European Commission

ESAC Opinion on the EURL ECVAM Eye Irritation Validation Study (EIVS) on EpiOcular™ EIT and SkinEthic™ HCE and a related Cosmetics Europe study on HPLC/UPLC-spectrophotometry as an alternative endpoint detection system for MTT-formazan

ESAC Opinion No. 2014-03
of 17 November 2014

2014

EUR 28173 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

JRC103702

EUR 28173 EN

PDF ISBN 978-92-79-63195-5 ISSN 1831-9424 doi:10.2787/043697

Luxembourg: Publications Office of the European Union, 2016

How to cite: EURL ECVAM Scientific Advisory Committee (2014). ESAC Opinion on the EURL ECVAM Eye Irritation Validation Study (EIVS) on EpiOcular™ EIT and SkinEthic™ HCE and a related Cosmetics Europe study on HPLC/UPLC-spectrophotometry as an alternative endpoint detection system for MTT-formazan. ESAC Opinion No. 2014-03 of 17 November 2014; EUR 28125 EN; doi:10.2787/043697. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC103702.

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

EURL ECVAM Eye Irritation Validation Study (EIVS) on EpiOcular™ EIT and SkinEthic™ HCE and a related Cosmetics Europe study on HPLC/UPLC-spectrophotometry as an alternative endpoint detection system for MTT-formazan

ESAC Opinion No.	2014-03
Relevant ESAC Request No.	2013-03
Date of Opinion	17/11/2014

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 EURL ECVAM Eye Irritation Validation Study (EIVS) on EpiOcular™ EIT and SkinEthic™ HCE and a related Cosmetics Europe study on HPLC/UPLC-spectrophotometry as an alternative endpoint detection system for MTT-formazan.

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)

Ispra, 17 November 2014

Summary of the ESAC Opinion

The ESAC was requested (see Annex 2) to conduct a scientific peer review of and provide an ESAC opinion on

(1) the EURL ECVAM Eye Irritation Study (**EIVS**); co-sponsored by EURL ECVAM and Cosmetics Europe) evaluating the EpiOcular Eye Irritation Test (EIT) and SkinEthic Human Corneal Epithelium (HCE), both based on Reconstructed human Tissue (RhT); with the ESAC review addressing the study design, results and conclusions. The review examined the scientific basis of the assays, their reproducibility and transferability as assessed during the study, as well as their predictive capacity. The latter was examined in the context of the test methods' envisaged use within an *in vitro* testing strategy (Scott et al., 2010) based on strategic combinations of alternative test methods to replace the Draize eye test, either as the initial step in a Bottom-Up Approach or a second or subsequent step in a Top-Down Approach to discriminate chemicals (substances and chemical mixtures) not requiring classification for serious eye damage/eye irritancy (No Category) from chemicals requiring classification and labelling as Category 1 and Category 2 (UN 2011; EU 2008). The term "strategic combinations" here refers to the placement of methods in a sequential manner in view of either identifying first negatives (high sensitivity: bottom-up approach) or first positives (high specificity: top-down approach).

(2) The usefulness, based on a study conducted by Cosmetics Europe and submitted to EURL ECVAM, of HPLC/UPLC-photometry as an alternative endpoint detection system in particular for highly coloured chemicals interfering with the conventional formazan endpoint measurement by OD-photometry.

EIVS

The ESAC concludes that the study was generally well planned and executed. Briefly, reproducibility, transferability and predictive capacity of both test methods were assessed in a three laboratory ring trial testing 104 chemicals plus an additional 8 solids for assessing the optimised EpiOcular™ EIT protocol. The chemicals represented a wide range of irritancy scores and chemical and use classes. The aims of the study were well described and clear target criteria for test method performance had been predefined. Using appropriate provisions for retesting, the study generated almost complete data sets as a solid basis for evaluating test method performance. However, the ESAC does not agree with the approach chosen by the Validation Management Team (VMT) to calculate the predictive capacity based on the individual test runs: the ESAC believes that confidence limits based on the number of test chemicals are more appropriate. The ESAC believes that the approach chosen by the Validation Management Team (VMT), over-estimated the sample size, produced overly narrow confidence limits, and resulted in a seemingly high degree of certainty with respect to the point estimates of sensitivity and specificity. In addition rather than relying only on single test runs to calculate accuracy, the ESAC believes that, in case single runs are used for analysing predictive capacity, reliance is better placed on bootstrapping or worst-case/best-case scenarios of predictive capacity.

EpiOcular™ EIT

The EpiOcular™ EIT test method was found to be highly reproducible within laboratories, irrespective of the protocol used. The ESAC considers that the VMG predictive capacity criteria are deemed to have been met with the 60% cut-off value using protocol V8.0. This version was the result of optimisation phase during the validation study aiming to improve results for solids while maintaining the accuracy obtained for liquid test materials.

The ESAC concludes that the EIVS study findings with respect to transferability, reproducibility and predictive capacity, taking account of the supplementary analysis undertaken by the ESAC WG, justify the EpiOcular™ EIT test method (SOP V 8.0) being considered for use within a test strategy to determine the eye irritation potential of chemicals, specifically to detect non-irritants as part of a Top-Down or Bottom-Up approach (Scott et al, 2010).

However, the ESAC believes that confidence in the test method would be further increased if supplementary studies confirm that the test method correctly classifies a small but representative sample of labelled products or mixtures as defined by REACH and from various sectorial use classes: for products containing mixtures of substances, such as plant protection products or cosmetic products, the testing of the mixture is considered more representative of the expected exposure situation in actual practice than testing its constituent substances separately.

SkinEthic™ HCE

The ESAC agrees with the VMG's view that this test method does not meet the performance criteria that would be required to consider it as part of a test strategy. ESAC also agrees that with the VMG that there is potential for improvement but that the necessary work is beyond the scope of the validation study.

The Cosmetics Europe Report

The Cosmetics Europe Report is not a Validation Study Report (VSR) relating to a predictive alternative test method, but is a report of a study which evaluated the performance of a bio-analytical endpoint measurement system that overcomes a limitation of the conventional OD-photometry measurements.

The ESAC believes the evidence presented tends to confirm that for chemicals compatible with these RhT methods, the measurements obtained of formazan values by OD-photometry and the suitably qualified HPLC/UPLC endpoint detection systems are comparable. ESAC appreciates that an advantage of the HPLC/UPLC endpoint detection measurement is that it can also be used in the case of strongly coloured chemicals incompatible with formazan measurement by OD-photometry (e.g. colorants used in cosmetics and other consumer products). The report also provides plausible arguments that this endpoint measurement system is suitable for use with other RhT test methods relying on measurement of the same endpoint.

