European Commission

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

2016

EUR 28174 EN

Contact information
Name: João BARROSO
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
E-mail: Joao.BARROSO@ec.europa.eu
Tel.: +39 0332 78 5329

JRC Science Hub
https://ec.europa.eu/jrc

JRC103703

EUR 28174 EN

PDF ISBN 978-92-79-63196-2 ISSN 1831-9424 doi:10.2787/991443

Luxembourg: Publications Office of the European Union, 2016

How to cite: EURL ECVAM Scientific Advisory Committee (2016). 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.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC OPINION

on the

Ocular Irritection® Test Method for Prediction of Serious Eye Damage/Eye Irritation Potential of Chemicals

ESAC Opinion No.	2016-01
Relevant ESAC Request No.	2016-01
Date of Opinion	24/06/2016

Table of contents

Abstract

ESAC, the EURL ECVAM Scientific Advisory Committee, advises EURL ECVAM on scientific issues. Its main role is to conduct independent peer review of validation studies of alternative test methods and to assess their scientific validity for a given purpose. The committee reviews the appropriateness of study design and management, the quality of results obtained and the plausibility of the conclusions drawn. ESAC peer reviews are formally initiated with a EURL ECVAM Request for ESAC Advice, which provides the necessary background for the peer-review and establishes its objectives, timelines and the questions to be addressed. The peer review is normally prepared by specialised ESAC Working Groups. These are typically composed of ESAC members and other external experts relevant to the test method under review. These experts may be nominated by ESAC, EURL ECVAM and partner organisations within the International Cooperation on Alternative Test Methods (ICATM). ESAC ultimately decides on the composition of these Working Groups. ESAC's advice to EURL ECVAM is formally provided as 'ESAC Opinions' and 'Working Group Reports' at the end of the peer review. ESAC may also issue Opinions on other scientific issues of relevance to the work and mission of EURL ECVAM but not directly related to a specific alternative test method.

The ESAC Opinion expressed in this report relates to the peer-review of the Ocular Irritection® test method for prediction of serious eye damage/eye irritation potential of chemicals.

Ispra, 24 June 2016

ESAC Opinion

In April 2016, the EURL ECVAM Scientific Advisory Committee (ESAC) (Annex 1) received from EURL ECVAM a request for scientific advice on the external prospective and retrospective validation of the Ocular Irritection® (OI) test method for serious eye damage/eye irritation testing (Annex 2). ESAC established a working group (WG) (Annex 1) which delivered an ESAC WG report dated 6 June 2016 (Annex 3).

The ESAC WG was established to conduct a peer review of, and provide scientific advice on a multi-laboratory trial involving three laboratories supplemented by additional retrospective data and analysis of the OI assay, a test method claimed to have a wide applicability domain. In particular the ESAC WG was asked to consider the relevance (biological/mechanistic relevance and predictive capacity; the latter in the context of an Integrated Approach to Testing and Assessment (IATA) (OECD, 2008)), and the reliability (transferability; within and between laboratory reproducibility) of the test method.

In addition the ESAC WG was asked to comment on draft Performance Standards (PS).

The analysis and conclusions of the ESAC WG were based primarily on the EURL ECVAM Test Submission Template (TST) and supporting documents supplied by SeCAM on behalf of the test manufacturer, and supplementary information made available by SeCAM during and after an 11 May 2016 teleconference.

Details of the validation study were previously published by Eskes *et al*. (2014).

The prospective and retrospective OI assay validation study pre-defined objectives were:

- To formally evaluate the usefulness of the OI assay to reliably discriminate chemicals not requiring classification for serious eye damage/eye irritancy (No Category) from chemicals requiring classification and labelling (Category 1 and Category 2) according to the UN GHS Classification and Labelling of Chemicals (UN GHS. UN, 2015) and as implemented by the EU CLP regulation (EU CLP. EC, 2008a). This test method is not intended to differentiate between GHS Categories 1 (irreversible effects) and 2A-B (reversible effects).
    - N.B. The test method was not designed, intended, or evaluated to test gases or aerosols.
- A further post-hoc evaluation was performed to evaluate the usefulness of the OI assay to discriminate chemicals inducing serious eye damage (UN GHS Category 1) from other classes.
- In addition, to produce evidence and analysis to support the test method being incorporated into a tiered testing strategy (so-called Bottom-Up/Top-Down testing strategy, Scott L. *et al.*, 2010). The ultimate purpose of such a tiered testing strategy being to replace the traditional *in vivo* Draize eye test [Method B.5 of EC Regulation 440/2008 (EC, 2008b) or OECD TG 405 (OECD, 2002)].

At its 42nd meeting, held on the 9th and 10th June 2016 at EURL ECVAM, Ispra, Italy, the non-Commission members of ESAC unanimously endorsed the following statement which was based on the ESAC WG report:

The OI assay evolved from the Eytex® test method (Kelly, 1989; Gordon, 1992) following recommendations made by Balls *et al*. (1995). In 1996, the system underwent substantial revisions to take into account the recommendations made after the earlier multi-laboratory trials including the development of a single protocol, clear procedures for surfactant testing, and a well-defined applicability domain.

The test method is premised on the assumption that eye irritants produce physicochemical changes in the corneal macromolecular matrix, and that these changes can be mimicked in a macromolecular matrix of plant origin as measured by changes in the turbidity of the assay matrix caused by alterations in protein conformation and structural organisation of the matrix. The degree of turbidity is proportional to the ocular irritation potential of the test material – as judged against positive and negative controls, and calibration materials. The change in matrix turbidity is measured by the light scattering detected by a spectrometer set to a wavelength of 405 nm and is converted to a numerical Irritection Draize Equivalent (IDE). The highest estimated IDE score, termed the Maximum Qualified Score (MQS), is then used in the prediction model to categorise the ocular hazard potential of test chemicals according to the UN GHS (UN, 2015) and EU CLP classification systems (EC, 2008a).

From the information available it is not clear why the changes induced in the test method matrix by chemicals should be specific or restricted to ocular irritants, rather than being also applicable to other classes of irritants (e.g. skin irritants). The supporting documentation claims a mechanistic basis based on the resemblance of the assay macromolecular matrix with the human cornea. However, in addition to the assay matrix never having been fully chemically characterised and not being of mammalian origin, there is no detailed description of the nature of the healthy human corneal matrix, or how it changes after exposure to ocular irritants. The documentation does not explain why or how the raw material was originally selected to produce the test kit macromolecular matrix as a surrogate for the human cornea. It also does not explain how the physical chemical alterations occurring in the kit matrix upon exposure to irritant chemicals compare to effects observed with the *in vivo* Draize eye test. ESAC therefore considers that the OI assay's mechanistic relevance to predict adverse ocular effects of chemicals in humans is poorly defined.

Neither the raw materials used to produce the OI assay, nor the test kit matrix itself (before or after the SOP filtration step) have been chemically defined or specified. Furthermore, no information is available about the batch-to-batch chemical consistency of the raw material, as to date the manufacturer has only produced test kits from one batch of the plant extract.

ESAC therefore believes that, at this stage, the assay must be considered more "correlative" than mechanistically relevant.

The OI assay has a reasonably well established and defined applicability domain. It is not applicable to very acidic (pH < 4.0) and very alkaline (pH > 9.0) materials, oils and water-insoluble organic chemicals; and non-ionic surfactants can cause assay interference. The test method also has limitations for the testing of intensely coloured materials generating high OD readings for blanks and samples. In addition, volatile ketones have been found to result in under-estimation of irritancy due probably to evaporation. Finally, a number of false negatives (urea at concentrations > 5 %) and false positives (sorbitol at concentrations > 5 %, manganese violet, aluminium chlorohydrate, aluminium zirconium chlorohydrate, aluminium chloride, titanium oxide, zinc oxide, silver salts, ferrous sulphate, zinc sulphate) have been identified.

The Prediction Model (PM) of the OI assay is based on that initially developed for the Eytex assay to predict *in vivo* Maximum Average Scores (MAS). In this validation study the previously existing MQS cut-off of 12.5 was used to distinguish classified from non-

classified chemicals (i.e. Bottom-Up approach) and the previously existing MQS cut-off of 30.0 was used to distinguish Category 1 from non-Category 1 chemicals (i.e. Top-Down approach).

The OI assay is not intended to provide insights into persistence of chemically-induced ocular injuries, which could result in the underprediction of chemicals classified *in vivo* as Category 1 due only to persistence of effects if the test method is used to identify Category 1 chemicals, e.g., as a first step in a Top-Down approach; gases and aerosols were not evaluated; and the available documentation provides little information on the test method's performance with mixtures. In addition, if this test method is to be used to identify non-irritants within an Integrated Approach to Testing and Assessment (IATA) then ESAC believes other test methods may be needed to assess the vascular and inflammatory components of the adverse outcome pathway for eye irritation, and to take account of the lack of epithelial barrier function.

In addition to qualified and non-qualified runs, the submitter makes use of a third category, excluded runs, which impacts on the calculation of the test performance when comparisons are made with the reported performance of other test methods.

For the prospective ring trial reliance was placed on a statistical power analysis (sample size calculation) estimating the minimum requirements for the assay's use within a Bottom-Up approach (n=50). To make provision for unforeseeable events, 56 chemicals were used in the ring trial (five of which were subsequently excluded from the analysis of the test method performance). The chemicals selection was nevertheless biased in favour of assay use in a Bottom-Up approach.

The predictive capacity of the OI assay as reported by the test manufacturer also took account of retrospective data from 45 chemicals, eight of which were also used in the prospective ring trial.

The OI assay seems to be easily transferable to another laboratory with only general working expertise required for the lab personnel.

In the analysis of the reliability of the assay (assessed through WLR and BLR), non-qualified and excluded test results were considered as concordant when the three laboratories obtained the same outcome.

ESAC believes that including "concordant" non-qualified and excluded test results in the analysis is not sound. Taking this into account, the WLR and BLR values and their respective Wilson two-sided 95 %-Confidence Intervals (CIs) were recalculated by ESAC including only qualified results (Table 1).

Table 1. Calculation of WLR and BLR NOT considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay. The values in brackets correspond to Wilson two-side 95 %-CIs. IVI, InVitro International laboratory; IISV, Institute for In Vitro Sciences laboratory; RP, Res Pharma.

	Within-laboratory Reproducibility						Between-laboratory Reproducibility	
	Cut-off 12.5			Cut-off 30.0			Cut-off 12.5	Cut-off 30.0
	IVI	IISV	RP	IVI	IISV	RP		
concordant predictions	45	42	41	46	43	43	42	43
discordant predictions	6	8	10	5	7	8	8	7
% concordance	88.2 % (76.6-94.5 %)	84.0 % (71.5-91.7 %)	80.4 % (67.5-89.0 %)	90.2 % (79.0-95.7 %)	86.0 % (73.8-93.1 %)	84.3 % (72.0-91.8 %)	84.0 % (71.5-91.7 %)	86.0 % (73.8-93.1 %)

The OI reproducibility (WLR and BLR) appears to be adequate although lower than that of Reconstructed human Cornea-like Epithelium (RhCE)-based test methods. Nevertheless, since all the macromolecular matrices used in the validation study were produced from the same bulk plant-extract raw material (see below), it remains unknown how the variability of the assay may be affected by the use of different batches and suppliers of this raw material.

Predictive capacity results are summarised below (Tables 2 and 3). Due to the unbalanced number of repetitions between the prospective and the retrospective datasets available, the majority of predictions available for each chemical (mode) was preferred by the study Validation Management Group (VMG) as the means of expressing the predictive capacity of the assay. ESAC is however of the opinion that the calculations of Predictive Capacity should reflect in the best way possible the real-life testing situation. Since with the OI assay one single test result will be used to derive one final prediction, ESAC considers that the majority of predictions is sub-optimal to express the Predictive Capacity of the test method. In ESAC's opinion, more truthful point estimates can be obtained by using a 'weighted' calculation that considers for each chemical the proportion of correct classification of repeat tests or by resampling of the multiple data generated in the validation study. Both the weighted calculations (with two-sided 90 %-Confidence Intervals (CIs) obtained from resampling) and predictive capacity values obtained from the majority of predictions (with Wilson two-sided 95 %-CIs calculated by ESAC) are given in Tables 2 and 3 for comparison. Published predictive capacity values for other serious eye/damage eye irritation *in vitro* methods are also provided for comparison.

When used for the identification of UN GHS non-classified versus classified materials (based on the existing cut-off of 12.5) the OI assay showed an overall sensitivity of 90.7 % (two-sided 90 %-CIs: 87.0 % and 93.5 %), a specificity of 58.9 % (two-sided 90 %-CIs: 53.5 % and 65.1 %) and an overall accuracy of 75.3 %, based on weighted calculations (Table 2). Some organic functional groups were found possibly to correlate with the observed mispredictions. In particular, acrylate, carboxamide, and cycloalkenes. If, despite their small number, chemicals containing these functional groups were excluded from analyses, the obtained dataset resulted in a sufficiently large dataset (n=79) to still derive sound conclusions. Such findings are comparable to the currently accepted OECD test methods as shown in Table 2.

Table 2. Predictive capacity of the OI for the identification of UN GHS/EU CLP non-classified chemicals and comparison to published values for other *in vitro* methods.

	OI n=88 weighted (resampling 90 %-CI)	**OI* n=79 weighted (resampling 90 %-CI)**	**OI n=88 majority of predictions (Wilson 95 %-CI)**	**OI* n=79 majority of predictions (Wilson 95 %-CI)**	**BCOP (TG437) n=196 majority of predictions**	**ICE (TG 438) n=152 majority of predictions**	**STE** (TG 491) n=101**	**EpiOcular EIT (TG 492) n=112 weighted**	**CM*** (Draft TG) n=45 weighted**
Accuracy	75.3 % n.a.	80.5 % n.a.	76.1 % (67/88) (66.3-83.8 %)	81.0 % (64/79) (71.0-88.1 %)	68.9 % (135/196)	82.2 % (125/152)	90.1 % (91/101)	80 % (n=112)	68 % (n=45)
Sensitivity	90.7 % (87.0-93.5 %)	96.4 % (92.9-100 %)	93.3 % (42/45) (82.1-97.7 %)	97.6 % (40/41) (87.4-99.6 %)	100.0 % (107/107)	98.6 % (72/73)	98.1 % (53/54)	96 % (n=57)	100 % (n=22)
Specificity	58.9% (53.5-65.1 %)	62.8 % (55.3-71.1%)	58.1 % (25/43) (43.3-71.6 %)	63.2 % (24/38) (47.3-76.6 %)	31.5 % (28/89)	67.1 % (53/79)	80.9 % (38/47)	63 % (n=55)	32 % (n=23)

*excludes chemicals containing the *acrylate, carboxamide and cycloalkenes organic* functional groups.

