review. EURL-ECVAM Report EUR 26797 EN. Luxembourg: Publications Office of the European Union. Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC91361/echa_jrc_sla_report_public_05-09-14_withcover%20ipo.pdf.

Schneider K., Schwarz M., Burkholder I., Kopp-Schneider A., Edler L., Kinsner-Ovaskainen A., Hartung T., Hoffmann S. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol. Lett. 189, 138-144.

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.

Settivari R.S., Amado R.A., Corvaro M., Visconti N.R., Kan L., Carney E.W., Boverhof D.R., Gehen S.C. (2016). Tiered application of the neutral red release and EpiOcular™ assays for evaluating the eye irritation potential of agrochemical formulations. Regul. Toxicol. Pharmacol. 81, 407-420.

Solimeo R., Zhang J., Kim M., Sedykh A., Zhu H. (2012). Predicting Chemical Ocular Toxicity Using a Combinatorial QSAR Approach. Chem. Res. Toxicol. 25, 2763-2769.

Spielmann H., Gerner I., Kalweit S., Moog R., Wirnsberger T., Krauser K., Kreiling R., Kreuzer H., Luepke N.P., Miltenburger H.G., Mueller N., Muermann P., Pape W., Siegemund B., Spengler J., Steiling W., Wiebel, F.J. (1991). Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. Toxicology *In Vitro* 5, 539-542.

Steiling W., Bracher M., Courtellemont P., Silva O. de (1999). The HET-CAM, a useful *in vitro* assay for assessing the eye irritation properties of cosmetic formulations and indigrents. Toxicology *In Vitro* 13, 375-384.

Stürer A., Seidel C., Sauer O., Koch I., Zilker T., Hermanns-Clausen M., Hruby K., Hüller G., Tutdibi E., Heppner H.J., Desel H. (2010). Poisons Centres' Data for Expert Judgement within Classification, Labelling and Packaging Regulation: Solid Household Automatic Dishwashing Products do not Cause Serious Eye Damage. Clinical Toxicology 48, 245.

Takahashi Y., Koike M., Honda H., Ito Y., Sakaguchi H., Suzuki H., 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.

Tandon R., Bartok M., Zorn-Kruppa M., Brandner J.M., Gabel D., Engelke M. (2015). Assessment of the eye irritation potential of chemicals: A comparison study between two test methods based on human 3D hemi-cornea models. Toxicology *In Vitro* 30, 561-568.

Tsakovska I., Netzeva T., Worth A.P. (2005) Evaluation of (Q)SARs for the prediction of Eye Irritation/Corrosion Potential - physico-chemical exclusion rules. JRC Report EUR 21897 EN, 42pp. European Chemicals Bureau, Ispra, Italy. Available at: https://eurl-ecvam.jrc.ec.europa.eu/ laboratories-research/predictive_toxicology/doc/Evaluation_of_Eye_Irritation_QSARs.pdf

Tsakovska I., Gallegos Saliner A., Netzeva T., Pavan M., Worth A.P. (2007) Evaluation of SARs for the prediction of eye irritation/corrosion potential - structural inclusion rules in the BfR decision support system. SAR QSAR Environ Res 18, 221-235.

UN (2015). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30/Rev.6, Sixth Revised Edition, New York and Geneva: United Nations. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev06/06files_e.html.

US EPA (2015). Use of an alternate testing framework for classification of eye irritation potential of EPA pesticide products. Office of Pesticide Programs, US Environmental Protection Agency, Washington DC, 20460, 3-2-2015. Available at: http://www.epa.gov/sites/production/files/2015-05/documents/eye_policy2015update.pdf.

U.S. CPSC. 2015b. Test for Eye Irritants. 16 C.F.R. §1500.42. Available at: https://www.gpo.gov/fdsys/pkg/CFR-2016-title16-vol2/pdf/CFR-2016-title16-vol2-sec1500-42.pdf

Verma R.P., Matthews E.J. (2015a). Estimation of the chemical-induced eye injury using a weight-of-evidence (WoE) battery of 21 artificial neural network (ANN) c-QSAR models (QSAR-21): part I: irritation potential. Regul. Toxicol. Pharmacol. 71, 318-330.

Verma R.P., Matthews E.J. (2015b). Estimation of the chemical-induced eye injury using a Weight-of-Evidence (WoE) battery of 21 artificial neural network (ANN) c-QSAR models (QSAR-21): part II: corrosion potential. Regul. Toxicol. Pharmacol. 71, 331-336.

Verma R.P., Matthews E.J. (2015c). An *in silico* expert system for the identification of eye irritants. SAR QSAR Environ. Res. 26, 383-395.