The ESAC concludes that the use of HPLC/UPLC would overcome the current limitation of RhT test methods with respect to problems encountered with strongly-coloured chemicals which interfere with the measurement of formazan levels by OD-photometry.

1. Mandate of the ESAC

The 38th ESAC plenary (19 June 2013) established an ESAC Working Group (WG)[1] to undertake a detailed scientific review of an EURL ECVAM/Cosmetics Europe Eye Irritation Validation Study (EIVS) evaluating two *in vitro* test methods based on Reconstructed human Tissue (RhT) models for assessing acute eye irritation caused by chemicals: the EpiOcular™ EIT and the SkinEthic™ HCE.

A revised EURL ECVAM Request for ESAC advice (Annex 2) was adopted at the 39th ESAC plenary (11/12 March 2014) extending the ESAC mandate to include a scientific review of a Cosmetics Europe study of HPLC/UPLC-photometry as an alternative formazan endpoint measurement system.

2. Detailed Opinion of the ESAC

Taking into account (a) the report of the detailed review undertaken by the ESAC WG; (b) the information made available to ESAC by EURL ECVAM (Including the EIVS VSR, the Cosmetics Europe report, and other supporting documentation), and; (c) the ECVAM request for ESAC advice (Annex 2) - the ESAC has adopted the following opinion.

2.1 Background, regulatory and scientific rationale

Both the EIVS and the Cosmetics Europe reports describe the potential role of the RhT-based test methods, and the endpoint measurement system, in the context of the regulatory testing.

EIVS

The ESAC considers that the VSR provides plausible arguments supporting the use of these RhT test methods, when combined with other *in vitro* methods, within a tiered approach using strategic combinations of alternative test methods. This tiered approach (Scott et al, 2010) is supposed to be integrated into future more complete "Integrated Testing Strategies" (ITS) or "Integrated Approaches to Testing and Assessement" (IATA) to replace the Draize eye test. The tiered approach is argued to have the potential to be at least as relevant and reliable as the current *in vivo* test system for the identification of non-irritants, either as the initial step in a Bottom-Up Approach or a second or subsequent step in a Top-Down Approach test to discriminate chemicals not requiring classification for serious eye damage/eye irritancy (No Category) from chemicals requiring classification and labelling (UN GHS Category 1 and Category 2).

Taking account of the probably rather small prevalence of chemicals inducing serious eye damage (Adriaens et al, 2014) RhT test methods validated for this purpose would significantly reduce animal testing by identifying the large number of chemicals not requiring classification.

[1] The WG was established according to the ESAC Rules of Procedure and included experts nominated by international partner organisations of EURL ECVAM collaborating under the ICATM (International Collaboration on Alternative Test Methods) framework.

SkinEthic™ HCE

The tissue component of the test kit is manufactured by controlled culture of immortalised human corneal epithelial cells to produce a multi-layered epithelium resembling *in vivo* corneal epithelium. Although derived from cultured human corneal cells the cells used have been transformed and immortalised. The ESAC offers no opinion on whether these cells are inherently more biologically relevant in the context of eye irritation testing than other epithelial cells (e.g. RhT models derived from normal untransformed human keratinocytes).

EpiOcular™ EIT

The EpiOcular™ EIT tissue construct is prepared from normal human-derived epidermal keratinocytes grown under defined conditions. The resulting 3D tissue is a non-keratinized multi-layered and stratified (but non-cornified) epithelium intended to model properties of human corneal epithelium.

The Cosmetics Europe Report

The Cosmetics Europe Report is not a Validation Study Report (VSR): the study evaluated the performance of a bio-analytical endpoint measurement system (which is not itself a test method).

The Scientific Committee on Consumer Safety (SCCS) of the European Commission Directorate General for Health and Consumers previously advised that *"...for coloured substances, a different endpoint, not involving Optical Density (OD) quantification, should be envisaged. Analytical methods such as HPLC/UPLC might be more appropriate to detect formazan in the in vitro assay..."* (McNamee *et al*. 2009). An alternative endpoint detection measurement for RhT/MTT reduction product test methods that overcomes this current limitation would extend the applicability domain of these RhT tests.

The Cosmetics Europe study proposes the use of suitably qualified HPLC/UPLC endpoint measurement in this context. The ESAC believes The Cosmetics Europe Report describes a bio-analytical system that has the potential to overcome a current limitation on the use of these and other RhT test methods.

2.2 Design and conduct of the study

2.2.1 Definition of the study objectives

EIVS

The ESAC believes these are clearly defined.

The Cosmetics Europe Report

The ESAC considers the study objectives to have been sufficiently clearly defined.

2.2.2 Study design for the prospective part

EIVS

The ESAC concludes that, subject to specific qualifications set out elsewhere in this Opinion, the study design was generally appropriate and sufficiently robust with respect

to determining the relevance and reliability of these test methods for the purposes defined in the VSR.

SkinEthic™ HCE

The protocol for the SkinEthic™ HCE test method described the following testing strategy**:**

(1) The first step was an "Eye irritation Peptide Reactivity Assay" (EPRA).
(2) For EPRA reactive chemicals the SkinEthic™ HCE Short-time Exposure protocol (HCE SE).
(3) For EPRA non-reactive chemicals the SkinEthic™ HCE Long-time Exposure protocol (HCE LE).

However, for the purposes of the EIVS study all chemicals were tested using both the HCE SE and HCE LE protocols.

104 coded chemicals were blind-tested using both the SE and LE protocols in three runs with three replicate tissues per run, in three laboratories. A test strategy combining, in a sequential manner, the Eye Irritation Peptide Reactivity Assay (EPRA) with both the SE and LE variants of the SkinEthic™ HCE test method was also assessed.

EpiOcular™ EIT

In the ring trial 104 coded chemicals were blind-tested (three runs with two replicate tissues per run, in three laboratories) using EpiOcular™ EIT SOP (V 6.0), using separate protocols for liquid and solid chemicals, and two prediction models (50% and 60% cell viability).

With the original EpiOcular™ EIT protocol (SOP V6.0) all of the VMG acceptance criteria were met for liquid chemicals: however, some of the VMG predictive capacity acceptance criteria were not met for solid chemicals.

The VMG considered that test method performance might be improved if a. Test optimisation was undertaken by the test developer within EIVS.

To improve the performance of the test methods with solid chemicals, test optimisation was undertaken within EIVS to produce a balanced increase in sensitivity offset by some decrease in specificity using 11 of the most challenging EIVS solid chemicals. An amended optimised for solid chemicals protocol, V 8.0, was used for a post-optimisation validation study of the EpiOcular™ EIT optimised in one laboratory with 60 solid chemicals. The post-optimisation findings are incorporated and analysed in the VSR, and form the basis of the VMG conclusions on the performance of the EpiOcular™ EIT.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC believes the Cosmetics Europe study was well-designed, and is satisfied that a sufficient number of RhT MTT-reduction product test methods, chemicals, and laboratories were used.