** Only water-soluble chemicals or chemicals forming a uniform suspension, and excluding highly volatile substances and solid substances other than surfactants;

*** Only water-soluble surfactants and surfactant-containing formulations

For the identification of the UN GHS /EU CLP Category 1 chemicals a post-hoc evaluation of the entire dataset comprising both, the prospective validation dataset and the additional existing data from retrospective studies was carried out. When used for the identification of UN GHS Category 1 versus non-Category 1 chemicals (based on the cut-off of 30.0) the OI assay showed an overall specificity of 80.9 % (two-sided 90 %-CIs: 76.8 % and 84.1 %), a sensitivity of 53.3 % (two-sided 90 %-CIs: 50.0 % and 60.0 %), and an accuracy of 74.7 %, based on weighted calculations (Table 3). These values are

compared with those of the currently accepted OECD *in vitro* test methods for eye hazard assessment in Table 3 below.

Further investigations conducted by the VMG to better understand possible reasons for misclassification suggested that chemicals having the presence of the organic functional groups carboxylic acid and sulphate seemed to risk possible underpredictions of Category 1 chemicals as non-Category 1 chemicals. Excluding chemicals having these functional groups from analyses resulted still in a sufficiently large dataset to make sound conclusions (n=74). In this case, a specificity of 80.0 % (48/60), a sensitivity of 71.4 % (10/14) and a concordance of 78.4 % (58/74) were obtained, based on the majority of predictions.

Table 3. Predictive capacity of the OI for the identification of UN GHS / EU CLP Category 1 chemicals and comparison to published values for other *in vitro* methods.

	OI n=88 weighted (resampling 90 %-CI)	OI n=88 majority of predictions (Wilson 95 %-CI)	BCOP (TG 437) n=191 majority of predictions	ICE (TG 438) n=140 majority of predictions	FL (TG 460) n=151 weighted	STE (TG 491) n=120*	CM (Draft TG) n=68* weighted
Accuracy	74.7 % n.a.	73.9 % (65/88) (63.8-81.9 %)	78.5 % (150/191)	85.7 % (120/140)	77.5 % (117/151)	85.0 % (102/120)	88 % (n=68)
Specificity	80.9 % (76.8-84.1 %)	80.9 % (55/68) (70.0-88.5 %)	74.6 % (94/126)	93.8 % (106/113)	93.2 % (96/103)	98.8 % (83/84)	98 % (n=42)
Sensitivity	53.3 % (50.0-60.0 %)	50.0 % (10/20) (29.9-70.1 %)	86.2 % (56/65)	51.9 % (14/27)	43.8 % (21/48)	52.8 % (19/36)	73 % (n=26)

* water-soluble chemicals or chemicals forming a uniform suspension.

Overall the rationale provided for the assay limitations with respect to the applicability of the OI assay to particular types of chemicals is weak since the exclusion of chemicals is based on single or a very limited number of representative chemicals.

When used to identify Category 1 chemicals (e.g., as a first step in a Top-Down Approach), the OI assay shows rather low sensitivity (influenced in part by its limitation to detect persistence of effects as mentioned above) and only moderate specificity. Thus, other than having a long shelf-life and short testing time, the OI assay does not seem to be an advance on other validated test methods for this purpose.

Whilst the performance of the test method for identifying non-classified chemicals appears to be adequate, there are a number of uncertainties related with the method that may call into question its regulatory acceptance, not least a lack of control over the precise chemical composition, stability and structure of the test matrix (to date only produced from one batch of the plant-extract over a period of 25+ years); the unknown variability of the method if performed with matrices produced from different bulks of the plant-extract raw material (to date only matrices produced from a single bulk have been evaluated), and a lack of transparency about the details of the software used in the OI assay to analyse the data. On this basis, ESAC is also not able to offer a reasoned opinion on what might constitute Essential Test Method Components for any assay of this class to be included in Performance Standards: it is not clear how an essential test method component can be defined without specifying the test matrix origin and its chemical composition.

ESAC also notes that the Performance Standards provided for peer review are incomplete as they lack a list of reference chemicals and target values for reproducibility and predictive capacity. The chemicals used in the validation study of the OI assay would be suitable only for similar test methods with the same precise applicability domain.

On the basis of the above, ESAC does not recommend the use of the OI assay for regulatory testing purposes. Nevertheless, considering that (i) the assay is relatively easy and fast to perform, (ii) its performance appears to be acceptable, (iii) it is easily shipped and stored, and (iv) it has a long shelf-life, ESAC considers that the assay may be useful for screening purposes within the applicability domain established in the SOP. Users of the OI assay should nevertheless take into consideration that the applicability domain of the method has been imperfectly defined, being purely empirical and not biologically justified.

References

- Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, *et al.* (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of the ECVAM Workshop 5*. ATLA* 23**:** 129-147.
- EC (2008a). Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures (the 'CLP Regulation').
- EC (2008b). COUNCIL REGULATION (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)
- 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. *Toxicol in Vitro* 28: 1046-1065.
- Gordon VC (1992). The scientific basis of the EYTEX system. *ATLA* 20: 537–548.
- Kelly CP (1989). EYTEX: an in vitro method of predicting ocular safety. *Pharmacopeial Forum* 15: 4815–5824.
- OECD (2002). Acute Eye Irritation/Corrosion. OECD Guideline for the Testing of Chemicals No. 405, OECD, Paris.
- OECD (2008). OECD Environment Health and Safety Publications Series on Testing and Assessment No. 88 Workshop on Integrated Approaches to Testing and Assessment. OECD, Paris.
- Scott L, Eskes C, Hoffmann S, Adriaens E, Alépée N, Bufo M, et al. (2010). A proposed eye irritation testing strategy to reduce and replace *in vivo* studies using Bottom-Up and Top-Down approaches. *Toxicol In Vitro* 24: 1-9.
- UN (2015). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Sixth revised edition, UN New York and Geneva.

Annex 1

COMPOSITION OF ESAC AND ESAC WORKING GROUP

Composition of ESAC and ESAC Working Group

EURL ECVAM Scientific Advisory Committee (ESAC)

- Dr. Neil CARMICHAEL (ESAC Chair)
- Prof. Jürgen BORLAK
- Dr. Harvey CLEWELL
- Prof. Lucio G. COSTA
- Dr. Kristina KEJLOVÁ
- Prof. David John KIRKLAND
- Prof. Annette KOPP-SCHNEIDER
- Dr. Renate KRÄTKE
- Prof. Claus-Michael LEHR
- Dr. José Maria NAVAS
- Prof. Aldert PIERSMA
- Dr. Jonathan RICHMOND
- Dr. Erwin L. ROGGEN
- Dr. Dorothea SESARDIC

ESAC Working Group (WG)

- Dr. José Maria NAVAS (ESAC Member, WG Chair)
- Dr. Kristina KEJLOVÁ (ESAC Member)
- Prof. Annette KOPP-SCHNEIDER (ESAC Member)
- Dr. Renate KRÄTKE (ESAC Member)
- Dr. Jon RICHMOND (ESAC Member)
- Dr. Dave ALLEN (NICEATM; ICATM nomination by NICEATM/ICCVAM)
- Prof. Kyung-Min LIM (College of Pharmacy, Ewha Womans University; ICATM nomination by KoCVAM)

EURL ECVAM (Secretariat)

- Dr. João BARROSO (ESAC Coordinator)
- Prof. Maurice WHELAN (Head of Unit)

Annex 2

EURL ECVAM REQUEST FOR ESAC ADVICE

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

ESAC Request 2016-01

EURL ECVAM Scientific Advisory Committee (ESAC)

EURL ECVAM REQUEST FOR ESAC ADVICE

on the

Ocular Irritection® test method for prediction of serious eye damage/eye irritation potential of chemicals

Title page information	
Abbreviated title of ESAC request	Ocular Irritection® validation
ESAC REQUEST No.	2016-01
Template used for preparing request	EP 3.02
Date of finalising request	27/05/2016
Date of submitting request to ESAC	27/05/2016
Request discussed through	Written procedure previous to ESAC 42
Opinion expected at (date)	ESAC 42 (June 2016)
File name of this request	**ER2016-01_ESAC_REQUEST_OCULAR_IRRITECTION.doc**

TABLE OF CONTENTS

1. TYPE OF REQUEST

Request Type	Identify request ("YES")
R1 ESAC Peer Review **of a Prevalidation Study or Validation Study**	**YES, external validation study** **(i.e. not coordinated by EURL ECVAM)**
If R1)applies please specify further:	
►Prevalidation Study	**NO**
►Prospective Validation Study	**YES (the submission includes a prospective validation study with 56 test chemicals as well as retrospective data on an additional 45 test chemicals, for a total of 93 unique chemicals)** Background The Ocular Irritection® is an *in vitro* macromolecular test method, representing a refinement of the former Eytex® method (Kelly, 1989; Gordon, 1992) following recommendations made by Balls et al. (1995). It intends to predict the ocular hazard effects of chemicals based on the premise that corneal opacity may result from the disruptive effects ocular irritants may have on the highly organised structure of the cornea through interaction with some of its components (e.g. proteins and carbohydrates). This assay thus mimics the biochemical phenomena of corneal protein denaturation and disruption caused by irritant chemicals acting on the cornea. The test method uses as test system a macromolecular matrix composed of a mixture of plant proteins, plant glycoproteins, plant carbohydrates, plant lipids and low molecular weight plant components, which mimics the highly ordered structure of the transparent cornea. The Ocular Irritection® underwent an external prospective and retrospective validation study to assess its usefulness and limitations to identify chemicals not requiring classification for serious eye damage/eye irritation ('No Category') and chemicals inducing serious eye damage ('Category 1'), according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) (UN, 2015) and the European Union Regulation on Classification, Labelling and Packaging of chemicals (EU CLP) (EC, 2008), in the framework of a Bottom-Up/Top-Down testing strategy (Scott et al., 2010). In December 2013, after completion of the validation study, EURL ECVAM received a full submission on the Ocular Irritection® test method. EURL ECVAM evaluated the submission and concluded that the test method appears to be promising as a partial replacement to identify 'Category 1' and 'No Category' chemicals. However, EURL ECVAM noted a number of shortcomings in the method definition and in the information provided concerning its performance (reproducibility and predictive capacity) and, in August 2014, the test submitter was requested to

	address and clarify these issues in a revised submission. Upon request of the test submitter, a meeting between EURL ECVAM and the Validation Management Group (VMG) of the Ocular Irritection® validation study was organised in October 2014 to discuss the various points raised by EURL ECVAM in its assessment report. Following this meeting, the submitter updated its submission with additional (non-testing) information and further biostatistical analyses as requested by EURL ECVAM, and provided a full revised submission in April 2015. EURL ECVAM evaluated this revised submission, and the updated assessment report is made available to the ESAC in the accompanying documents of the current request.
►Retrospective Validation Study	**YES (see above)**
►Validation Study based on Performance Standards	**NO**

R2 Scientific Advice on a test method submitted to EURL ECVAM for validation **(e.g. the test method's biological relevance etc.)**	**NO**
R3 Other Scientific Advice **(e.g. on test methods, their use; on technical issues such as cell culturing, stem cells, definition of performance standards etc.)**	**NO**

2. TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED

External prospective and retrospective validation of the in vitro Ocular Irritection® (OI) test method for prediction of serious eye damage/eye irritation potential of chemicals

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

3.1. Background on serious eye damage/eye irritation and current testing strategies

3.1.1. Serious eye damage/eye irritation

Serious eye damage/eye irritation is an adverse effect that produces changes in the eye following exposure of the anterior surface of the eye to a test substance. According to UN GHS (UN, 2015), *serious eye damage* is the production of tissue damage in the eye, or serious physical decay of vision, following application of a test substance to the anterior surface of the eye, which is not fully reversible within 21 days of application. *Eye irritation* is the production of changes in the eye following the application of a test substance to the anterior surface of the eye, which are fully reversible within 21 days of application.

UN GHS includes three main categories for the classification of chemicals: Category 1 (abbr. Cat 1) for "serious eye damage", Category 2 (abbr. Cat 2) for "eye irritation" and No Category (abbr. No Cat) for "not-classified".

Currently, serious eye damage/eye irritation can be determined through *in vivo* and *in vitro* assays. The traditional *in vivo* test is the Draize eye test, which uses rabbits as model for ocular toxicity (OECD Test Guideline (TG) 405) (OECD, 2012a). Validated *in vitro* alternative methods are available based on organotypic assays (Bovine Corneal Opacity and Permeability test, Isolated Chicken Eye test), adopted in 2009 and revised in 2013 (TG 437, TG 438) (OECD, 2013a, b), on cell-based assays (Fluorescein Leakage, Short Time Exposure and Cytosensor Microphysiometer) adopted in 2012 (TG 460) (OECD, 2012b) and in 2015 (TG 491) (OECD, 2015a) or currently undergoing regulatory acceptance (draft TG on the Cytosensor Microphysiometer), and on Reconstructed human Cornea-like Epithelium (EpiOcular™ EIT) adopted in 2015 (TG 492) (OECD, 2015b).