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 Pharmacology 19, 276-360.

WHO (2000). Working Group report on Evaluation and use of epidemiological evidence for environmental health risk assessment: WHO Guideline Document. Environm. Health Perspect. 108, 997-1002.

Wilhelmus K.R. (2001). The Draize eye test. Survey of Ophthalmology 45, 493-515.

Yamaguchi H., Kojima H., Takezawa T. (2016). Predictive performance of the Vitrigel-eye irritancy test method using 118 chemicals. J. Appl. Toxicol. 36, 1025-1037.

Young J.R., How M.J., Walker A.P., Worth W.M.H. (1988). Classification as corrosive or irritant to skin of preparations containing acidic or alkaline substances, without testing on animals. Toxicology *In Vitro* 2, 19-26.

Zhu H., Bouhifd M., Kleinstreuer N., Kroese E.D., Liu Z., Luechtefeld T., Pamies D., Shen J., Strauss V., Wu S., Hartung T. (2016). Supporting read-across using biological data. ALTEX 33,167-182.

Zorn-Kruppa M., Houdek P., Wladykowski E., Engelke M., Bartok M., Mewes K.R., Moll I., Brandner J.M. (2014). Determining the Depth of Injury in Bioengineered Tissue Models of Cornea and Conjunctiva for the Prediction of All Three Ocular GHS Categories. PLoS One 9, e114181.

ANNEX I – EXAMPLE OF MATRIX FOR WEIGHT OF EVIDENCE ANALYSES

For those modules having available data, entries are filled in the respective cases. For the rest of the entries, NA shall be indicated in column 2. It is recommended to use short and conclusive wording. For assessment of the evidence, refer to the Part 2 of this guidance document. Note that WoE should be assessed before any new experimental data is generated.

Module	**Title of document / full reference; or data not available (NA)**	**Study result and/or positive or negative evidence obtained**	**Data quality,** according to the Klimisch score, when appropriate *	**Adequacy and relevance,** short statement	**Coverage of relevant para-meters and observations,** Yes/No	**Consistency with other information****	**Conclusive remark*****
1. Existing human data							
2. Existing data on skin corrosion							
3.In vivo animal study							
4. *In vitro* data from OECD adopted test methods							
5. Other animal data from non OECD adopted test methods							
6. Other data from non-OECD adopted alternative test methods							
7.Physico-chemical properties							
8.Non-testing methods ((Q)SAR, grouping, bridging & additivity approaches)							
Overall conclusion	1. WoE allows decision/assessment of the potential of the test chemical to induce serious eye damage and eye irritation. The substance should be classified as UN GHS No Cat., Cat. 2 (2A or 2B), Cat. 1, **or**						

	2. WoE does not allow decision/assessment of the potential of the test chemical to induce serious eye damage and eye irritation. Recommendation or specification of the most appropriate additional testing strategy to be undertaken.

) An electronic tool supporting the quality assessment of in vivo and vitro data through the application of consistent criteria leading to scored results has been developed by EURL ECVAM (described in Schneider et al., 2009). The ToxRTool can be downloaded from the EURL ECVAM page:* https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool*) For example: "This data (any entry except 3 and 4) is consistent with the existing in vitro studies".*

****) For example: "The existing human data suggest that the substance is an eye irritant. Due to poor reporting of this data, and low quality in terms of exposure information, the data is inconclusive, and has a low weight in the final evaluation.* "

ANNEX II – EXAMPLES OF WEIGHT OF EVIDENCE EVALUATIONS

***Disclaimer**: the examples presented below do not imply acceptance or endorsement by any Member Country or OECD. They are intended only to provide an illustration on how individual information sources may be reported and combined in a WoE approach to derive a final classification.*

Example 1: Weight of evidence analyses for classification of a Soluble Liquid (SL) Agrochemical formulation, DD-001, for effects on eyes

	Full Reference	Study result	Data quality Klimisch score	Adequacy and relevance	Coverage of relevant parameters/ observations Yes/No	Consistency	Conclusive remark
1. Existing human data	Not available	-	-	-	-	-	-
2.*In vivo* animal study	Non available	-	-	-	-	-	-
3. *In vitro* data from OECD adopted test methods	Arthur, 2015. *In vitro* Eye irritation test (EpiOcular method). OECD, 492 compliance.	The relative tissue viability (mean) in the EpiOcular™ was 67 %. The mixture does not require classification for effect on eyes according to the prediction model.	1	Key study conducted according to GLP. OECD 492 allows discrimination between materials not requiring classification from those requiring classification (Cat. 2/ Cat. 1).	Yes	Consistent with existing *in vitro* studies.	Key data. Data supports that the mixture does not require classification as eye irritant. Proposed classification: GHS not classified
4. Other animal data from non OECD	Non available	-	-	-	-	-	-