In the first phase of the study, three laboratories qualified HPLC/UPLC systems using appropriate qualification criteria in accordance with the May 2001 US Food and Drug Administration (FDA) guidance for industry "Bioanalytical Method Validation".

Next, three laboratories tested up to 26 chemicals using one of three RhT test methods measuring three different toxicological endpoints (EpiOcular™ EIT as per the EIVS protocol v6.0 for eye irritation, SkinEthic™ RhE as per OECD TG 431 for skin corrosion, and the EpiSkin™ as per OECD TG 439 for skin irritation). The three laboratories with the suitably qualified HPLC/UPLC systems for tested the samples and applied the Prediction Models using both endpoint measurements.

Chemical selection:

EIVS

The ESAC believes the method by which the test chemicals were selected, and the number and nature of the test chemicals, were appropriate.

The Cosmetics Europe Report

The ESAC believes that the 26 chemicals selected were adequate for this study, and represent the almost the full dynamic range of the relevant toxicological endpoints.

2.2.3 Study design for the retrospective part

Not applicable.

2.2.4 Statistical analysis used in both part of the study

The ESAC believes that whist the data management systems for both studies were robust, elements of the statistical analysis were not best practice.

EIVS

Almost complete datasets are available for both RhT test methods, but ESAC has some concerns about some of the statistical methods applied - see 2.3.3 below.

The Cosmetics Europe Report

The ESAC believes that all of the reference and new data relied upon are of sufficient quality.

However, the statistical method by which the formazan measurements by OD-photometry and HPLC/UPLC were correlated in the study report is not ideally suited to situations where the true value of the variable of interest in unknown, and in producing this Opinion the ESAC has taken account of an alternative approach developed for use in these situations (Bland and Altman, 1986).

2.3 Study results and conclusions

2.3.1 Standardised use

EIVS

The ESAC considers that the SOPs used in the EIVS validation study are sufficiently detailed and complete, with SOP V 8.0 being considered the definitive protocol for EpiOcular™ EIT.

The Cosmetics Europe Report

The ESAC considers the qualification process and parameters described for suitably qualified HPLC/UPLC to be appropriate and sufficiently clear.

2.3.2 Within- and between laboratory reproducibility

Within-Laboratory Reproducibility (WLR)

EIVS

Both RhT test methods proved to be highly reproducible within laboratories, satisfying the acceptance WLR criteria established by the VMG.

EpiOcular™ EIT

The possible short-comings of the VMG WLR calculations are described elsewhere in this Opinion. The ESAC WG, having re-analysed the data set on the basis of the number of chemicals tested estimated the 95% Confidence Interval (95%-CI) for the WLR using the 60% cut off to be 89.0% to 94.4%, satisfying the VMG acceptance criteria (WLR ≥ 85%).

The WLR of the EpiOcular™ EIT optimised for solid chemicals protocol (SOP V8.0) at the one laboratory that conducted the supplementary testing satisfied the VMG acceptance criteria.

The Cosmetics Europe Task Force Eye Irritation Report

During the Phase 1 (qualification) reproducibility was addressed with both intra- and inter-day evaluations of the formazan calibration reproducibility calculated by regression analysis. Whilst the ESAC accepts that variations on this method of assessing agreement between two test methods are commonly used, the use of correlation for this purpose can be misleading and make it difficult to recognise differences, and the significance of the differences, between the findings with different measurement systems when the true value of the variable of interest is unknown. The ESAC takes account of an alternative approach (Bland and Altman, 1986), developed for use in these situations, which also confirms the limits of agreement are good.

Within laboratory reproducibility (WLR) of formazan measurement by HPLC/UPLC was evaluated in one laboratory: the results obtained were 100% concordant.

Between-Laboratory Reproducibility (BLR)

EIVS

The ESAC notes that all of the BLR values reported in the VSR are significantly above the acceptance criteria set by the VMG (BLR ≥ 80%), but that there was no BLR study conducted using the EpiOcular™ EIT optimised protocol for solids (SOP V 8.0).

The Cosmetics Europe Task Force Eye Irritation Report

There was 100% concordance between OD-photometry and HPLC/UPLC formazan endpoint measurements for non-interfering chemicals.

2.3.3 Predictive Capacity

EIVS

The ESAC believes that the VMG's determination of confidence limits for the predictive capacity based on the number of test runs, rather than the number of chemicals tested, produces confidence limits which may be inappropriately narrow, and has taken account revised calculations based on the number of chemicals tested.

Furthermore, the ESAC considers that accuracy figures should reflect the way decisions will be made in practice. Specifically, as chemical classification will be based upon one qualifying test run, the study results should have been evaluated to produce an accuracy range determined by the best- and worst-cases based on data from single runs rather than aggregating data from different test runs with the same chemical. Additional analysis was undertaken along these lines.

SkinEthic™ HCE

The SkinEthic™ HCE, SE and LE protocols, failed to meet the 'definitely acceptable' criteria for sensitivity and overall accuracy defined by the VMG (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%).

The SkinEthic™ HCE sensitivity, as judged by the VMG criteria (specificity < 50%; sensitivity < 80%, overall accuracy < 65%), was 'definitely unacceptable' [2].

EpiOcular™ EIT

ESAC believes that with SOP V 8.0 and a 60% cut off all of the VMG acceptance criteria can be deemed to have been met.

When used for liquid chemicals, the original EpiOcular™ EIT SOP V6.0 protocol the VSR reported results satisfying all of the VMG predictive capacity acceptance criteria. A 60% cut-off resulted in a better sensitivity. For solid chemicals the results with protocol version (V6.0) did not satisfy all of the VMG acceptance criteria: six chemicals were under-predicted, one of which (misclassified by two of the laboratories) is classified *in vivo* as Category 1.

When the solid chemicals were tested in one laboratory by the protocol optimised for solid chemicals (SOP V 8.0) and a 60% cut-off all of the VMT definitely acceptable values were achieved.

The ESAC notes that with respect to accuracy using the best-case/worse-case calculation model even the worst-case values satisfy VMG criteria for sensitivity and accuracy, and that the worst-case specificity is very close to the VMG acceptance criterion. As the ESAC considers that the true predictive capacity falls somewhere between the best- and worst-case calculations based on all test runs, the VMG predictive capacity criteria are deemed to have been met with the 60% cut-off value.