3.1.2. Testing strategies composed of in vitro methods

It is generally accepted that, in the foreseeable future, no single *in vitro* test method will be able to replace the *in vivo* Draize eye test (TG 405) (OECD, 2012a) in its capacity to predict averse ocular effects for the full range of potency and for a broad spectrum of chemical classes (wide applicability domain). However, appropriate combinations of several alternative test methods within (tiered) testing strategies may be able to replace the Draize eye test.

A possible conceptual framework for such (tiered) testing strategies was developed within an EURL ECVAM workshop (Scott *et al*., 2010). The framework is based on alternative serious eye damage/eye irritation methods that vary in their capacity to detect chemicals inducing serious eye damage (Cat 1) and/or chemicals not requiring classification for serious eye damage/eye irritation (No Cat). According to this framework, the entire potency range of effects may be resolved by arranging tests in a tiered (sequential) strategy that may be operated bidirectional, i.e. from either end. As such, the strategy intends to classify chemicals following two possible types of approaches colloquially referred to as 'Top-Down' and 'Bottom-Up' (Scott *et al.*, 2010). In the Top-Down approach, the testing aims to first identify Cat 1 chemicals, discriminating these from the rest, i.e. a combination of Cat 2 and No Cat chemicals. Conversely, in the Bottom-Up approach, the testing aims to first identify No Cat chemicals, discriminating these from all chemicals requiring classification, i.e. Cat 1 and Cat 2 chemicals combined. Ocular irritant chemicals (Cat 2) will be resolved in a last tier in both approaches.

In the international regulatory context, this framework is established within an Integrated Approach to Testing and Assessment (IATA), which is currently under development at the OECD. The IATA includes several modules considering already existing data, physicochemical properties, (Q)SARs and other *in silico* tools and empirical testing tools (including *in vitro* methods and, as a last resource, *in vivo* testing). The IATA integrates the use of the Top-Down and Bottom-Up approaches when generation of new testing data is necessary i.e., if no conclusion can be drawn from existing and *in silico* data. In the European regulatory context, the ECHA *Guidance on Information Requirements and Chemical Safety Assessment* (IR&CSA), Chapter R.7a: Endpoint specific guidance, provides guidance on the integration of testing and non-testing data for the assessment of eye damage/irritation and is regularly updated (ECHA, 2015).

3.2. Background on the Ocular Irritection® test method

The Ocular Irritection® (OI) is a test method that uses as test system a macromolecular matrix of plan origin intending to mimic the highly organised structure of the transparent cornea. The topical application of irritant substances induces changes of the matrix (turbidity) which are measured by optical density. The increase in optical density is used to predict the ocular hazard effects of chemicals based on the premise that corneal opacity observed *in vivo* may result from the disruptive effects ocular irritants may have on the highly organised structure of the cornea through interaction with some of its components (e.g. proteins and carbohydrates).

3.2.1. Study objectives and design

The OI aims to identify both No Cat and Cat 1 chemicals in the framework of a Bottom-Up/Top-Down testing strategy. It is not intended to identify Cat 2 chemicals on its own. The OI underwent an external prospective and retrospective validation study. The goal of the study was initially to assess the relevance (predictive capacity), reliability (transferability and reproducibility within and between laboratories) and limitations/applicability domain of the OI to identify No Cat chemicals. A post-hoc evaluation of the OI to identify Cat 1 chemicals was however also conducted.

The submission of the OI external validation study comprises both retrospective and prospective data. In the prospective part of the validation study 56 chemicals with existing reference *in vivo* Draize eye test data were tested in a ring trial involving three laboratories, of which 4 chemicals provided "excluded" data in all tests in all laboratories. The prospective data were used to evaluate the reproducibility (within- and between-laboratories), predictive capacity (sensitivity, specificity and accuracy) and limitations/applicability domain of the OI for the purposes of identifying No Cat (Bottom-Up) or Cat 1 (Top-Down) chemicals. The assessment of the OI predictive capacity and limitations was also complemented with retrospective *in vitro* data available for 45 unique chemicals that had been generated with the same protocol as the one used in the prospective validation study. Out of these 45 chemicals, 37 were in addition to the prospective validation dataset and 8 were also tested in the prospective validation study.

If found valid for its proposed uses, the OI may be formally incorporated into an OECD Test Guideline (it is currently already included in the OECD work programme under the leadership of Italy). Additionally, the OI may also be incorporated in the draft IATA currently under development at the OECD.

3.2.2. Summary of study results

The study results are presented in detail in the test submission and the EURL ECVAM assessment report. They are also summarised in the following paragraphs.

(a) Within- and between laboratory reproducibility (WLR and BLR)

The submitter and EURL ECVAM had slightly different views on the data that should be used to calculate within- and between laboratory reproducibility (WLR and BLR). The submitter and the VMG deemed that consistent occurrence of "non-qualified"/"excluded" runs could be considered as concordant. In contrast, EURL ECVAM recommended that the submitter discards "non-qualified"/"excluded" runs from the calculations of reproducibility. However, the impact in terms of differences between these two views is limited (<2%) as indicated in the table below provided by the test submitter.

WLR and BLR assessment provided by the submitter

Calculation of WLR (and BLR) when considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay

	Within-laboratory						Between-laboratory	
	Cut-off 12.5 (identification of No Cat)			Cut-off 30.0 (identification of Cat 1)			Cut-off 12.5 (identification of No Cat)	Cut-off 30.0 (identification of Cat 1)
	IVI	IIVS	RP	IVI	IIVS	RP		
concordant predictions	45	42	41	46	43	43	42	43
discordant predictions	6	8	11	5	7	8	10	9
concordant excluded	4	5	4	4	5	4	4	4
concordant NQ	1	1	0	1	1	0	0	0
overall concordant	50	48	45	51	49	46	46	47
% concordance	89.3	85.7	80.4	91.1	87.5	83.9	82.1	83.9

Calculation of WLR and BLR when EURL ECVAM's suggestions are taken up i.e., NOT considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay								
	Within-laboratory						Between-laboratory	
	Cut-off 12.5 (identification of No Cat)			Cut-off 30.0 (identification of Cat 1)			Cut-off 12.5 (identification of No Cat)	Cut-off 30.0 (identification of Cat 1)
	IVI	IIVS	RP	IVI	IIVS	RP		
concordant predictions	45	42	41	46	43	43	42	43
discordant predictions	6	8	11	5	7	8	10	9
% concordance	88.2	84.0	78.8[a]	90.2	86.0	84.3	80.8[b]	82.7[c]

[a] The value calculated by EURL ECVAM corresponds to 80.4% (41/51)
[b] The value calculated by EURL ECVAM corresponds to 84.0% (42/50)
[c] The value calculated by EURL ECVAM corresponds to 86.0% (43/50)

(b) Predictive capacity

According to EURL ECVAM on the basis of weighted calculations, the results are the following:

- Prospective validation data considering chemicals with qualified data in at least one laboratory (n=52)

Prospective validation data: Evaluation of the OI predictive capacity for both top-down and bottom-up approaches using a weighted calculation

Predictive Capacity	TOP-DOWN	BOTTOM-UP
Sensitivity	42.4% (4.67/11)	87.4% (20.11/23)
False Negatives	57.6% (6.33/11)	12.6% (2.89/23)
Specificity	80.2% (32.89/41)	55.6% (16.11/29)
False Positives	19.8% (8.11/41)	44.4% (12.89/29)
Overall Accuracy	72.2% (37.56/52)	69.7% (36.22/52)
Total Mispredictions	27.8% (14.44/52)	30.3% (15.78/52)

- Combined prospective and retrospective validation data considering chemicals with qualified data in at least one laboratory (n=89)

Combined prospective and retrospective validation data: Evaluation of the OI predictive for both top-down and bottom-up approaches using a weighted calculation

Predictive Capacity	TOP-DOWN	BOTTOM-UP
Sensitivity	53.4% (10.67/20)	90.7% (41.71/46)
False Negatives	46.7% (9.33/20)	9.3% (4.29/46)
Specificity	81.0% (55.89/69)	59.8% (25.71/43)
False Positives	19.0% (13.11/69)	40.2% (17.29/43)
Overall Accuracy	74.8% (66.56/89)	75.8% (67.42/89)
Total Mispredictions	25.2% (22.44/89)	24.2% (21.58/89)

Additionally, the submitter followed EURL ECVAM suggestion and provided results after resampling, for both bottom-up and top-down approach, for the combined prospective and retrospective validation data:

Predictive capacity for the bottom-up approach, calculated by the submitter following resampling. The * sign means including chemical #23. The reduced applicability domain excludes acrylate, carboxamide and cycloalkenes functional groups

	Bottom-up (n=88 or 89 chemicals)*				*Bottom-up for a reduced applicability domain suggested by the submitter (n=79 or 80* chemicals)*			
	Resampling 5%-quantile (n=43+45)	*Point estimate: mode of repeat tests*	*Point estimate: weighted*	*Resampling 95%-quantile (n=43+45)*	*Resampling 5%-quantile (n=38+41)*	*Point estimate: mode of repeat tests*	*Point estimate: weighted*	*Resampling 95%-quantile (n=38+41)*
Specificity	*53.5%*	*58.1% (25/43)*	*58.9% (25.31/43)*	*65.1%*	*55.3%*	*63.2% (24/38)*	*62.8% (23.88/38)*	*71.1%*
Sensitivity	*87.0%*	*93.3% (42/45)*	*90.7% * (41.71/46)*	*93.5%*	*92.9%*	*97.6% (40/41)*	*96.4% * (40.49/42)*	*100%*
Accuracy	*na*	*76.1% (67/88)*	*75.3% (67.02/89)*	*na*	*na*	*81.0% (64/79)*	*80.5% (64.37/80)*	*na*

The reduced applicability domain that was proposed by the submitter was concluded on the basis of an OECD QSAR Toolbox analysis of organic functional groups (OFG) and correlation of these with false predictions. However, according to EURL ECVAM, the evaluation of predictive capacity for this reduced applicability domain cannot be considered robust enough as some of the limitations (e.g., acrylates) were assumed on the basis of a single chemical representing that particular OFG being tested.

Predictive capacity for the top-down approach, calculated by the submitter following resampling approach. The * sign means including chemical #23

	Top-down (n=88 or 89 chemicals)*			
	Resampling 5%-quantile (n=68+20)	*Point estimate: mode of repeat tests*	*Point estimate: weighted*	*Resampling 95%-quantile (n=68+20)*
Specificity	*76.8%*	*80.9% (55/68)*	*80.9% (55.81/69)*	*84.1%*
Sensitivity	*50.0%*	*50.0% (10/20)*	*53.3% (10.67/20)*	*60.0%*
Accuracy	*na*	*73.9% (65/88)*	*74.7% (66.48/89)*	*na*

Finally, also ROC analyses were performed upon the suggestion of EURL ECVAM to obtain a better picture of the overall performance of the OI assay in the prospective study:

ROC analysis and areas under ROC curves (AUROC) using data from the prospective validation study. Left part: submitter's input in its revised submission; Right part: EURL ECVAM's analysis

	Submitter's ROC analysis				*EURL ECVAM's ROC analysis*			
	Bottom-Up Approach		*Top-Down Approach*		*Bottom-Up Approach*		*Top-Down Approach*	
	n	*AUROC*	*n*	*AUROC*	*n*	*AUROC*	*n*	*AUROC*
IVI	*51*	*0.7813*	*51*	*0.6977*	*51*	*0.8017*	*51*	*0.6890*
IIVS	*50*	*0.8263*	*50*	*0.7366*	*50*	*0.8523*	*50*	*0.7362*
RP	*51*	*0.7915*	*51*	*0.7136*	*51*	*0.8009*	*51*	*0.7061*

References

Balls, M., Blaauboer, B.J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H. and Zucco, F. (1995) Practical aspects of the validation of toxicity test procedures. ECVAM Workshop Report 5. ATLA 23, 129-147.

ECHA (2015) Guidance on information requirements and Chemical Safety Assessment, Chapter R.7a: Endpoint specific guidance. Version 4.1: http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf

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

OECD (2012a) Test Guideline 405. OECD Guideline for the Testing of Chemicals: Acute Eye Irritation/Corrosion. Paris, France: Organisation for Economic Cooperation and Development. Section 4, OECD Publishing. doi: 10.1787/9789264185333-en. Available at: [http://www.oecd-ilibrary.org/environment/test-no-405-acute-eye-irritation-corrosion_9789264185333-en;jsessionid=151kmp6bvvsqe.x-oecd-live-03].

OECD (2012b) Test No. 460: Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264185401-en. Available at: [http://www.oecd-ilibrary.org/environment/test-no-460-fluorescein-leakage-test-method-for-identifying-ocular-corrosives-and-severe-irritants_9789264185401-en;jsessionid=151kmp6bvvsqe.x-oecd-live-03].

OECD (2013a) Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264203846-en. Available at: [http://www.oecd-ilibrary.org/environment/test-no-437-bovine-corneal-opacity-and-permeability-test-method-for-identifying-i-chemicals-inducing-serious-eye-damage-and-ii-chemicals-not-requiring-classification-for-eye-irritation-or-serious-eye-damage_9789264203846-en;jsessionid=151kmp6bvvsqe.x-oecd-live-03].

OECD (2013b) Test No. 438: Isolated Chicken Eye Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264203860-en. Available at: [http://www.oecd-ilibrary.org/environment/test-no-438-isolated-chicken-eye-test-method-for-identifying-i-chemicals-inducing-serious-eye-damage-and-ii-chemicals-not-requiring-classification-for-eye-irritation-or-serious-eye-damage_9789264203860-en;jsessionid=151kmp6bvvsqe.x-oecd-live-03].

OECD (2015a) Test 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, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264242432-en. Available at: [http://www.oecd-ilibrary.org/environment/test-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_9789264242432-en;jsessionid=151kmp6bvvsqe.x-oecd-live-03].

OECD (2015b) Test 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, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264242548-en. Available at: [http://www.oecd-ilibrary.org/environment/test-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_9789264242548-en;jsessionid=151kmp6bvvsqe.x-oecd-live-03].