adopted test methods							
5. Other data from non-OECD adopted alternative test methods	Arthur, 2015. *In vitro* Eye irritation test (Neutral red release method).	The EC_{50} in the NRR assay was 630. The mixture did not show eye irritation potentially both according to original interpretation criteria (Reader, 1989) and proposed revised criteria (Settivari, 2016).	2	Internal screening, non-GLP compliant but performed in a GLP facility in the spirit of GLP. Supportive information, limitation due to lack of predictivity for GHS cat 2 agrochemicals formulations.	Yes	Consistent with existing *in vitro* studies (low cytotoxicity).	Supportive data. Data supports that the mixture does not require classification as eye irritant. Proposed classification: GHS not classified
6. Existing data on skin corrosion	Smith, 2011. Acute skin irritation study in the White Zealand Rabbit.	Not corrosive nor irritant to the skin. Mean scores (at 24, 48, 72 hours): - Erythema: 0.7, 1.0, 1.3 - Oedema: 0.0, 0.3, 0.0 Recovery by day 7	1	GLP compliant. Study confirms low skin irritation potential.	Yes	Consistent with other *in vitro* evidence.	Supportive data. Effects on skin except for skin corrosion do not allow assessment for effects on eyes. Data supports that the mixture does not require classification as severe eye irritant.
7.Physico-chemical properties	Acosta, 2001. Determination of pH, acidity and alkalinity measurement according to OECD 122.	pH is 5.2 is therefore not pH-extreme	1	Supportive information because pH alone does not allow assessment of the eye irritancy.	Yes	-	Supportive data. Data supports that the mixture does not require classification as severe eye irritant.
8.Non-testing methods	Chatfield, 2014. Additivity	GHS or CLP classification for all ingredients (2 active substances and 4 co-formulants)	Not applicable as Klimisch	Supportive information. Usable for Classification	Yes	Consistent with existing *in vitro*	Key data. Data supports that the mixture does not require

((Q)SAR, grouping, bridging & additivity approaches)	approach, requested in the European assessment report, Part C, confidential information.	is available from the corresponding MSDS. There are no GHS cat 1 classified ingredients and 2 ingredients (surfactants) classified as GHS cat 2A, accounting for a total of 6.72% w/w of the mixture composition. No classification for eye irritation is triggered according to GHS criteria (UN, 2015).	score is applicable to assessing the reliability of toxicological studies	purposes in EU.		studies.	classification as eye irritant. Proposed classification: GHS not classified
Overall conclusion	**No human data are available.** **pH and skin effects do not lead to a direct UN GHS Cat. 1 classification.** ***In vitro* data on two independent cytotoxicity based assays indicate that the test item has low cytotoxic potential and classification is not required.** **Non testing data (additivity approach based on concentration thresholds), support that classification is not required.** **In conclusion, a WoE evaluation of the consistency, quality and relevance of all available data allows a decision on the eye irritation/serous eye damage potential of the Agrochemical formulation, DD-001. DD-001 should not be classified for eye hazards.**						

Note: This example has been developed only to illustrate how the classification of an untested mixture could be derived and justified. It does not contain any recommendation for a testing strategy.

Example 2: Weight of evidence analyses for classification of a Hand Dish Washing Liquid W07 for effects on eyes

	Full Reference	**Study result**	**Data quality Klimisch score**	**Adequacy and relevance**	**Coverage of relevant parameters/ observations Yes/No**	**Consistency**	**Conclusive remark**
Existing human data on company-owned mixture W07	Poison Control Centre data collected over a 12 months period	9 cases of mild to moderate eye effects only were reported out of all sold products*. In the cases where follow-up information was available, all ocular effects were fully reversible within a few days. **This is an example, in reality the number of cases will need to be identified relative to the number of products sold in a specific geographical area.*	Not applicable to Poison Control Centre data as Klimisch score is applicable to assessing the reliability of toxicological studies.	Supportive information. Limitation due to unknown dose and exposure duration. No criteria for C&L based on human data.	No, not in every case all relevant parameters are covered (e.g. exposure conditions, detailed tissue effects).	Consistent with existing *in vitro* studies and other human experience, which identify the hand dish washing liquid W07 as inducing fully reversible ocular effects.	Supportive data.