The Cosmetics Europe Task Force Eye Irritation Report

Where comparisons can be made between HPLC/UPLC data and *in vivo* classification data, the HPLC/UPLC data did not under-estimate the *in vivo* classification. In case of EpiOcular™ EIT data, for test substances that do not exhibit colour interference or direct MTT reduction, the tissue viability values obtained are almost identical whether using absorbance (OD-photometry) or HPLC/UPLC, with the OD-photometry values being consistently, marginally above the values obtained with the HPLC/UPLC test method.

[2] The VMG defined a three tier system for judging the acceptability of performance of the assays by using a three category system – "definitely acceptable" and "definitely unacceptable", with the values between these figures being borderline acceptable. The same approach had been chosen by EURL ECVAM for validating the skin corrosion assays (Fentem et al., 1998).

2.3.4 Applicability and possible limitations

EIVS

The EpiOcular™ EIT and SkinEthic™ HCE test methods were not developed, designed, or evaluated to differentiate between UN GHS / EU CLP Category 1 (serious eye damage) and Category 2 (eye irritation) classifications. There is limited information available on the performance of the test methods with chemical mixtures.

EpiOcular™ EIT

The ESAC believes that the available evidence on the **relevance** and **reliability** of the EpiOcular™ EIT supports its consideration for use within an integrated, tiered approach using strategic combinations of alternative test methods to replace the Draize eye test, either as the initial step in a Bottom-Up Approach or a second or subsequent step in a Top-Down Approach, to discriminate chemicals (substances and chemical mixtures) 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) and as implemented by the EU CLP regulation (EU CLP).

With respect to the chemicals that were misclassified when the EpiOcular™ EIT findings were judged against the Draize eye test reference data, the ESAC has identified no specific additional limitations or exclusions.

SkinEthic™ HCE

The ESAC believes that the reported predictive capacity of the SkinEthic™ HCE test method does not justify it being considered for regulatory use.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC believes that the evidence presented suggests that for chemicals compatible with the three RhT methods used within this study comparable classifications are obtained using both OD-photometry and HPLC/UPLC endpoint detection systems; that with these RhT test methods HPLC/UPLC endpoint detection measurement can be used in the case of strongly coloured chemicals which are incompatible with formazan measurement by OD-photometry; and that there are plausible and reasoned arguments for why this endpoint measurement system might be applicable to other RhT test methods relying on measurement of the same endpoint.

2.3.5 Identified gaps between study design and study conclusions

EIVS

On the basis of the EIVS data the VMG concluded that the use of 2 tissue replicates in similar or modified RhT/MTT-based test method aiming at identifying chemicals not requiring classification for serious eye damage/eye irritation is statistically and scientifically justified. However, the ESAC does not accept that this can be generalised to all RhT test methods on the basis of the available evidence.

The Cosmetics Europe Task Force Eye Irritation Report

Nil of note.

2.4 Potential regulatory use of the test method

EIVS

EpiOcular™ EIT

The ESAC believes that the available evidence on the **relevance** and **reliability** of the EpiOcular™ Eye Irritation Test (EIT) supports its consideration for use within an integrated, tiered approach using strategic combinations of alternative test methods to replace the Draize eye test, either as the initial step in a Bottom-Up Approach or a second or subsequent step in a Top-Down Approach, to discriminate chemicals (substances and chemical mixtures) 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) and as implemented by the EU CLP regulation (EU CLP).

SkinEthic™ HCE

The ESAC believes that the predictive capacity of the SkinEthic™ HCE test method reported in the EIVS VSR does not currently justify it being considered for regulatory use.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC judges that the Cosmetics Europe Report provides evidence, and sets out plausible and reasoned arguments, for in the case of regulatory testing this endpoint measurement system being considered for RhT test methods relying on the quantification of formazan.

2.5 Recommendations

EIVS

SkinEthic™ HCE

ESAC agrees with the VMG and the test method developer's decision to not advance the SkinEthic™ HCE test method in this validation on the basis of insufficient predictive capacity (in particular specificity) for considering the test method, at present, for a bottom up test strategy within a tiered approach for inclusion into an Integrated Testing Strategy (ITS) / Integrated Approach to Testing and Assessment (IATA). ESAC understands that the test method developer is currently optimising the test method protocol before re-entering the validation process. ESAC supports the decision for optimisation, including considering developing and evaluating different protocols for liquid chemicals and solid chemicals.

EpiOcular™ EIT

1. The ESAC recommends that the EpiOcular™ EIT test method SOP V 8.0 using the 60% cut off be considered for use within an integrated, tiered approach using strategic combinations of alternative test methods to replace the Draize eye test, either as the initial step in a Bottom-Up Approach or a second or subsequent step in a Top-Down Approach, to discriminate chemicals (substances and chemical mixtures) 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) and as implemented by the EU CLP regulation (EU CLP).

2. Based on the data generated during validation, the ESAC recommends that EpiOcular™ EIT be considered applicable to a wide range of liquid and solid chemicals. However, confidence in the ability of the EpiOcular™ EIT test method to detect solid irritants (using SOP version 8.0) would be increased if more chemicals were tested under blind conditions by the other two laboratories.
3. The ESAC recommends that the SOP should be amended to better define the protocol to be used for test chemicals with an unclear physical state, specifically that all viscous, waxy and gel-like chemicals are placed in a water bath for 15 minutes at 37°C before deciding if they should be tested with the liquids or the solids protocol; and that the test chemical should be applied directly from the water bath and should not be brought to room temperature before testing.

General

4. The ESAC recommends that a range of labelled chemical mixtures be identified for use as test substances within future eye irritation validation studies.

The Cosmetics Europe Task Force Eye Irritation Report

5. The ESAC recommends that the use of suitably qualified HPLC/UPLC systems for the detection and quantification of formazan be proposed for inclusion in relevant regulatory guidelines to extend the applicability domain of RhT test methods to include coloured chemicals interfering with formazan measurement by OD-photometry. However, the HPLC/UPLC provided marginally but systematically lower values for formazan levels than OD photometry. This may impact on some of the prediction models which currently rely on formazan quantification by OD-photometry