Scott, L., Eskes, C., Hoffmann, S., Adriaens, E., Alepée, N., Bufo, M., Clothier, R., Facchini, D., Faller, C., Guest, R., Harbell, J., Hartung, T., Kamp, H., Varlet, B.L., Meloni, M., McNamee, P., Osborne, R., Pape, W., Pfannenbecker, U., Prinsen, M., Seaman, C., Spielmann, H., Stokes, W., Trouba, K., Berghe, C.V., Goethem, F.V., 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. Toxicol In Vitro 24, 1-9.

United Nations (UN) (2015) Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Sixth revised edition, UN New York, USA and Geneva, Switzerland. Available at: [http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev06/English/ST-SG-AC10-30-Rev6e.pdf].

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

Objective *Why does EURL ECVAM require advice on the current issue?*	EURL ECVAM requests an ESAC opinion on the reliability (reproducibility within and between laboratories of results obtained *in vitro*) and relevance (predictive capacity of effects documented *in vivo*) of the Ocular Irritection® test method for prediction of eye irritation potential of chemicals. The opinion of ESAC should support EURL ECVAM with respect to the possible development of an EURL ECVAM recommendation on the Ocular Irritection® for serious eye damage/eye irritation testing outlining (1) the scientific basis of the assay, (2) its overall performance (transferability, reproducibility and predictive capacity) as assessed during the validation study and based on other (e.g. published) information, (3) its applicability and limitations, and 4) its proposed use. ESAC's advice should enable a conclusion on the potential adequacy of the Ocular Irritection® for routine testing of serious eye damage/eye irritation for regulatory purposes.

4.2 QUESTION(S) TO BE ADDRESSED

Questions *What are the questions and issues that should be addressed in view of achieving the objective of the advice?*	The ESAC peer review of the Ocular Irritection® should address the following aspects: (1) Scientific basis in relation to serious eye damage/eye irritation. (2) Clarity of the test definition, including: - purpose and need of the test method. - biological/mechanistic relevance in relation to the test system used and the endpoint measured. - protocol clarity and completeness. - clarity and adequacy of the prediction model and its development. (3) Clarity of the definition of the study objective(s). (4) Appropriateness of the study design and execution considering the study objective(s), including: - number and selection criteria for test chemicals (e.g., range of documented effects *in vivo*, etc.). - quality assurance of reference data (*in vivo*) for predictive capacity assessment. - number of participating laboratories. - number of replicates, number of repetitions, rules for retesting and handling of deviations. (5) Study management and conduct. (6) Results compilation and statistical analyses reporting, including: - appropriateness of calculation of WLR and BLR on the basis of the generated data.

	- appropriateness of calculation of Predictive Capacity on the basis of the generated data. - appropriateness of identification of limitations/applicability domain on the basis of the generated data. (7) Transferability and reproducibility (WLR/BLR). (8) Predictive capacity and relevance to a tiered (Top-Down/Bottom-Up) testing strategy when used for: - distinguishing chemicals not requiring classification from chemicals requiring classification as Category 1 (serious eye damage) or Category 2 (eye irritation). - distinguishing chemicals requiring classification as Category 1 (serious eye damage) from chemicals not requiring classification as Category 1 (i.e. Category 2 (eye irritation) or No Category (not classified)). (9) Applicability and any known limitations, assessed from the selection of the test chemicals (range of molecular class and physical properties) and analyses of possible reasons for misclassifications. (10) Completeness and adequacy of the Performance Standards proposed by the submitter, including the Essential Test Method Components, the list of Reference Chemicals and the Target Values for Reproducibility and Predictive Capacity. (11) Possible gaps, if any, between study design and study conclusions. (12) Whether the information provided in the submission is sufficient to substantiate the proposed use of the test method within a Bottom-Up/Top-Down testing strategy. (13) What additional work, if necessary, should be undertaken in future to further characterise the test method and its proposed use. ESAC's advice should conclude on the regulatory applicability of the Ocular Irritection® (i.e., for implementation as an EU test method and OECD Test Guideline).

4.3 TIMELINES

Timelines concerning this request *When does EURL ECVAM require the advice?*	**Timeline**	**Indication**
	Finalised ESAC Opinion required by:	**June 2016**
	Request to be presented to ESAC by written procedure (e.g. due to urgency) prior to the next ESAC	**YES**
	Request to be presented to ESAC at ESAC plenary meeting	**NO**

5. EURL ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC

5.1 EURL ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED

Specific structures required within ESAC to address the request *Does the advice require an ESAC working group, an ESAC rapporteur etc.?*	**Structure(s) required**	**Required according to EURL ECVAM? (YES/NO)**
	S1 ESAC Rapporteur	**NO**
	S2 ESAC Working Group	**ESAC members** - José M. Navas (Chair) - Kristina Kejlová - Annete Kopp-Schneider - Renate Kraetke - Jon Richmond **ICATM nominations** - Dave Allen (NICEATM/ICCVAM) - Kyung-Min Lim (College of Pharmacy, Ewha Womans University; nominated by KoCVAM)
	S3 Invited Experts	**NO**
	Ad S3: If yes – list names and affiliations of suggested experts to be invited and specify whether these are member of the EEP	
	If other than above (S1-S3):	

5.2 DELIVERABLES AS PROPOSED BY EURL ECVAM

Deliverables *What deliverables (other than the ESAC opinion) are required for addressing the request?*	**Title of deliverable other than ESAC opinion**	**Required? (YES/NO)**
	D1 ESAC Rapporteur Report and draft opinion	**NO**
	D2 ESAC Working Group Report and draft opinion	**YES**
	If other than above (D1-D2):	

6. LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC

Count	Description of document	Already available? (YES/NO)	File name
1	Ocular Irritection® test submission (TST) (latest version submitted to EURL ECVAM - clean)	Yes	1a_TST_OI_assay_30_12_2013_Rev10April2015_cleared.pdf
2	Ocular Irritection® test submission (TST) (latest version submitted to EURL ECVAM – with tracked changes)	Yes	1b_TST_OI_assay_30_12_2013_Rev10April2015_TC(vs. v.30Dec2013).docx
3	Letter from test method submitter summarising the main amendments made in the latest revised version of the Ocular Irritection® test submission	Yes	0a_AccompanyingLetter_OI_TST_finalSubmission_10April2015.pdf
4	Attestation letter from test method submitter clarifying the roles of each party in the submission	Yes	0b_AttestationLetter_8April2015.pdf
5	EURL ECVAM assessment report on the Ocular Irritection® test submission (updated version considering the last revised submission)	Yes	Ocular_Irritection_updated_assessment_report_2016-05-09_final.pdf
6	Standard Operating Procedure (SOP) of the Ocular Irritection®	Yes	1_Att1a_SOP_OI.pdf
7	DB-ALM protocol n. 157	Yes	2a_Att1b_DBAlm157_OI.pdf
8	Details of kit components, software setup, data calculation, and example of an OI assay report (CONFIDENTIAL)	Not on CIRCABC (provided by e-mail only)	2b_Att1c_CONFIDENTIAL_KitComp_SoftSetup_DataCalc_ExplOIReport_NEW_15Oct2014.pdf
9	Interpretation of Ocular Irritection® data (CONFIDENTIAL)	Not on CIRCABC (provided by e-mail only)	2c_Att1d_CONFIDENTIAL_InterpOIData_NEW_15Oct2014.pdf
10	Quality Audit Report	Yes	3a_Att1e_IVI_Quality_Audit.pdf
11	Quality Audit Questionnaire	Yes	3b_Att1e_IVI_Quality_Audit_Questionnaire.pdf

12	Identity and characteristics of the 56 chemicals tested in the prospective validation study (WLR, BLR)	Yes	4_Att3_ID_ProspChem_WLR_BLR.pdf
13	WLR and BLR results	Yes	5_Att4_WLV_and_BLV_Results_rev8April2015.pdf
14	Transferability test items and results	Yes	6_Att5_Transf_Results.pdf
15	Training and transferability report	Yes	7_Att6_Train_Transf_Report.pdf
16	Identity and characteristics of the chemicals tested in the prospective and retrospective validation study (PC)	Yes	8_Att11_ID_Prosp_Retrosp_Chem_PC.pdf
17	Predictive Capacity results	Yes	9a_Att12a_PC_Results_rev29March2015.pdf
18	MSDS of chemical n. 12, Tetraethylene glycol diacrylate, CAS 17831-71-9	Yes	9b_Att12b_MSDS_Chemical12_Tetraethylene glycol diacrylate_NEW_8April2015.pdf
19	Results of the evaluation on possible reasons for misclassification	Yes	9c_Att12c_ReasonsMissclassification_Rev8April2015.pdf
20	Project plan(s)	Yes	10_Att13_Project Plan_8March2011_Rev8April2015.pdf
21	Overview of test items used in the validation study	Yes	11_Att14_Overview_Chem_VS.pdf
22	Draft Performance Standards for validation of a similar or updated test method	Yes	12_Att16_DraftPerfStand_OI_NEW_10April2015.docx
23	Peer-reviewed publication of the validation study and its outcomes	Yes	13_Att17a_Eskesetal_OIValidation_TIV_2014_NEW_10April2015.pdf

7. TERMS OF REFERENCE OF THE ESAC WORKING GROUP

7.1 ESTABLISHMENT OF THE ESAC WORKING GROUP

The ESAC unanimously agreed by written procedure on the 18th of February 2016 on the composition of a new ESAC Working Group for the review of test methods in the area of serious eye damage/eye irritation.

7.2 TITLE OF THE ESAC WORKING GROUP

Full title:
ESAC Working Group on Eye Irritation Test Methods

Abbreviated title:
ESAC WG Eye Irritation

7.3 MANDATE OF THE ESAC WORKING GROUP

The EWG is requested to conduct a scientific review of the Ocular Irritection® validation study.. The review needs to address the questions put forward to ESAC by EURL ECVAM under section 4.2 of the current request.

The review should focus on the appropriateness of design and conduct of the study in view of the study objective and should provide an appraisal to which extent the conclusions of the test submitter are substantiated by the information generated during the study and how the information generated relates to the scientific background available.

7.4 DELIVERABLES OF THE ESAC WORKING GROUP

The ESAC WG is requested to deliver to the chair of the ESAC and the ESAC Coordinator a detailed **ESAC Working Group Report** outlining its analyses and conclusions and a **draft ESAC Opinion.** A template has been appended (Appendix 1) intended to facilitate the drafting of the WG report.

The conclusions drawn in the report should be based preferably on consensus. If no consensus can be achieved, the report should clearly outline the differences in the appraisals and provide appropriate scientific justifications.

7.5 PROPOSED TIMELINES OF THE ESAC WORKING GROUP

Item	Proposed date/time	Action	Deliverable
1	6 May 2016	Teleconference of the Working Group	Agree procedure
2	11-13 May 2016	Working Group meeting	Draft ESAC WG report and draft ESAC opinion
3	27 May 2016	Circulation of final WG report and draft ESAC opinion to ESAC	Final draft ESAC WG report and draft ESAC opinion
4	9-10 June 2016	Endorsement of WG report and ESAC opinion at ESAC42 meeting	Final ESAC WG report and ESAC opinion

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC WORKING GROUP

The review should address the **questions put forward to ESAC by EURL ECVAM** (see section 4.2) and the information requirements of the ESAC Working Group Template, where applicable. The ESAC Coordinator will provide guidance if needed.

When preparing the final ESAC WG report to address these questions, the ESAC WG is requested to use a pre-defined reporting template. This template (see appendix 1) follows EURL ECVAM's modular approach and addresses to which extent the standard information requirements have been addressed by the study. The template allows moreover for addressing the issues specific studies outlined in section 4.2. The Secretariat will provide guidance if necessary.

APPENDIX 1 REPORTING TEMPLATE

The appended ESAC WG template suggests a structure that is in close agreement with the EURL ECVAM information requirements ("modules") for scientific review following validation and allows at the same time for the description of the analysis and conclusions concerning more specific questions.

The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies). Depending on the study type and the objective of the study, not all sections may be applicable.

However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

The current template is

TEMPLATE_ESAC-WG_REPORT-v6.doc

Annex 3

ESAC WORKING GROUP PEER REVIEW CONSENSUS REPORT

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC Working Group Peer Review Consensus Report

on the

Ocular Irritection® Test Method for Prediction of Serious Eye Damage/Eye Irritation Potential of Chemicals

Title page information	
File name	ESAC_WG_Report_Ocular_Irritection.doc
Abbreviated title of ESAC request	Ocular Irritection® validation
Relating to ESAC REQUEST No.	2016-01
Request discussed through	Written procedure previous to ESAC 42
Report to be handed over to ESAC Chair and EURL ECVAM Coordinator by	José M. Navas

Version tracking

Date	Version	Author(s)	Description
13 May 2016	V1.0	ESAC WG	First ESAC WG agreed draft
19 May 2016	V2.0	ESAC WG	Second revised draft after commenting
22 May 2016	V3.0	ESAC WG	Third revised draft after commenting
27 May 2016	V4.0	ESAC WG	Fourth revised draft after commenting
01 June 2016	V5.0	ESAC WG	Fifth revised draft after commenting
07 June 2016	V6.0	ESAC WG	Final ESAC WG approved draft sent to ESAC for endorsement

Table of Contents

ESAC Working Group

Full title: ESAC Working Group on Eye Irritation Test Methods

Abbreviated title: ESAC WG Eye irritation

The ESAC WG was established in March 2016 by written procedure to assist in the production of an ESAC Opinion by undertaking a peer review of a three laboratory ring trial of the Ocular Irritection® (OI) Assay, a test method with a claimed wide applicability domain, developed for the prediction of the eye irritation potential of liquid and solid chemicals, specifically to distinguish chemicals requiring official classification for eye irritation or serious eye damage according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS. UN, 2015) from chemicals not requiring classification (no classification category; 'non-irritants').

Following a teleconference on 6 May 2016 the ESAC WG met at EURL-ECVAM 11-13 May 2016 to conduct its peer review.