	Full Reference	**Study result**	**Data quality Klimisch score**	**Adequacy and relevance**	**Coverage of relevant parameters/ observations Yes/No**	**Consistency**	**Conclusive remark**
Existing human data on similar mixtures	MAGAM II Multicentre multi-national prospective, study of human eye exposures reported to poisons centres, over a 24 months period	28 reported cases related to hand dish washing liquids: mild to moderate but no severe eye irritation after exposure. In the cases where follow-up information was available, all ocular effects were fully reversible within a few days.	Not applicable to Poison Control Centre data as Klimisch score is applicable to assessing the reliability of toxicological studies.	Supportive information. Scoring based on Poison Control Centre severity scoring system complemented by MAGAM reported symptoms. No criteria for C&L based on human data. Information provided as a product category containing different products *vs.* an individual named product.	Although not in every case all relevant parameters are available (e.g. exposure conditions), tissue observations are conducted typically by an ophthalmologist and reported in a standardized way.	Consistent with existing *in vivo* and *in vitro* studies, which identify the hand dish washing liquid W07 as inducing fully reversible ocular effects.	Supportive data.

	Full Reference	**Study result**	**Data quality Klimisch score**	**Adequacy and relevance**	**Coverage of relevant parameters/ observations Yes/No**	**Consistency**	**Conclusive remark**
In vitro data on eye irritation corrosion	Isolated Chicken Eye Test OECD 438 with histopathology as an additional endpoint, 2015	No Prediction can be Made based on a combination of the endpoint categories of II;II;III. This combination of endpoint categories is much lower than those used to identify classification as Cat. 1. Not identified as UN GHS Cat. 1 based on criteria developed by Cazelle *et al*. (2014) for histopathological evaluation of non-pH-extreme detergents and cleaning products.	1	Key and supportive study conducted according to GLP.	Yes	Consistent with existing *in vitro* studies and human experience data which ***does not*** identify the hand dish washing liquid W07 as a UN GHS Cat. 1.	Key and supportive data. OECD 438 study with histopathology as an additional endpoint.
In vitro data on eye irritation corrosion	Reconstructed human Cornea-like Epithelium (RhCE) Test Method OECD 492, 2016	Tissue viability in the EpiOcular™ EIT was 45 %, identifying that the mixture requires classification for effect on eyes	1	Key study conducted according to GLP. Study allows judgement on need or no need for classification. OECD 492 allows discrimination between materials not requiring classification from those requiring classification (Cat. 2/ Cat. 1).	Yes	Consistent with existing *in vitro* studies and human experience data	Key data.

	Full Reference	Study result	Data quality Klimisch score	Adequacy and relevance	Coverage of relevant parameters/ observations Yes/No	Consistency	Conclusive remark
In vitro data on eye irritation corrosion	Bovine Corneal Opacity and Permeability Test OECD 437, 2015	No Prediction can be Made based on *In Vitro* Irritancy Score (IVIS) of 10.3. The IVIS is far below the threshold of 55.1 for classification as Cat. 1	1	Key study conducted according to GLP. Study allows judgement on severity of effects but not persistence of effects and it does not allow identification of Cat. 2 specifically.	Yes	Consistent with existing *in vitro* studies and human experience data	Key data.
In vitro data on skin irritation	*In Vitro* Skin Irritation: Reconstructed Human Epidermis (RHE) Test Method OECD 439, 2014	Tissue viability in EpiSkin™ test method was 75 %, identifying that the tested mixture does not require classification for skin irritation.	1	Study confirms low skin irritation potential.	Yes	Consistent with *existing in vitro* studies and human experience data	Supportive data. Effects on skin except for skin corrosion do not allow assessment for effects on eyes.
Physico-chemical properties	Determination of pH, acidity and alkalinity measurement according to OECD 122	pH is 6.0, W07 is therefore not pH-extreme	1	Supportive information because pH alone does not allow assessment of the eye irritancy.	Yes		Supportive data.
Overall conclusion	**Human data indicates only mild to moderate and fully reversible effects.** ***In vitro* data indicates classification required but mixture not identified as UN GHS Cat. 1.** **pH and skin effects do not indicate corrosive effects.** **In conclusion, a WoE evaluation of the consistency, quality and relevance of all available data allows a decision on the eye**						

	Full Reference	Study result	Data quality Klimisch score	Adequacy and relevance	Coverage of relevant parameters/ observations Yes/No	Consistency	Conclusive remark
	irritation/serious eye damage potential of the Hand Dish Washing Liquid W07. The Hand Dish Washing Liquid W07 should be classified as UN GHS Cat. 2.						

Note: This example has been developed only to illustrate how the classification of an untested mixture could be derived and justified. It does not contain any recommendation for a testing strategy. However, the BCOP has very recently been included in a testing strategy for antimicrobial cleaning products (AMCPs) under the U.S. EPA classification and labelling system (Clippinger et al., 2016).