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. 2014 Mar;88(3):701-23.
- Bland JM, Altman DG. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, i, 307-310.
- Fentem JH, Archer GE, Balls M, Botham PA, Curren RD, Earl LK, Esdaile DJ, Holzhütter HG, Liebsch M. (1998) The ECVAM International Validation Study on In Vitro Tests for Skin Corrosivity. 2. Results and Evaluation by the Management Team. Toxicol In Vitro, 12(4):483-524.
- Food and Drug Administration (2001) Guidance for Industry: Bioanalytical Method Validation. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research.
- Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A., Zuang V. (2004). A modular approach to the ECVAM principles on test validity. Altern. Lab. Anim. 32, 467-472.
- Mc Namee P, Faller IC, Goebel C, Pfuhler S, Tieber T, Egham I. Use of HPLC/UPLC instead of photometry for evaluation of MTT in in vitro RhT assays for irritation; assessment of colored materials, in Altex, volume 26, special issue, Abstracts 7th World Congress, Rome, 2009, p281.
- Nguyen DH, Beuerman RW, DeWever B, Rosdy M: Three-dimensional construct of the human corneal epithelium for in vitro toxicology. In: Alternative toxicological methods, 1st ed., chapter 14, pp.147-159, Ed. Salem H and Katz SA, CRC, New York, 2003
- 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.
- OECD (2013), In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris
- OECD (2013). Test Guideline No. 439: In Vitro Skin Irritation – Reconstructed Human Epidermis Test Method.
- SCCS/1392/10: Memorandum (addendum) on the in vitro test EPISKIN™ for skin irritation testing, adopted by the SCCS during the 9th plenary meeting of 14 December 2010.
- 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 (2009), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Third 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. Edward CARNEY
- 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. Jon RICHMOND (ESAC Member, WG Chair)
- Dr. Kristina KEJLOVÁ (ESAC Member)
- Prof. Annette KOPP-SCHNEIDER (ESAC Member)
- Dr. Renate KRÄTKE (ESAC Member)
- Dr. Tohru MARUNOUCHI (Fujita-Health University; ICATM nomination by JaCVAM)
- Dr. Jill MERRILL (US FDA; ICATM nomination by NICEATM/ICCVAM)

*) One member of the ESAC WG (JM) had been the ICCVAM liaison to the EIVS VMG during the validation study. She had not been a member of the core Validation Management Group (VMG). Although all decisions during the validation study were taken by the core VMG, JM advised and commented on some elements of the study design. The ESAC WG does not believe that this liaison role constitutes a significant conflict of interest with respect to the work of the ESAC WG.

EURL ECVAM (Secretariat)

- Dr. Claudius GRIESINGER (ESAC Coordinator)
- Dr. João BARROSO
- Dr. Michael SCHÄFFER
- 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 2013-03

EURL ECVAM Scientific Advisory Committee

(ESAC)

EURL ECVAM REQUEST FOR ESAC ADVICE

on the

EURL ECVAM-coordinated Eye Irritation Validation Study (EIVS)

Title page information	
Abbreviated title of ESAC request	EURL ECVAM Eye Irritation Validation Study
ESAC REQUEST No.	2013-03
Template used for preparing request	EP 2.01
Date of finalising request	10/06/2013; updated 18/2/2014: inclusion of review mandate: review objective &questions as well as list of documents to be made available to ESAC
Date of submitting request to ESAC	1) Preliminary request: ESAC38, 18/19 June 2013 2) Final request: 14/3/2014
Request discussed through	ESAC38 (preliminary); ESAC 39 (final, including review mandate)
Opinion expected at (date)	~~ESAC meeting 1Q 2014~~ **Amended: through written procedure 2Q 2014.**
File name of this request	

TABLE OF CONTENTS

1. TYPE OF REQUEST

Request Type	Identify request ("YES")
R1 ESAC Peer Review **of a Prevalidation Study or Validation Study**	**YES**
If R1)applies please specify further:	
►Prevalidation Study	**NO**
►Prospective Validation Study	**YES** In December 2008, EURL ECVAM, together with Cosmetics Europe, initiated a **prospective validation study (=Eye Irritation Validation Study, EIVS)** on two *in vitro* test methods for acute eye irritation caused by chemicals (substances and chemical mixtures). Both test methods, the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE) are based on Reconstructed human Tissue (RhT). The goal of the study was to assess their **relevance** (predictive capacity) and **reliability** (transferability; within and between laboratory reproducibility). Both test methods did not meet all of the acceptance criteria set by the VMG. **For SkinEthic™ HCE none of the protocols under validation (SE and LE) nor their combination in a test strategy with the Eye irritation Peptide Reactivity Assay (EPRA) was able to meet all of the acceptance criteria.** For the EpiOcular™ EIT some difficulties were encountered with regard to meeting the acceptance criteria when using the protocol for **solid chemicals, while the liquids protocol was able to meet all of these criteria**. As there was scope for improvement, the VMG allowed optimization within the EIVS. The study on the optimised EpiOcular™ EIT solids protocol has been completed in 2013. In contrast, the optimisation work on the SkinEthic™ HCE protocol will take longer and will thus be carried out outside the EIVS. It is foreseen however that this study concerning the SkinEthic™ HCE will be forwarded to ESAC review at a later stage.
►Retrospective Validation Study	**NO**
►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

Prospective validation study on two *in vitro* test methods using Reconstructed human Tissue (RhT) for the detection of acute eye irritation caused by chemicals: EpiOcular™ Eye Irritation Test (EIT) and SkinEthic™ Human Reconstructed Corneal Epithelium (HCE)

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

1) Background of Serious Eye Damage/Eye Irritation and current testing strategies

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, 2013), *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. Currently, serious eye damage/eye irritation is determined through in vivo and in vitro assays. The traditional Draize eye test uses rabbits as model for ocular irritation (OECD Test Guideline 405; OECD, 2012). 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) and on cell-based methods (Fluorescein Leakage, Cytosensor Microphysiometer and Short Time Exposure), adopted in 2012 (TG 460) or currently undergoing regulatory acceptance (draft TGs on the Cytosensor Microphysiometer and the Short Time Exposure assays).

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 in its capacity to predict ocular irritation for the full range of irritancy and for a broad spectrum of chemical classes (wide applicability domain). However, strategic 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 either chemicals inducing serious eye damage (GHS 'Category 1') or chemicals not requiring classification for serious eye damage/eye irritation ("non-irritants") (GHS 'No Category'). According to this framework, the entire range of irritancy may be resolved by arranging tests in a tiered (sequential) strategy that may be operated bidirectional, i.e. from either end: to detect first eye damage and resolve absence of irritancy ("Top-Down Approach") or to proceed inversely, starting with the identification of non-irritants first ("Bottom-Up Approach"). Ocular irritant chemicals (GHS cat. 2) will be resolved in a last tier in both approaches. When considering the tiered *in vitro* strategy described above, it should be kept in mind that ocular toxicity is already and will be evaluated in the future on the basis of integrated approaches making use also of *inter alia*, physicochemical properties, (Q)SARs and empirical testing tools. The tiered testing strategy suggested by Scott et al. (2010) should be seen as an empirical module within such an integrated approach. For example, the ECHA *Guidance on information requirements and Chemical Safety Assessment* provides guidance on the integration of testing and non-testing data for the assessment of eye damage/irritation and this guidance is

regularly updated (last version: ECHA, 2014).