The ESAC WG members appointed by ESAC were:

- Dr. José M. Navas (ESAC Member, WG Chair)
- Dr. Kristina Kejlová (ESAC Member)
- Prof. Annette Kopp-Schneider (ESAC Member)
- Dr. Renate Krätke (ESAC Member)
- Dr. Jon Richmond (ESAC Member)
- Dr. Dave Allen (Nomination by NICEATM/ICCVAM)
- Prof. Kyung-Min Lim (Nomination by KoCVAM)

ESAC Coordination:

- Dr. João Barroso (ESAC Coordinator)

Abbreviations used in the document

- **BLR** Between-laboratory reproducibility
- **CI** Confidence Interval
- **EIT** Eye Irritation Test
- **ESAC** EURL ECVAM Scientific Advisory Committee
- **ESAC WG** ESAC Working Group
- **EU CLP** European Union Regulation on Classification, Labelling and Packaging of Substances and Mixtures
- **EURL ECVAM** European Union Reference Laboratory for Alternatives to Animal Testing
- **GLP** Good Laboratory Practice
- **IATA** Integrated Approach to Testing and Assessment
- **IDE** Irritation Draize Equivalent
- **MAS** (in vivo, Draize test) Maximum Average Scores
- **MQS** Maximum Qualified Score (highest estimated IDE score)
- **OD** Optical density
- **OECD** Organisation for Economic Co-operation and Development
- **OI** Ocular Irritection® (OI) Assay
- **PM** Prediction Model
- **RhCE** Reconstructed Human Cornea-like Epithelium
- **ROC** Receiver Operation Characteristics
- **SOP** Standard Operating Procedure
- **TST** Test Submission Template
- **UN GHS** United Nations Globally Harmonized System for the Classification and Labelling of Chemicals.
- **VMG** Validation Management Group
- **WLR** Within-laboratory reproducibility

1. Study objective and design

1.1 Analysis of the clarity of the study objective's definition

(a) ESAC WG summary of the study objective as outlined in the Test Submission

The prospective Ocular Irritection® (OI) assay validation study pre-defined study objectives were:

- To formally evaluate the usefulness in terms of relevance (predictive capacity) and reliability (transferability, and reproducibility within and between laboratories) of the OI assay to reliably discriminate chemicals not requiring classification for serious eye damage/eye irritancy (No Category) from chemicals requiring classification and labelling (Category 1 and Category 2) according to the UN GHS Classification and Labelling of Chemicals (UN GHS. UN, 2015) and as implemented by the EU CLP regulation (EU CLP. EC, 2008a). This test method is not intended to differentiate between GHS Categories 1 (irreversible effects) and 2A-B (reversible effects).
    - N.B. The test method was not designed, intended, or evaluated to test gases or aerosols.
- A further post-hoc evaluation was performed to evaluate the usefulness of the OI assay to discriminate chemicals inducing serious eye damage (UN GHS Category 1) from other classes.
- In addition, to produce evidence and analysis to support the test method being incorporated into a tiered testing strategy (so-called Bottom-Up/Top-Down testing strategy, Scott L. *et al.*, 2010). The ultimate purpose of such a tiered testing strategy being to replace the traditional *in vivo* Draize eye test [Method B.5 of EC Regulation 440/2008 (EC, 2008b) or OECD TG 405 (OECD, 2002)].

(b) Appraisal of clarity of study objective as outlined in the Test Submission

The ESAC WG believes that the study objectives were sufficiently clear, and determined the way the study was designed, conducted, analysed, and reported in the TST. The WG notes that the original goal of the prospective validation study, as reflected for example in chemical selection, was restricted to discriminating eye irritants from non-classified (Bottom-Up approach) (Scott *et al*., 2010). The data and analysis submitted in the TST, however, also includes consideration of its use in a Top-Down approach.

1.2 Quality of the background provided concerning the purpose of the test method

The TST clearly describes the intended application of the OI assay as being for regulatory testing.

The TST discusses the possible use of this assay in future *in vitro* tiered testing strategies (Scott *et al*, 2010) within an integrated approach using strategic combinations of alternative test methods to replace the Draize eye test.

(a) Analysis of the scientific rationale provided in the Test Submission

The TST describes the relevance and scientific rationale of the OI assay as evaluating the ocular hazard effects of chemicals based on the premise that eye irritation and corneal damage after exposure to irritating chemicals results from the disruptive effects ocular irritants may have on the highly organized structure of corneal macromolecular matrix of proteins and carbohydrates.

According to Scott *et al*. (2010) mechanisms of eye irritation within the cornea include '*coagulation'* (the precipitation/denaturation of macromolecules, particularly proteins), '*saponification'* (the breakdown of lipids), and '*actions of macromolecules'*. The reported mechanisms leading to ocular irritation *in vivo* include denaturation of collagen, loss of glycosaminoglycans (polysaccharide), and

saponification of lipids such as in the case of alkalis; coagulation and precipitation of proteins such as in the case of acids; dissolution of lipids such as in the case of solvents (Eskes *et al*., 2010).

The OI assay manufacturer considers the identity of the raw material used to produce the macromolecular assay matrix to be confidential: however the ESAC WG finds it is already identified in a number of public domain resources.

The raw material for OI assay matrix is plant-based, and purchased in bulk in powder form from a third party: neither the supplier nor the assay manufacturer have chemically characterised, defined, or standardised the plant extract or the chemical composition of the assay matrix. The information available to the ESAC WG does not explain why or how the raw material was originally selected to produce the test kit macromolecular matrix as a surrogate for the human cornea.

The supporting documentation claims a mechanistic basis based on the resemblance of the assay macromolecular matrix with the human cornea. However, in addition to the assay matrix never having been fully chemically characterised and not being of mammalian origin, there is no detailed description of the nature of the healthy human corneal matrix, or how it changes after exposure to ocular irritants. The documentation also does not explain how the physical chemical alterations occurring in the kit matrix upon exposure to irritant chemicals compare to effects observed with the *in vivo* Draize eye test. Furthermore, it is not clear to the ESAC WG why the chemical changes induced in the assay matrix should be specific to ocular irritants, rather than being also applicable to other classes of irritants (e.g. skin irritants). The ESAC WG therefore considers that the OI assay's mechanistic relevance to predict adverse ocular effects of chemicals in humans is poorly defined and that the data presented in the test submission represent an empirically observed relationship (correlative relationship), rather than a proven mechanistic relationship, between changes to the matrix and the potential of test chemicals to produce ocular irritation.

The OI assay is not intended to provide insights into persistence of chemically-induced ocular injuries, which could result in the underprediction of chemicals classified *in vivo* as Category 1 due only to persistence of effects, if the test method is used to identify Category 1 chemicals, e.g., as a first step in a Top-Down approach. In addition, if this test method is to be used to identify non-irritants within an Integrated Approach to Testing and Assessment (IATA) then the ESAC believes other test methods may be needed to assess the vascular and inflammatory components of the adverse outcome pathway for eye irritation, and to take account of the lack of epithelial barrier function.

(b) Analysis of the regulatory rationale provided in the Test Submission

The TST identifies relevant regulatory requirements. In the view of the ESAC WG the relevant legislation and regulations are appropriately referenced; and the regulatory requirements and the role of the non-animal methods in the context of the regulatory requirements are adequately specified.

If validated and accepted for regulatory use, the OI assay could contribute to a reduction in animal testing by reliably identifying chemicals not requiring classification (No Category) and chemicals inducing serious eye damage (Category 1) when used within an appropriate non-animal testing strategy (Scott *et al*, 2010). This ESAC WG considers that in view of the high prevalence of non-classified chemicals (Adriaens *et al*, 2014), non-animal test methods validated for this purpose could contribute to reducing animal testing by identifying the much larger number of chemicals not requiring classification.

1.3 Appraisal of the appropriateness of the study design

The study and data reported in the TST generally comply with the principles and criteria set out in the OECD Guidance Document on the Validation and International Acceptance of New or Updated Test

Methods for Hazard Assessment (No. 34. OECD, 2005), and described in the generally accepted Modular Approach to validation (Hartung *et al*, 2004).

The ESAC WG notes in particular:

- the comprehensive SOP for test method implementation;
- the chemical selection for the prospective study was structured in favour of assay use in a Bottom-Up approach;
- appropriate training facilitating transfer to a naïve laboratory was planned and undertaken;
- there was generally appropriate separation of responsibilities for the ring trial (project management, chemicals management, data management);
- defined controls and calibration materials were used;
- descriptions were provided of how WLR, BLR, specificity, sensitivity, accuracy were to be calculated;
- the degree of independence used for data statistics analysis.

However, the ESAC WG also notes the following shortcomings:

- there is no quality assurance system to define and control the chemical composition of the raw material used to produce the assay matrix, and no detailed chemical specification or chemical characterisation of the assay matrix before test kit batch release;
- there is a lack of transparency associated with the software and algorithms used in the test method;
- information required for the users and potential users to interpret some test results is considered confidential;
- the independent statistician involved in the study was also the vice-chair of the Validation Management Group (VMG);
- in addition to qualified and non-qualified runs the submitter makes use of a third category, excluded runs, which impacts on the calculation of the test performance when comparisons are made with the reported performance of other test methods.

For the prospective ring trial, reliance was placed on a statistical power analysis (sample size calculation) estimating the minimum requirements for the assay's use within a Bottom-Up approach (n=50). To make provision for unforeseeable events, 56 chemicals were used in the ring trial (five of which were subsequently excluded from the analysis of the test method performance).

The relevance and reliability of the OI assay as reported by the test manufacturer also took account of retrospective data from 45 chemicals, eight of which were also used in the prospective ring trial.

Accordingly, the analysis presented by the test method developer in the TST is based on the classification of the eye damage/irritation potential of a total of 88 chemicals (n= 56 – 5 + 45 – 8 = 88).

The test method performance acceptance criteria proposed by the test manufacturer are based on the reported performance of other validated *in vitro* methods used to assess chemical eye damage/irritation potential, not all of which would fulfil similar roles in a future integrated testing strategy.

In its scientific review of the study, taking into account the study findings, and the OI assay TST conclusions, the ESAC WG concludes that, subject to specific qualifications set out above and below, the study design was generally appropriate and robust, the acceptance criteria applied to the test results were appropriate. In addition, the study report provides sufficient data, evidence and analysis for the ESAC WG to conduct a peer review of to what extent the study objectives were satisfied.

1.4 Appropriateness of the statistical evaluation

Although the Study Plan specifies the production of a stand-alone statistical report, there is no formal separate statistical report included in the documents available to the ESAC WG. The available statistical information is contained in the TST and associated annexes and additional information supplied by the test manufacturer at the request of the ESAC WG.

The initial calculation to determine the minimal number of chemicals to be used in the ring trial was appropriate for a test intended to be used within a Bottom-Up approach (see 1.3 above).

In the case of the WLR and BLR the WG considers that although point estimates were calculated, Confidence Intervals (CIs) were not supplied, therefore the precision of the estimated WLR and BLR values cannot be fully assessed. The ESAC WG also believes that including "concordant" non-qualified and excluded test results in the analysis is not sound.

The predictive capacity appears to have been calculated correctly, and one-sided 95 %-CIs (similar to two-sided 90 %-CIs) based on resampling methods were supplied. However, the WG notes that validation studies for other test methods typically provide two-sided 95 %-CIs. Furthermore, there is a lack of clarity about how the resampling was performed.

The ESAC WG cannot say with certainty whether the cumulative effects of these perceived short-comings, including the provision made to "exclude" test runs, may have tended to over-estimate the performance of the assay in comparison to the datasets presented for evaluation from other validation studies.

2. Collection of existing data

2.1 Existing data used as reference data

The TST and annexes provide sufficient information on this point. The TST references and relies on an extensive collection of information generated before and during the development and pre-validation of the test method.

Chemical selection was carried out in collaboration with EURL ECVAM in order to avoid bias in the process, to safeguard the independence of the chemicals selection process, and to facilitate comparison of OI assay's performance to other assays evaluated or being evaluated for eye damage/irritation testing.

Databases consulted as sources for chemical selection included the ICCVAM (TSCA) list of reference chemicals (ICCVAM, 2007), the ECETOC database (ECETOC, 1998) and EURL ECVAM suggested chemicals.

The essential requirements for chemical selection for the ring trial were toxicological and physico-chemical properties, chemical class, as well as the availability of complete and quality assured supporting *in vivo* data to allow comparative evaluation of the predictive capacity of the test method as measured against the *in vivo* (Draize eye test) reference method.

According to the performed calculations, the minimum number of chemicals needed from a statistical point of view for the evaluation of the predictive capacity of the assay for use within a Bottom-Up approach was 50 (27 non-classified + 23 classified chemicals). Additionally, six more chemicals were included (4 non-classified and 2 classified ones) to compensate for unforeseeable events. In conclusion, a total of 56 chemicals were selected for the prospective ring trial, including 31 non-classified and 25 classified chemicals for eye damage/irritation (12 Category 1 and 13 Category 2 chemicals).

A wide range of chemical properties, use/function (surfactants were also included), and physico-chemical properties were taken into consideration for the selection. In addition, the US EPA eye hazard categories were also described (18 EPA Category IV, 20 EPA Category III, 6 EPA Category II, 7 EPA Category I and 5 having study criteria not met to be able to assign an unequivocal EPA Category).

The only other consideration was to ensure that the selected chemicals were consistent with the pre-defined applicability domain of the OI assay as defined in its SOP.

2.2 Existing data used as <u>testing</u> data

The TST analysis also took account of test results with additional 45 chemicals. Such results were generated with the same protocol as that used in the prospective validation study. All chemicals had good quality *in vivo* data and fell within the pre-defined OI assay applicability domain. 37 chemicals were not tested in the prospective validation dataset, but eight had also been tested during the prospective validation study.

As a consequence, and disregarding the four chemicals whose ring trial results were classified as “excluded” in all tests performed by the three laboratories and one other chemical for which "non-qualified" results were obtained in all tests performed by two of the three laboratories (with the third laboratory obtaining three qualified tests), a total of 88 single chemicals with both *in vivo* and *in vitro* data were used by the test manufacturer to evaluate the predictive capacity of the OI assay (n= 56 – 5 + 45 – 8 = 88. See 1.3 above).

These chemicals comprised 43 UN GHS / EU CLP Non-Classified and 45 UN GHS / EU CLP Classified including 20 Category 1 and 25 Category 2 chemicals (21 Cat. 2A and 4 Cat. 2B). The composite dataset had a majority of liquids as compared to solids (56 liquids, 25 solids, and 7 viscous materials). Furthermore, the distribution according to the US EPA classification categories included 26 EPA Cat IV, 26 EPA Cat III, 15 EPA Cat II, 13 EPA Cat I and 8 to which no unequivocal EPA Category could assigned.

2.3 Search strategy for retrieving existing data

See Section 2.1 above.

2.4 Selection criteria applied to existing data

See Sections 2.1 and 2.2 above.

3. Quality aspects relating to data generated during the study

3.1 Quality assurance systems used when generating the data

The TST claims laboratories participating in the prospective ring trial worked in accordance with OECD GLP principles, such as, but not limited to, the use of SOPs, adequate data recording, and record keeping. No further information was supplied.

3.2 Quality check of the generated data prior to analysis

The data management procedures and statistical tools and methods applied were approved by the Ocular Irritection® test method VMG. The independent statistician who analysed the data was also vice-chairman of the VMG.

According to the project plan, raw data produced by the participating laboratories were entered into spreadsheets provided by the lead laboratory. The study biostatistician collated, audited and analysed the data. The data compiled were also audited by the participating laboratories. However, there is no formal separate statistical report included in the documents available to the ESAC WG.

4. Quality of data used for the purpose of the study (existing and newly generated)

4.1 Overall quality of the evaluated testing data (newly generated or existing)

The data were of sufficient quality to apply the predetermined acceptance criteria and prediction model.

However, the ESAC WG felt it necessary to conduct supplementary analysis in view of the perceived shortcomings in the manufacturer's data analysis. See 1.4 above.

4.2 Quality of the reference data for evaluating relevance[1]

Draize eye test reference data, as those used in the present study, are considered those of the best quality for the purpose of the validation exercise. The quality of the reference data used here is equivalent to that of data used in previous *in vitro* test method eye damage/irritation validation studies.

4.3 Sufficiency of the evaluated data in view of the study objective

The sample size exceeded the minimum number of chemicals as determined by the power calculation for Bottom-Up analysis. See sections 1.2 and 2.2 above.

The assumptions made for the sample size calculation (with an assumed sensitivity of 95 % with lower confidence bound of 75 %, and a specificity of 50 % with lower confidence bound of 25 %) were close to the values obtained in the final evaluation of the predictive capacity (sensitivity was slightly lower and specificity slightly higher than expected; see 10.1 below). In the prospective validation study 9 % of tests (46 out 504) were coded as excluded or non-qualified (representing, out of 56 chemicals, 4 with excluded results in 9 out of 9 tests, 1 with exclude in results in 4 out of 9 tests and 1 with "non-qualified" results in 6 out of 9 tests). The 5 chemicals with a majority of excluded of non-qualified results were then excluded from predictive capacity analyses due to issues encountered with the blank check (blank max failed or blank not flat). See sections 1.3, 1.4, and 2.2 above.

[1] OECD guidance document No. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

5. Test definition (Module 1)

5.1 Quality and completeness of the overall test definition

The Ocular Irritection® assay is based on the Eytex® method (Kelly, 1989; Gordon, 1992) taking account of recommendations made by Balls *et al*. (1995). It comprises single protocol, clear procedures for surfactant testing, and a defined applicability domain.

The major underlying assumption behind the Ocular Irritection® assay is that eye irritation and corneal injury after exposure to irritant chemicals result from the disruptive effects those chemicals have on the highly organized matrix structure of corneal proteins and carbohydrates. The test method utilises a plant based matrix (derived from jack bean *Canavalis enisformis*. Loprieno, 1995; see: link 1; link2), which is claimed to become turbid after exposure to ocular irritants.

The chemical composition of the Ocular Irritection® assay matrix is not fully characterized or specified in the test submission. Supplementary information supplied by SeCAM (SeCAM Services & Consultation on Alternative Methods Sagl, Switzerland) indicates neither the chemical composition of the raw material or the resulting test matrix have been fully chemically characterised or defined either by the supplier or test kit manufacturer.

To perform the assay, test materials are applied to the surface of the Ocular Irritection® assay plant-extract matrix, with a membrane disk used to ensure proper coverage, in defined quantities and for a set time. Chemicals that produce ocular irritation are expected to produce physico-chemical changes including alterations in protein conformation and the degree of hydration resulting in turbidity in the matrix, with the turbidity degree being proportionate to the ocular irritation potential of the test material – as judged against positive and negative controls, and calibration materials. The change in matrix turbidity produced by five concentrations of the test material (and the control and calibration samples) is measured by the light scattering detected by a spectrophotometer set to a wavelength of 405 nm. The software supplied with the assay then calculates a dose response curve for the test material (the precise algorithms used are not described in detail in the TST). The turbidity as estimated by the software is converted to a numerical Irritection Draize Equivalent (IDE), and the highest estimated IDE score, termed the Maximum Qualified Score (MQS), is then used in the prediction model to categorise the ocular hazard potential of the test chemicals according to the UN GHS and EU CLP classification systems (UN, 2015; EC, 2008a).

The Ocular Irritection® assay software automatically performs a number of qualification checks against pre-defined parameters to ensure assay performance. The software has been designed to accept the sample data only if specific acceptance criteria as detailed in the TST are met. Results may be shown as qualifying, non-qualifying, or excluded. Users are provided with instructions to deal with ambiguous or equivocal results.

The Prediction Model (PM) is based on that initially developed for the Eytex assay to predict *in vivo* Maximum Average Scores (MAS). The PM as used for this validation study was confirmed as optimal with the Ocular Irritection® assay for application to UN GHS/EU CLP classification using a set of 23 chemicals (9 non-classified, 12 Category 2 and 2 Category 1 chemicals based on the UN GHS/EU CLP classification system) (Eskes *et al*., 2014). In the prospective validation study the PM was used primarily to distinguish classified from non-classified chemicals using the previously existing cut-off of 12.5 (i.e. Bottom-Up approach) (Table 5.1.1).

Table 5.1.1. Maximal Qualified Score (MQS) values and correspondence with the degree of ocular irritancy and the UN GHS/EU CLP classification.

Maximal Qualified Score (MQS)	UN GHS / EU CLP classification
0 – 12.5	No Category
> 12.5	Categories 1 and 2

After the ring trial a post-hoc evaluation of the data was carried out to evaluate the predictive capacity of the assay using a cut-off value of 30.0 for the identification of UN GHS/EU CLP Category 1 (i.e. Top-Down approach) (Table 5.1.2).

Table 5.1.2. Maximal Qualified Score (MQS) values and correspondence with the degree of ocular irritancy and the UN GHS/EU CLP classification.

Maximal Qualified Score (MQS)	UN GHS / EU CLP classification
0 – 30	No Category and Category 2
> 30	Category 1

From the information available it is not clear why the changes to the test method matrix should be specific or restricted to ocular irritants, as opposed to other categories of irritants. Nevertheless the empirical evidence tends to confirm a correlative relationship with the *in vivo* results rather than a proven biological/mechanistic relationship as claimed by the submitter.

5.2 Quality and completeness of the documentation concerning SOPs and prediction models

The SOPs are detailed and complete and the PM is defined (see section 5.1). However, for the evaluation of ambiguous and equivocal results those using the assay must refer to a supplementary, confidential facts sheet provided by the manufacturer.

6. Test materials

6.1 Sufficiency of the number of evaluated test items in view of the study objective

See sections 1.3, 2.1, and 2.2 above.

6.2 Representativeness of the test items with respect to applicability

See sections 2.1 and 2.2 above.

7. Within-laboratory reproducibility (WLR) (Module 2)

7.1 Assessment of repeatability and reproducibility in the same laboratory

The within-laboratory reproducibility was assessed in two ways as the software provides Maximum Qualified Score, which ranges from 0 to 51 being truncated at the upper border of 51, and simultaneously also indicates if a sample is not qualified (NQ) or excluded. The rules for assessing such cases are defined in the Ocular Irritection SOP.

Firstly, the analyses focused on the chemicals with non-truncated Maximal Qualified Scores for which the SD were calculated. The standard deviation of the three samples per chemical was considered as an appropriate measurement of within-laboratory reproducibility. Considering conservatively an SD below 4, which according to the submitter represents a low variability in relation to the dynamic response range of the OI assay that spans from 0 to 51, as an indicator for acceptable within-laboratory reproducibility, the calculation resulted in 93.3 % (IVI), 84.4 % (IIVS) and 91.3 % (RP) WLR. These values are reported in the TST as a point estimate rather than with CIs. Although the ESAC WG has not calculated the CIs, based on the data sets they should have been acceptably narrow.

Secondly, the WLR was evaluated individually for each laboratory and for both cut-offs (i.e., 12.5 for Bottom-Up and 30.0 for Top-Down) in terms of concordance of the predictions of the three independent experiments carried out per chemical in each laboratory. In this case, the WLR values were calculated using data from qualified tests, but also with chemicals that all three laboratories consistently coded as non-qualified or excluded being deemed to be concordant and therefore included in the calculation. The concordances based on predictions are shown in Table 7.1.1. The ESAC WG discovered two errors in the TST in the reporting of the WLR for RP: a total of 9 and not 8 discordant predictions and a total of 47 and not 46 concordant predictions were obtained. The reported WLR value of 83.9 % for RP is nevertheless correct.

Table 7.1.1. Calculation of WLR when considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay. IVI, InVitro International laboratory; IIVS, Institute for In Vitro Sciences laboratory; RP, Res Pharma laboratory.

	Within-laboratory Reproducibility					
	Cut-off 12.5			Cut-off 30.0		
	IVI	IIVS	RP	IVI	IIVS	RP
concordant predictions	45	42	41	46	43	43
discordant predictions	6	8	11	5	7	9
concordant excluded	4	5	4	4	5	4
concordant NQ	1	1	0	1	1	0
overall concordant	50	48	45	51	49	47
% concordance	**89.3 %**	**85.7 %**	**80.4 %**	**91.1 %**	**87.5 %**	**83.9 %**

However, the ESAC WG believes that including “concordant” non-qualified and excluded test results in the analysis is not sound. Taking this into account, the WLR values and their respective Wilson two-sided 95 %-CIs were recalculated by the ESAC WG including only qualified results. The ESAC WG places reliance on the WLR values calculated on this basis (Table 7.1.2.).

Table 7.1.2. Calculation of WLR and BLR NOT considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay. The values in brackets correspond to Wilson two-side 95 %-CIs. IVI, InVitro International laboratory; IIVS, Institute for In Vitro Sciences laboratory; RP, Res Pharma.

	Within-laboratory Reproducibility					
	Cut-off 12.5			Cut-off 30.0		
	IVI	IIVS	RP	IVI	IIVS	RP
concordant predictions	45	42	41	46	43	43
discordant predictions	6	8	10	5	7	8
% concordance	88.2 % (76.6-94.5 %)	84.0 % (71.5-91.7 %)	80.4 % (67.5-89.0 %)	90.2 % (79.0-95.7 %)	86.0 % (73.8-93.1 %)	84.3 % (72.0-91.8 %)

7.2 Conclusion on within-laboratory reproducibility as assessed by the study

See 7.1

The OI WLR appears to be similar or lower than what was observed in the other available *in vitro* methods for serious eye damage/eye irritation testing, especially the RhCE-based test methods. A higher variability was obtained than what could be expected from a purely chemically defined test method.

8. Transferability (Module 3)

8.1 Quality of design and analysis of the transfer phase

This was generally well planned and conducted.

InVitro International, the test manufacturer, has well established training procedures, which last 2 days and encompass the conduct of the Ocular Irritection® assay; an explanation of the Ocular Irritection® assay software; and the evaluation, review and discussion of results and troubleshooting. This training was provided by IVI to IIVS and ResPharma before the prospective chemical testing was undertaken.

For the transferability study, four chemicals with different physical properties (two liquids, one solid and one surfactant) having different *in vivo* classifications (two UN GHS / EU CLP Category 2, and two Non-classified chemicals) were used, being tested three separate times using three different Ocular Irritection® kits by the operators.

The transferability of the updated Ocular Irritection® protocol was therefore evaluated in the three laboratories: the first, InVitro International (IVI, California, US) had over 20 years of experience with this class of assay; the second, ResPharma (RP, Italy) also had experience with this class of assay; and the third, the Institute for In vitro Sciences (IIVS, Maryland, US) has considerable experience with in vitro test systems, but no previous experience with this class of assay. The fact that the three laboratories were based on two different continents also allowed information to be gathered on issues that may arise from distribution of the test kits.

The inter-laboratory concordance of classifications obtained during the transferability study was 100 %, the within-laboratory variability was low (SD ≤ 1.7) and the variability between laboratories within an acceptable range (SD ≤ 3.6).

8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study

According to the information provided the OI assay seems to be easily transferable to another laboratory with only general working expertise required for the lab personnel.

However, the ESAC WG noted that the participating laboratories were allowed to freely communicate and meet during the training and transfer phases of the study. Moreover, three of the four chemicals used in training were also used in the transfer study and three of the chemicals used in the training and transfer phase were subsequently included in the prospective validation study. The ESAC WG believes that it would have been preferable to use completely new chemicals instead.

9. Between-laboratory reproducibility (BLR) (Module 4)

9.1 Assessment of reproducibility in different laboratories

As with the WLR, the BLR was evaluated in two ways.

First, when MQSs were available for experiments conducted at all three laboratories with a test chemical (n=45), a mean MQS for the chemical was calculated for each laboratory, with the SD of the three laboratory means being considered as an indicator of BLR. For 38 of the 45 chemicals the SD was < 4. The overall mean (SD: 2.5, CV: 17 %) and median (SD: 2.1; CV: 14 %) obtained between-laboratories was also reported.