2) Background of the EpiOcular™ Test method

The EpiOcular™ tissue construct is a non-keratinized multi-layered and stratified (but not cornified) epithelium prepared from normal human-derived epidermal keratinocytes. It is intended to model properties of the corneal epithelium. The reconstructed tissue is prepared in inserts with a porous membrane (MTI-003) through which the nutrients pass to the cells. A cell suspension is seeded into the MTI-003 membrane in specialized medium. After a period of initial cell proliferation, the medium is removed from the top of the tissue so that the epithelial surface is in direct contact with the air. This allows the test chemical to be directly applied to the epithelial surface, mimicking exposure of the corneal epithelium *in vivo / in situ*. This topical exposure is essential for modelling the same kind of progressive injury expected *in vivo*, allowing moreover the application of both solid and liquid chemicals. In the EpiOcular™ EIT, liquids and solids are applied using different exposure and post-exposure incubation times (i.e., liquids: 30 min exposure followed by 120 min post-treatment incubation, and solids: 90 min exposure followed by 18 hours post-treatment incubation- the optimized protocol for solids requires now an exposure time of 6 hrs. Post-treatment time remains unchanged.).

2) Background of the SkinEthic™ HCE Test method

To construct SkinEthic™ HCE tissues, immortalized human corneal epithelial cells are cultured in a chemically defined medium and seeded on a polycarbonate membrane at the air–liquid interface. The tissue construct obtained is a multi-layered epithelium resembling the *in vivo* corneal epithelium. As *in vivo*, columnar basal cells are present, including Wing cells. The model is characterized by the presence of ultra-structural features such as intermediate filaments, mature hemi-desmosomes and desmosomes that characterize the relevant epithelium *in situ*. Specific cytokeratins 64kD (K.3) have also been described (Nguyen D.H. *et al.*, 2003).

The SkinEthic™ HCE test method, when submitted to EURL EVCAM, comprised a small **test strategy** based on **three separate assays:**
(1) the "Eye irritation Peptide Reactivity Assay" (EPRA) followed by
(2) either the SkinEthic™ HCE Short-time Exposure protocol (HCE SE) or
(3) the SkinEthic™ Long-time Exposure protocol (HCE LE),
depending on the reactivity measured in the initial EPRA (method (1) of the strategy).
The HCE SE uses 10 min exposure without post-treatment incubation, while the HCE LE is based on 1 h exposure followed by 16 h post-treatment incubation. Following treatment with a test chemical as described above (using EpiOcular™ EIT, SkinEthic™ HCE SE or HCE LE), the relative tissue viability is determined against the negative control-treated constructs by the reduction of the vital dye MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-111 diphenyltetrazolium bromide). Tissues treated with chemicals inducing eye irritation or serious eye damage (UN GHS/EU CLP Category 2 or Category 1, respectively) are expected to show a decrease in viability below a certain threshold in respect to the negative control.

Study objectives and design

In December 2008, EURL ECVAM, together with Cosmetics Europe, initiated a prospective validation study on two *in vitro* test methods, using Reconstructed human Tissue (RhT) test systems (EpiOcular™ Eye Irritation Test (EIT) and SkinEthic™ Human Reconstructed Corneal Epithelium (HCE)) for the detection of acute eye irritation caused by chemicals (substances and chemical mixtures).

The goal of the present study was to assess the relevance (predictive capacity) and reliability (transferability and reproducibility within and between laboratories) of the EpiOcular™ EIT and

SkinEthic™ HCE test methods. If found valid, both test methods may be incorporated into a tiered test strategy (so-called Bottom-Up/Top-Down test strategy, as defined in an EURL ECVAM workshop held in 2005, Scott *et al.*, 2010) as e.g. the **initial step in a Bottom-Up approach or the second step in a Top-Down Approach.** The test methods are not intended to differentiate between GHS Category 1 (irreversible effects) and 2A-B (reversible effects). This differentiation would be left to another tier of the Bottom-up/Top-down testing strategy (Scott et al., 2010).
The ultimate purpose of a tiered testing strategy comprising these and other *in vitro* assays (e.g. BCOP, ICE, FL, CM) is to replace the *traditional in vivo* Draize eye irritation test [Method B.5 of EC Regulation 440/2008 (EC, 2008a) or OECD TG 405 (OECD, 2012)] as an empirical testing tool to generate information on the eye irritation potential of chemicals.

More specifically, EIVS aimed to assess the validity of the SkinEthic™ HCE SE, LE and test strategy and of the EpiOcular™ EIT as stand-alone (independent) test methods to reliably discriminate chemicals not classified for serious eye damage/eye irritation ("non-irritants") from chemicals requiring classification (in the framework of a Bottom-Up/Top-Down test strategy, Scott L. *et al.,* 2010), defined according to the United Nations Globally Harmonized System (GHS) of Classification and Labelling of Chemicals (UN GHS: No Category versus Category 1/Category 2A/Category 2B; UN, 2013) and as implemented in the European Commission Regulation (EC) No 1272/2008 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 (EU CLP: No Category versus Category 1/Category 2).

Relevance and reliability were assessed by testing, in a ring trial, a statistically significant number of coded test chemicals (substances and mixtures), supported by complete and quality assured *in vivo* Draize eye irritation data for comparative evaluation of results.

Study results

The study results were analyzed by an independent statistician (Carina Rubingh, Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO, i.e. Dutch Organization for Applied Scientific Research).

SkinEthic™ HCE
Two different protocols of SkinEthic™ HCE, a short-time exposure protocol (SE), and a long-time exposure protocol (LE) where fully assessed in EIVS. A test strategy combining SE and LE with the Eye irritation Peptide Reactivity Assay (EPRA) was also assessed for its predictive capacity.

Reproducibility: Based on the results for the fraction of complete test sequences (about 100% in total), the SkinEthic™ HCE SE and LE protocols are highly reproducible: within-laboratory reproducibility was about 94% (SE) and about 96% (LE), while between laboratory reproducibility was about 92% for both SE and LE. The acceptance criteria for these characteristics were fulfilled and were above the acceptance criteria set by the VMG (i.e. WLR ≥ 85% and BLR ≥ 80%).

Predictive capacity: The values are as follows:

a) SE: **89% specificity**, **43% sensitivity** and **66% overall accuracy**.

b) LE: **66% specificity, 72% sensitivity** and **69% overall accuracy**.

c) Test Strategy with EPRA (Reactive → SE; Non-Reactive → LE): **77% specificity, 55% sensitivity** and **66% overall accuracy**.