In addition, the BLR was evaluated for both cut-offs (i.e., 12.5 for Bottom-Up and 30.0 for Top-Down) in terms of the percentage of concordant predictions among the three laboratories based on the majority laboratory classification. The calculation was performed on the dataset of 56 chemicals tested (31 non-classified, 25 classified as UN GHS/EU CLP Cat 1 (12) or Cat 2 (13). Thirty two were liquids and 24 solids. Again, for the analysis in the TST non-qualified and excluded test results were included in this analysis. Using this approach, 42 (cut-off 12.5) and 43 (cut-off 30.0) of the 56 test chemicals were concordantly classified in all laboratories. Four chemicals were concordantly “excluded” in all laboratories, resulting in a between-laboratory reproducibility of 82.1 % (46/56) for the Bottom-Up approach (cut-off 12.5) and of 83.9 % (47/56) for the Top-Down approach (cut-off 30.0). Results are shown in Table 9.1.1.

Table 9.1.1. Calculation of BLR when considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay.

	Between-laboratory Reproducibility	
	Cut-off 12.5	Cut-off 30.0
concordant predictions	42	43
discordant predictions	10	9
concordant excluded	4	4
concordant NQ	0	0
overall concordant	46	47
% concordance	**82.1 %**	**83.9 %**

As in the case of the WLR, the ESAC WG believes that including “concordant” non-qualified and excluded test results in the analysis is not sound. Taking this into account, the BLR values and their respective Wilson two-sided 95 %-CIs were recalculated by the ESAC WG including only qualified results. According to this approach,, the resulting BLR concordance would be of 84.0 % (42/50) for the Bottom-Up approach, and of 86.0 % (43/50) for the Top-Down approach (Table 9.1.2). See section 9.2.

Table 9.1.2. Calculation of BLR NOT considering consistent occurrence of "non-qualified" and "excluded" runs as concordant in the reproducibility evaluation of the OI assay. The values in brackets correspond to Wilson two-side 95 %-CIs.

	Between-laboratory Reproducibility	
	Cut-off 12.5	Cut-off 30.0
concordant predictions	42	43
discordant predictions	8	7
% concordance	84.0 % (71.5-91.7 %)	86.0 % (73.8-93.1 %)

In addition the ESAC WG notes that three of the chemicals used to the BLR calculation where also used for the training, transfer and proficiency testing.

9.2 Conclusion on between-laboratory reproducibility as assessed by the study

The ESAC WG believes that the reproducibility of the Ocular Irritection® assay is best evaluated using the BLR values not including concordant “non-qualified” and “excluded” results and including only “qualified” results. On the basis of the prospective validation data, using a total of 50 chemicals qualified for BLR, the corrected values would be 84 % (42/50) and 86 % (43/50) for the Bottom-Up and the Top-Down approaches, respectively.

This slightly improves the BLR for both the Top-Down (from 83.9 % to 86 %) and the Bottom-Up (82.1 % to 84 %) approaches.

The WG recognised that three chemicals used in the training and transfer phase were also used to calculate the reproducibility of the test method (BLR).

The ESAC WG noted that the WLR values for the laboratories IIVS (Bottom-Up: 84 % (42/50), Top-Down: 86 % (43/50)) and RP (Bottom-Up: 80.4 % (41/51), Top-Down: 84.3 % (43/51)) (Table 7.1.2 above) are equal or lower than the BLR value, and that the BLR is similar or lower than which is observed in the other available *in vitro* methods for serious eye damage/eye irritation testing, especially the RhCE-based test methods. Although acceptable, a lower variability would have been expected for this assay as compared to other biological assays constituted of living cells or tissues.

10. Predictive capacity and overall relevance (Module 5)

10.1 Adequacy of the assessment of the predictive capacity in view of the purpose

Altogether, data from a total of 88(+1) single chemicals having parallel *in vivo* and *in vitro* data were used to evaluate the predictive capacity of the Ocular Irritection®, including 51(+1) chemicals from the prospective validation study and 37 additional chemicals having existing *in vitro* data generated with the same OI SOP.

Due to the unbalanced number of repetitions between the prospective and the retrospective datasets available, the majority of predictions available for each chemical (mode) was preferred by the VMG as the means of expressing the predictive capacity of the assay. The concordance of these predictions with the expected result as defined by the *in vivo*-based classifications (segregated into UN GHS / EU CLP classified vs. non-classified chemicals, and UN GHS / EU CLP Category 1 vs. non-Category 1 chemicals) were assessed by means of 2x2 contingency tables, and then the specificity, sensitivity and overall accuracy were calculated.

For specificity and sensitivity one-sided 95 %-CIs were calculated using the mid-p approach (Agresti and Gottard, 2005) with the R package 'PropCIs'. This analysis was complemented by a) deriving point estimates of specificity, sensitivity and concordance using an often called 'weighted' approach that considered for each chemical the proportion of correct classification of repeat experiments, and by b) a resampling approach providing probability distributions for specificity and sensitivity, from which the 5 % and 95 %-quantiles were used to describe the expected range of the two parameters (i.e., similar to two-sided 90 %-CIs). Finally, receiver operation characteristics (ROC) curves were produced to more fully characterize the predictive capacity of the Ocular Irritection® assay, and to assess the appropriateness of the MQS cut-offs i) 12.5 to identify GHS non-classified chemicals in a Bottom-Up approach, and ii) 30.0 to identify GHS Category 1 chemicals in a Top-Down approach.

The ESAC is of the opinion that the calculations of Predictive Capacity should reflect in the best way possible the real-life testing situation. Since with the OI assay one single test result will be used to derive one final prediction, the ESAC considers that the majority of predictions is sub-optimal to express the Predictive Capacity of the test method. In ESAC's opinion, more truthful point estimates can be obtained by using the weighted calculation or by resampling of the multiple data generated in the validation study. Both the weighted calculations (with two-sided 90 %-CIs obtained from resampling) and predictive capacity values obtained from the majority of predictions (with Wilson two-sided 95 %-CIs calculated by the ESAC WG) are given in Tables 10.1.1, 10.1.2 and 10.1.3 below for comparison. Published predictive capacity values for other serious eye/damage eye irritation *in vitro* methods are also provided for comparison in Tables 10.1.1 and 10.1.2.

Out of the 56 chemicals tested in the ring study, five chemicals were not included in the analyses of predictive capacity performed due to the fact that they were either excluded (4 chemicals) or not-qualified (1 chemical) by the Ocular Irritection® software in the majority of the available tests/experiments (6 or 9 out of 9). These represent one Cat. 1 chemical, two Cat. 2 chemicals and two Non-Classified chemicals. As a consequence, out of the 56 tested chemicals a total of 51 chemicals were considered in the analyses of the predictive capacity of the OI assay. These represent 20 solids, 27 liquids and 4 viscous chemicals; 29 UN GHS No Category, 11 Category 1 and 11 Category 2 chemicals.

Predictive capacity for the identification of UN GHS / EU CLP Non-classified chemicals

When used for the identification of UN GHS non-classified versus classified materials (based on the existing cut-off of 12.5) the OI assay showed an overall sensitivity of 90.7 % (with two-sided 90 %-CIs: 87.0 % and 93.5 %), a specificity of 58.9 % (with two-sided 90 %-CIs: 53.5 % and 65.1 %) and an overall accuracy of 75.3 %, based on weighted calculations (Table 10.1.1). Some organic functional groups were found possibly to correlate with the observed mispredictions. In particular, acrylate, carboxamide, and cycloalkenes. If, despite their small number, chemicals containing these functional groups were excluded from analyses, the obtained dataset resulted in a sufficiently large dataset (n=79) to still derive sound conclusions. Such findings are comparable to the currently accepted OECD test methods as shown in Table 10.1.1.

Table 10.1.1. Predictive capacity of the OI for the identification of UN GHS/EU CLP non-classified chemicals and comparison to published values for other *in vitro* methods.

	OI n=88 weighted (resampling 90 %-CI)	**OI* n=79 weighted (resampling 90 %-CI)**	**OI n=88 majority of predictions (Wilson 95 %-CI)**	**OI* n=79 majority of predictions (Wilson 95 %-CI)**	**BCOP (TG437) n=196 majority of predictions**	**ICE (TG 438) n=152 majority of predictions**	**STE** (TG 491) n=101**	**EpiOcular EIT (TG 492) n=112 weighted**	**CM*** (Draft TG) n=45 weighted**
Accuracy	75.3 % na	80.5 % na	76.1 % (67/88) (66.3-83.8 %)	81.0 % (64/79) (71.0-88.1 %)	68.9 % (135/196)	82.2 % (125/152)	90.1 % (91/101)	80 % (n=112)	68 % (n=45)
Sensitivity	90.7 % (87.0-93.5 %)	96.4 % (92.9-100 %)	93.3 % (42/45) (82.1-97.7 %)	97.6 % (40/41) (87.4-99.6 %)	100.0 % (107/107)	98.6 % (72/73)	98.1 % (53/54)	96 % (n=57)	100 % (n=22)
Specificity	58.9% (53.5-65.1 %)	62.8 % (55.3-71.1%)	58.1 % (25/43) (43.3-71.6 %)	63.2 % (24/38) (47.3-76.6 %)	31.5 % (28/89)	67.1 % (53/79)	80.9 % (38/47)	63 % (n=55)	32 % (n=23)

* excludes chemicals containing the *acrylate*, *carboxamide* and *cycloalkenes* organic functional groups;
** Only water-soluble chemicals or chemicals forming a uniform suspension, and excluding highly volatile substances and solid substances other than surfactants;
*** Only water-soluble surfactants and surfactant-containing formulations.

Predictive capacity for the identification of UN GHS / EU CLP Category 1 chemicals

For the identification of the UN GHS /EU CLP Category 1 chemicals a post-hoc evaluation of the entire dataset comprising both, the prospective validation dataset and the additional existing data from retrospective studies was carried out. When used for the identification of UN GHS Category 1 versus non-Category 1 chemicals (based on the cut-off of 30.0) the OI assay showed an overall specificity of 80.9 % (with two-sided 90 %-CIs: 76.8 % and 84.1 %), a sensitivity of 53.3 % (with two-sided 90 %-CIs: 50.0 % and 60.0 %), and an accuracy of 74.7 %, based on weighted calculations (Table 10.1.2). These values are compared with those of the currently accepted OECD in vitro test methods for eye hazard assessment in Table 10.1.2 below.

Further investigations conducted by the VMG to better understand possible reasons for misclassification suggested that chemicals having the presence of the organic functional groups carboxylic acid and sulphate seemed to risk possible underpredictions of Cat. 1 chemicals as non-Cat. 1 chemicals. Excluding chemicals having these functional groups from analyses resulted still in a sufficiently large dataset to make sound conclusions (n=74). In this case, a specificity of 80.0 % (48/60), a sensitivity of 71.4 % (10/14) and a concordance of 78.4 % (58/74) were obtained, based on the majority of predictions.

Table 10.1.2. Predictive capacity of the OI for the identification of UN GHS / EU CLP Category 1 chemicals and comparison to published values for other *in vitro* methods.

	OI n=88 weighted (resampling 90 %-CI)	**OI n=88 majority of predictions (Wilson 95 %-CI)**	**BCOP (TG 437) n=191 majority of predictions**	**ICE (TG 438) n=140 majority of predictions**	**FL (TG 460) n=151 weighted**	**STE (TG 491) n=120***	**CM (Draft TG) n=68* weighted**
Accuracy	74.7 % na	73.9 % (65/88) (63.8-81.9 %)	78.5 % (150/191)	85.7 % (120/140)	77.5 % (117/151)	85.0 % (102/120)	88 % (n=68)
Specificity	80.9 % (76.8-84.1 %)	80.9 % (55/68) (70.0-88.5 %)	74.6 % (94/126)	93.8 % (106/113)	93.2 % (96/103)	98.8 % (83/84)	98 % (n=42)
Sensitivity	53.3 % (50.0-60.0 %)	50.0 % (10/20) (29.9-70.1 %)	86.2 % (56/65)	51.9 % (14/27)	43.8 % (21/48)	52.8 % (19/36)	73 % (n=26)

* water-soluble chemicals or chemicals forming a uniform suspension.

Full predictive capacity point estimates as recalculated by EURL ECVAM on the basis of a weighted calculation are summarised in Table 10.1.3.

Table 10.1.3. Evaluation of the OI assay predictive capacity for both Top-Down and Bottom-Up approaches. For each approach weighted calculations are provided.

Predictive Capacity	TOP-DOWN		BOTTOM-UP	
	Prospective data	Combined prospective and retrospective data	Prospective data	Combined prospective and retrospective data
Sensitivity	42.4 % (4.67/11)	53.4 % (10.67/20)	87.4 % (20.11/23)	90.7 % (41.71/46)
False Negatives	57.6 % (6.33/11)	46.7 % (9.33/20)	12.6 % (2.89/23)	9.3 % (4.29/46)
Specificity	80.2 % (32.89/41)	81.0 % (55.89/69)	55.6 % (16.11/29)	59.8 % (25.71/43)
False Positives	19.8 % (8.11/41)	19.0 % (13.11/69)	44.4 % (12.89/29)	40.2 % (17.29/43)
Overall Accuracy	72.2 % (37.56/52)	74.8 % (66.56/89)	69.7 % (36.22/52)	75.8 % (67.42/89)
Total Mispredictions	27.8 % (14.44/52)	25.2 % (22.44/89)	30.3 % (15.78/52)	24.2 % (21.58/89)

10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

In the view of the ESAC WG, the Ocular Irritection® assay shows rather low sensitivity and only moderate specificity for use within a Top-Down Approach. Much like other validated *in vitro* methods for identification of Category 1, the OI assay does not address persistence of chemically-induced ocular injuries and consequently is not able to correctly predict chemicals classified *in vivo* as Category 1 due only to persistence of effects. Therefore, the use of the OI test method as a first step in a Top-Down approach could result in the underprediction of this type of chemicals, which in part explains the poor sensitivity obtained in the validation study.

In addition, if this test method is to be used to identify non-irritants within an Integrated Approach to Testing and Assessment (IATA) then the ESAC believes other test methods may be needed to assess the vascular and inflammatory components of the adverse outcome pathway for eye irritation, and to take account of the lack of epithelial barrier function.

Three chemicals used in the training and transfer phase were also used to calculate the predictive capacity.

11. Applicability domain (Module 6)

11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

The manufacturer's original claim was that the Ocular Irritection® assay is applicable to substances and mixtures, based on over 20 years of experience of testing cosmetics, consumer products, their ingredients and pharmaceuticals

As a result of this long-standing experience, the manufacturer believes that the Ocular Irritection® assay has a well-established and defined applicability domain. It is not applicable to very acidic (pH < 4.0) and very alkaline (pH > 9.0) materials, oils and water-insoluble organic chemicals; and non-ionic surfactants can cause assay interference. The test method also has limitation for the testing of intensely coloured materials generating high OD readings for blanks and samples. In addition volatile ketones have been found to result in under-estimation of irritancy due probably to evaporation. Finally, a number of false negatives (Urea at concentrations > 5 %) and false positives (Sorbitol at concentrations > 5 %, Manganese violet, Aluminum chlorohydrate, Aluminum zirconium chlorohydrate, Aluminum chloride, Titanium oxide, Zinc oxide, Silver salts, Ferrous sulfate, Zinc sulfate) have been identified.

These materials were not tested within the validation study.

In addition to the pre-established applicability domain described above, the validation study showed that for the identification of UN GHS / EU CLP non-classified chemicals, despite the small number, chemicals containing the acrylate, carboxamide or cycloalkene organic functional groups seemed to correlate with mispredictions, where acrylate seemed to correlate with possible under-predictions (1 out of 1, and reported to possibly polymerize under light), cycloalkene with over-predictions (4 out of 5), and carboxamide with both under- and over-prediction (1/2 and 3/3 respectively).

Overall the developer provides a weak rationale for the assay limitations with respect to the applicability of the Ocular Irritection® assay to particular types of chemicals. In some cases the exclusion of chemicals is based on their experience with single or a very limited number of representative chemicals. The WG is not able to decide whether the number of chemicals is large and comprehensive enough to generalize their conclusions.

11.2 Quality of the description of applicability domain, limitations, exclusions

See 11.1 above.

It should be noted that the possible additional limitations in the applicability domain identified within the validation study are based on the testing of a limited number of chemicals. The selection by the VMG of the false negative and false positive threshold values used to decide if a limitation should be defined for a given organic functional group is rather subjective and the final conclusions appear to rely on single or a very small number of substances without showing a solid rationale for the

proposed exclusions. Moreover, the assay is considered as not appropriate for assessing a number of metal salts, what raises concerns about other metal salts not appearing in the list.

The WG recommends providing a sound scientific reason for the exclusion of (groups of) chemicals and the presentation of possible assay limitations so that final users can decide whether the assay is appropriate for testing their substances before wasting the kits.

12. Performance standards (Module 7)

12.1 Adequacy of the proposed Essential Test Method Components

The WG cannot offer a reasoned opinion on this.

In no small part this is due to the fact that the assay matrix has never been chemically characterised, and there is no batch-to-batch chemical analysis to ensure consistency; in addition there is a lack of transparency about the algorithms applied by the software.

The ESAC WG therefore cannot determine what might constitute essential test method requirements for any assay of this class.

12.2 Adequacy of the Reference Chemicals

No list of reference chemicals is supplied in the draft PS. The chemicals used for this validation study would be suitable only for similar test methods with the same precise applicability domain.

12.3 Adequacy of proposed performance target values

No performance target values are proposed in the draft PS. These will depend on the selected reference chemicals.

13. Readiness for standardised use

13.1 Assessment of the readiness <u>for regulatory purposes</u>

The performance of the test method for identifying non-classified chemicals (e.g., as first step in a Bottom-Up approach) appears to be adequate. However, there are a number of uncertainties related with the method that may call into question its acceptance, not least a lack of control over the precise chemical composition and structure of the test matrix. However, after consultation with the manufacturer, they informed the WG that to date all test kits had been produced from of a single batch of plant-extract, and that they still have sufficient in reserve to produce test kits for several more years, although it is not clear to the ESAC WG how stable the plant-extract raw material can be for 25+ years.

13.2 Assessment of the readiness <u>for other uses</u>

The method is already in wide use for non-regulatory testing.

13.3 Critical aspects impacting on standardised use

The main factors are the undefined chemical matrix composition (to date only produced from one batch of the plant-extract), the unknown variability of the method if performed with matrices produced from different bulks of the plant-extract raw material (to date only matrices produced from a single bulk have been evaluated), and a lack of transparency about the details of the software.

13.4 Gap analysis

- Test methods of this class could form components of future integrated testing strategies for determining the eye damage/irritation potential of chemicals (Scott *et al*, 2010). The other components of such a testing strategy and the precise role of these test methods have yet to be formally defined.
- The ESAC WG notes that there is limited information on the performance of this test method with chemical mixtures and no information about gases and aerosols.
- The ESAC WG acknowledges that including a wide range of chemical mixtures in validation studies currently raises several problems, e.g. availability of *in vivo* data, selection of test mixtures, and continuity of supply. However, most of the substances which have to be classified are mixtures and there is a need to confirm that *in vitro* methods can be used for the classification of chemical mixtures. The ESAC WG recommends the inclusion of a broader range of chemical mixtures in future validation studies, and proposes consideration of the use of reference data available for the classification of mixtures, using the additivity approach recommended by the UN GHS (UN, 2015) as well as the CLP (EC, 2008a), and/or the use of mixtures already assessed and identified as Cat.1, Cat.2, or No Cat.
- The chemical matrix is not chemically defined and the stability of the plant-extract raw material over time is also not known as far as the ESAC WG is aware, even though the same bulk is being used for over 25 years. The batch release Quality Assurance undertaken by the manufacturer addresses only test performance not composition. However, taking into account the use of a single batch of plant-extract for manufacturing the kits the WG cannot give any appraisal on its possible effect on study results.
- The ESAC WG understands that to date all test kits have been produced from a single batch of the plant-based raw material; and that this raw material and the final test kit matrix have not been chemically characterised or defined. The ESAC WG also understands that the precise chemical composition of the raw material is likely to vary bath-to-batch. This raises questions about the composition of the test kit matrix, and the performance of the test method, once the current batch of the raw material is exhausted.
- Much like other validated *in vitro* methods for identification of Category 1, the OI does not address persistence of chemically-induced ocular injuries and consequently is not able to correctly predict chemicals classified *in vivo* as Category 1 due only to persistence of effects. Therefore, the use of the OI test method as a first step in a Top-Down approach could result in the underprediction of this type of chemicals.
- If this test method is to be used to identify non-irritants within an Integrated Approach to Testing and Assessment (IATA) then the ESAC believes other test methods may be needed to assess the vascular and inflammatory components of the adverse outcome pathway for eye irritation, and to take account of the lack of epithelial barrier function.

14. Other considerations

The ESAC WG notes that due to the variability of individual animal responses within the *in vivo* Draize eye test there is ≥ 12 % probability, if chemicals are retested, of chemicals currently classified as UN GHS Category 2 by the *in vivo* test being identified as UN GHS No Category (Adriaens *et al*, 2014). As *in vivo* Draize eye test data served as reference data for chemical selection and Predictive Capacity within validation studies, the reported performance of the *in vivo* test should be borne in mind when evaluating the reported performance, and validity, of alternative methods and testing strategies for detecting chemically-induced eye damage/irritation.

15. Conclusions on the study

15.1 ESAC WG summary of the results and conclusions of the study

1. The ESAC WG considers that the performance of the test method appears to be appropriate to identify chemicals not requiring classification for serious eye damage/eye irritation under UN GHS (UN, 2015). However, there are a number of uncertainties related to the method that may call its regulatory acceptance into question.
2. For use in a Top-Down approach, other than having a long shelf-life and short testing time, does not seem to be an advance on other validated test methods for this purpose. Moreover, the test method is not intended to provide insights into persistence of chemically-induced ocular injuries, which could result in the underprediction of chemicals classified *in vivo* as Cat 1 due only to persistence of effects, if the OI test method is used in a Top-Down approach.
3. Beyond the applicability domain identified before the validation study, from the wide range of chemical types, chemical classes, molecular weights, LogP, chemical structures, etc., tested in the validation study, no clear, unambiguous additional limitations regarding applicability apart from those mentioned above were identified.
4. The ESAC WG has noted that currently there is only a limited range of chemical mixtures available for use as test chemicals within eye damage/irritation validation studies, and would like to see more data presented with respect to the test method performance in the case of chemical mixtures requiring the classification for eye damage/irritation potential.

15.2 Extent to which study conclusions are justified by the study results alone

In reaching its conclusions the ESAC WG has also taken account of the larger body of information and knowledge set out in the technical annexes supplied, the references cited in the study documents, and the answers given directly by the manufacturer to questions formulated directly (orally and in written form) by the WG. It is on consideration of both the study findings, and that larger body of knowledge of information, that the ESAC WG established and confirmed the plausibility of the conclusions set out above.

15.3 Extent to which conclusions are plausible in the context of existing information

See 15.2 above.

16. Recommendations

16.1 General recommendations

On the basis of the above, the ESAC WG does not recommend the use of the OI assay for regulatory testing purposes. Nevertheless, considering that (i) the assay is relatively easy and fast to perform, (ii) its performance appears to be acceptable, (iii) it is easily shipped and stored, and (iv) it has a long shelf-life, the ESAC WG considers that the assay may be useful for screening purposes within the applicability domain established in the SOP. Users of the OI assay should nevertheless take into consideration that the applicability domain of the method has been poorly defined, being purely empirical and not biologically justified.

16.2 Specific recommendations (e.g. concerning improvement of SOPs)

The OI test method is currently not intended to provide insights into persistence of chemically-induced ocular injuries. Since this is a current limitation of nearly all validated *in vitro* methods for identification of Category 1, it would be very useful if an OI protocol could be developed to differentiate between persistent and reversible ocular effects.

In addition, if this test method is to be used to identify non-irritants within an Integrated Approach to Testing and Assessment (IATA) then the ESAC WG believes other test methods may be needed to assess the vascular and inflammatory components of the adverse outcome pathway for eye irritation, and to take account of the lack of epithelial barrier function.

The ESAC WG recommends insights into test performance with chemical mixtures be obtained by testing mixtures based on their current classification, or based on the known properties of their components.

The ESAC WG believes that potential problems with the reliability of historical Draize eye test data must also be taken into account when evaluating the predictive capacity of alternative test methods and testing strategies.

The ESAC WG recommends replacing "…and 125 µl/mg sample applied…" by "…and 125 µl (liquids) or mg (solids) sample, applied…" in the two instances this sentence occurs in figure 3.1 of the SOP to improve clarity. Similarly, "µl/mg" in the test method quick work flowchart (page 21 of the SOP) should be replaced by "µl (liquids) or mg (solids)". In the way the volumes or weights of test items that should be applied to the matrix are currently reported in the SOP (as "µl/mg"), the "/" by convention puts the two units in relation to each other and thus could be potentially misunderstood as " µl per mg".

17. References

- Adriaens E, Barroso J, Eskes C, Hoffmann S, McNamee P, Alépée N, Bessou-Touya S, De Smedt A, De Wever B, Pfannenbecker U, Tailhardat M, Zuang V (2014). Retrospective analysis of the Draize test for serious eye damage/eye irritation: importance of understanding the in vivo endpoints under UN GHS/EU CLP for the development and evaluation of in vitro test methods. *Arch Toxicol* 88: 701-723.
- Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, et al. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of the ECVAM Workshop 5. *ATLA* 23: 129-147.
- EC (2008a). Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures (the 'CLP Regulation').
- EC (2008b). COUNCIL REGULATION (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).
- ECETOC (1998). Eye Irritation Reference Chemicals Data Bank. Technical Report n. 48 (2), European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.
- Eskes C (2010). Guidance Document on the Application of Alternative methods in the Regulatory Assessment of Chemical Safety Related to Human Eye Irritation and Severe Irritation: Current Status and Future Prospects. FOPH, 65 pp.
- 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. Toxicol In Vitro 28: 1046-1065.
- Gordon VC (1992). The scientific basis of the EYTEX system. *ATLA* 20: 537–548.
- Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V (2004). A modular approach to the ECVAM principles on test validity. *ATLA* 32: 467-472.
- ICCVAM (2007). Test Method Evaluation Report on In vitro test methods for identifying ocular severe irritants and corrosives. ICCVAM-NICEATM. NIH publication n. 07-4517. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmer.htm
- Kelly CP (1989). EYTEX: an in vitro method of predicting ocular safety. *Pharmacopeial Forum* 15: 4815–5824.
- Loprieno N (1995). Alternative methodologies for the safety evaluation of chemicals in the cosmetic industry. ISBN 0-8493-8546-6. CRC Press, Inc., Boca Raton, US.
- OECD (2002). Acute Eye Irritation/Corrosion. OECD Guideline for the Testing of Chemicals No. 405, OECD, Paris.
- OECD (2005). Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment, Series on Testing and Assessment, No. 34. OECD, Paris.
- OECD (2008). OECD Environment Health and Safety Publications Series on Testing and Assessment No. 88 WORKSHOP ON INTEGRATED APPROACHES TO TESTING AND ASSESSMENT. OECD, Paris.

- Scott L, Eskes C, Hoffmann S, Adriaens E, Alépée N, Bufo M, *et al*. (2010). A proposed eye irritation testing strategy to reduce and replace in vivo studies using Bottom-Up and Top-Down approaches. *Toxicol In Vitro* 24: 1-9.
- UN (2015). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Sixth revised edition, UN New York and Geneva.

LA-NA-28174-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub
EU Science Hub - Joint Research Centre
Joint Research Centre
EU Science Hub

Publications Office

doi:10.2787/991443
ISBN 978-92-79-63196-2