The SkinEthic™ HCE was not able to meet the 'definitely acceptable' criteria for sensitivity and overall accuracy defined by the VMG (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%) in any of the options above, and the sensitivity was even considered to be 'definitely unacceptable' (specificity < 50%; sensitivity < 80%, overall accuracy < 65%).

EpiOcular™ EIT

The predictive capacity of the EpiOcular™ EIT method was analysed on the basis of two viability thresholds: 50% and 60%. Since concordance of predictions was used as measure for reproducibility, for both WLR and BLR two sets of values are available, one calculated on the basis of a 50% viability cut-off, one on the basis of a 60% viability cut-off.

Reproducibility: Based on the results for the fraction of complete test sequences (about 100% in total), the EpiOcular™ EIT test method is highly reproducible: within-laboratory reproducibility was about 94% (50% cut off) and about 95% (60% cut off), while between laboratory reproducibility was about 91% (50% cut off) and 93% (60% cut off). The acceptance criteria for these characteristics were fulfilled and were above the acceptance criteria set by the VMG (i.e. WLR ≥ 85% and BLR ≥ 80%).

Predictive capacity: Since the purpose of the test methods will be the identification of chemicals *not classified for serious eye damage/eye irritation* (UN GHS/EU CLP No Category) as an initial step of a Bottom-Up test strategy (Scott L. *et al.* 2010) and not the discrimination between UN GHS/EU CLP Categories 1 and 2, the VMG considered that it is acceptable to have a lower specificity than sensitivity (higher false positives than false negatives). Nevertheless, specificity should not be too low in order to allow for the correct identification of the majority of the chemicals not classified as irritant to the eye. The values are as follows:

a) 50% cell viability cut-off:

75% specificity, **81% sensitivity** and **78% overall accuracy** when considering all chemicals, irrespective of physical state (solid/liquid) for which different protocols are available. When considering liquids only, the values area slightly higher, when considering solids only they are slightly lower, indicating that the test methods performs better with chemicals in liquid state.

b) 60% cell viability cut-off:

71% specificity, 88% sensitivity and **79% overall accuracy** when considering all chemicals irrespective of physical state (solid/liquid). Similarly than for the 50% cut-off, predictions for liquid chemicals showed higher accuracy than for solid chemicals. Moreover, using a 60% cut-off, specificity and sensitivity for solids were basically equal (about 75% and 77%).

The liquids protocol met all of the 'definitely acceptable' criteria for specificity, sensitivity and overall accuracy defined by the VMG (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%) while the sensitivity of the solids protocol was 'definitely unacceptable' according to the acceptance criteria defined by the VMG (sensitivity < 80%). Therefore, the EpiOcular™ EIT solids protocol underwent optimization by the method developer and further validation within EIVS.

The overall predictive capacity of EpiOcular™ EIT obtained in EIVS (including the main study and the post-optimization validation study) is summarised in table 1.

Table 1: *Predictive capacity values of the **EpiOcular™ EIT** for liquids and solids and on the basis of two viability cut-offs for the prediction model: 50% and 60% and considering the data obtained during the main study (A) and those compiling data of the main study and those of post-optimization validation. Data for liquids (1) and solids (2) were generated on the basis of different protocols. Results obtained*

with the solids protocol (2) did not meet the targets set by the VMG and the protocol underwent optimization within the scope of the EIVS. Subsequently, the optimized protocol was assessed through a validation set of chemicals (4). The combined performance of EpiOcular EIT on the basis of liquids and solids are presented before optimization of the solids protocol (3) and after optimization (5).

	50 % cut-off			60% cut-off		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
(A) Results obtained during main study (before optimization of solids protocol)						
(1) Liquids	82.5%	96.2%	69.8%	**81.9%**	**98.3%**	**66.7%**
(2) Solids	73.0%	66.7%	79.7%	75.9%	76.9%	74.8%
(3) Liquids and solids	77.9%	81.4%	74.5%	79.0%	87.6%	70.5%
(B) Results obtained through post-optimisation testing (solids) and during main study (liquids)						
(4) Solids	76.8%	88.2%	64.3%	**78.0%**	**93.5%**	**60.7%**
(5) Liquids and solids	79.4%	91.8%	66.5%	**79.7%**	**95.7%**	**63.0%**

All of the 'definitely acceptable' criteria for specificity, sensitivity and overall accuracy defined by the VMG (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%) were met by the liquids protocol and the optimized solids protocol using a 60% cut-off in the prediction model.

VMG decisions and conclusions

The experimental part of the EIVS was completed in March 2012 and the data generated with both test methods were analysed subsequently. For EpiOcular™ EIT, the protocol for liquids met all the acceptance criteria set by the VMG, whereas not all acceptance criteria were met by the protocol for solids, nor for any of the protocols or test strategy of SkinEthic™ HCE. Analysis of the EpiOcular™ EIT validation study data for solids identified that there was scope for improvement. Thus, the VMG allowed the optimization of the EpiOcular™ EIT solids protocol. The test method developer (MATTEK) successfully optimized the solids protocol between April and October 2012 and results were made available to the VMG. As a result, the optimized solids protocol was allowed to enter validation within the framework of the current study. The added experimental work relating to the validation study took place in the lead laboratory (Beiersdorf) between March and June 2013. The finalized data analysis was performed by the EURL ECVAM statistician also preparing the statistical report for the optimized EpiOcular™ protocol for solids. The report has been finalised and has been made available to ESAC (ESAC CIRCA) last week of February 2014. On the basis of the study results, the VMG made the following conclusions:

Conclusions and Recommendations of the VMG

- The VMG considers the EpiOcular™ EIT original liquids protocol and the optimised solids protocol as scientifically valid (reproducible and accurate) to identify chemicals not requiring classification for serious eye damage/eye irritation under UN GHS.

- The VMG recommends that the 60% cut-off (as submitted by the method developer) is used rather than the 50% cut-off due to higher sensitivity and accuracy, and lower number of total mispredictions.
- A wide range of chemical types, chemical classes, molecular weights, LogP, chemical structures, etc. were tested and no clear limitations regarding applicability could be identified. The VMG therefore recommends that EpiOcular™ EIT is considered applicable to the testing of all types of chemicals, until proven contrary.
- The VMG does not recommend the use of the SE or LE protocols of SkinEthic™ HCE to identify chemicals not requiring classification for serious eye damage/eye irritation under UN GHS.
- The VMG also does not recommend the use of EPRA to orient chemicals to the SE (reactive) or LE (non-reactive) protocols as proposed in the SkinEthic™ HCE test strategy.
- The VMG recommends the optimisation of the SkinEthic™ HCE test method considering different protocols for liquid chemicals and solid chemicals (as in EpiOcular™ EIT).
- The VMG recommends the use of positive control(s) and associated acceptance criteria that are strict enough to allow easy detection of inappropriate conduct of the assay (e.g., in the optimised SkinEthic™ HCE).
- The VMG concludes that the use of 2 tissue replicates in similar or modified RhT/MTT-based test method aiming at identifying chemicals not requiring classification for serious eye damage/eye irritation is statistically and scientifically justified.
- The VMG considers the current endpoint detection system using standard photometry as appropriate to assess direct MTT-reducers and coloured chemicals, when tissue viability falls within the linear range of the spectrophotometer (e.g., < 140%), or when the uncorrected viabilities already identify the test chemical as requiring classification.
- For coloured chemicals interfering too strongly with the MTT-reduction assay an alternative endpoint detection system should be used (e.g., HPLC/UPLC-photometry).

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 (2014) Guidance on information requirements and Chemical Safety Assessment, Chapter R.7a: Endpoint specific guidance. Online (last accessed 10/3/14): http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf

European Commission (EC) (2008a) REGULATION (EC) No 440/2008 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL 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). Official Journal of the European Union L142, 1-739.

European Commission (EC) (2008b) 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.

European Commission (EC) (2004) Directive 2004/73/EC of 29 April 2004 adapting to technical progress for the 29th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Union L152, 1-316.

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A. and Zuang. V. (2004) A modular approach to the ECVAM principles on test validity. ATLA 32, 467-472.

Nguyen, D.H., Beuerman, R.W., De Wever, B. and Rosdy, M. (2003) Three-dimensional construct of the human corneal epithelium for *in vitro* toxicology. In Alternatives Toxicological Methods, edited by Salem, H.and Katz S.A., CRC press, 47-159.

OECD (1999) OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring No. 5. Compliance of Laboratory Suppliers with GLP Principles. Paris, France: Organisation for Economic Cooperation and Development. Available at: [http://www.oecd.org/env/testguidelines].

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

OECD (2009),Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264076303-en.

OECD (2009),Test No. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: 10.1787/9789264076310-en.

OECD (2012),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

OECD (2012) 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: 10.1787/9789264070646-en. Available at: [http://www.oecd.org/env/testguidelines].

Pfannenbecker U, Bessou-Touya S, Faller C, Harbell J, Jacob T, Raabe H, Tailhardat M, Alépée N, De Smedt A, De Wever B, Jones P, Kaluzhny Y, Le Varlet B, McNamee P, Marrec-Fairley M, Van Goethem F.(2012): Cosmetics Europe multi-laboratory pre-validation of the EpiOcular™ reconstituted human tissue test method for the prediction of eye irritation. Toxicol In Vitro 27, 619-626.

Rubingh C (2013): Final Eye Irritation Validation Study on Human Tissue Models: Statistical Analysis and reporting on the EpiOcular™ EIT,

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) (2013) Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Fifth revised edition, UN New York, USA and Geneva, Switzerland. Available at: [http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf].

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

Objective *Why does EURL ECVAM require advice on the current issue?*	The opinion of ESAC on the EIVS should support EURL ECVAM with respect to the **development of an EURL ECVAM Recommendation** on the EpiOcular™ EIT test method for serious eye damage/eye irritation testing. The EURL ECVAM Recommendation will address (1) the scientific basis of the assay in relation to the adverse effect (serious eye damage/eye irritation), (2) the reproducibility and transferability of the assay as assessed during the study, (3) the predictive capacity of the assay as assessed during the study and its usefulness in an *in vitro* test strategy as proposed by Scott et al. (2010), either as an initial step in the Bottom-Up approach or the second step in the Top-Down Approach of this strategy, (4) its applicability (considering the two protocols available for solids and liquids) and its known limitations. N.B. the proposed *in vitro* test strategy should be understood as a testing tool potentially replacing the current *in vivo* test, but being used within a more integrated approach for ocular toxicity assessment, combining non-testing with testing data (e.g. REACH guidance, ECHA 2014). ESAC's advice should enable EURL ECVAM to conclude, within its EURL ECVAM Recommendation, on the **potential adequacy of the EpiOcular™ EIT test method for routine testing** of serious eye damage/eye irritation for regulatory purposes (e.g. EU test method, OECD Test Guideline), outlining the possible complementary information that will be required, in addition to EpiOcular™ EIT data, to make decisions on the presence/absence of ocular toxicity. The advice of ESAC should support EURL ECVAM with respect to an analysis of possible remaining **data gaps** that need to be addressed in view determining EpiOcular™ EIT's potential use and usefulness within the proposed test strategy supporting serious eye damage/eye irritation testing of chemicals (substances and mixtures). Moreover, the ESAC scientific peer review of the EIVS should provide scientific advice on the **plausibility and adequacy of the decisions of the Validation Management Group** that concluded on the necessity of conducting optimisation work of both test methods assessed in the EIVS (EpiOcular™ EIT and SkinEthic™ HCE).

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?*	1) **DESIGN & CONDUCT OF STUDY:** The ESAC is requested to review whether the EURL ECVAM Eye Irritation Validation Study (EIVS) was conducted appropriately in view of the objective of the study. The study objective concerned both EpiOcular™ EIT and SkinEthic™ HCE. The study objective was to assess: (1) **reproducibility** of the test method within one laboratory (WLR) (2) **transferability** of the test method to other laboratories (3) **reproducibility** of the test method between laboratories (BLR) (4) **predictive capacity** of the test method for distinguishing chemicals not requiring classification from chemicals requiring classification as Category 1 (serious eye damage) or Category 2 (eye irritation). (5) **the applicability domain and possible limitations** of the test method. The selection of the test chemicals and analyses of possible reasons for misclassifications in view of determining patterns reg. chemical class or physicochemical property should be carefully reviewed. When reviewing the design and conduct of the study, the following issues should be addressed in particular: (a) Clarity of the test definition (module 1) (b) Clarity of the definition of the study objective and study management (c) Appropriateness of the study design & execution in view of the study objectives, *inter alia:* ○ Is the **number of tested chemicals** sufficient for the purposes of the study? ○ Are the **reference data** used for assessing in particular the predictive capacity appropriate and of good quality? ○ Was the identification of **validation chemicals** conducted in an appropriate manner (i.e. presence or absence of selection criteria, justification etc.)? ○ Is the **adverse effect range of the selected chemicals** appropriate for the purpose of the study ○ In case of **gaps** (chemical class etc.) – are these justified? ○ Is the **number of laboratories** sufficient? (d) Appropriateness of the **study execution** (e.g. were there pre-defined test acceptance criteria, were these respected? How were exceptions / deviations handled? Were provisions specified for retesting? Was the number of repetitions sufficient? etc.) (e) Appropriateness of the **statistical analysis** used for analysing WLR, transferability, BLR and predictive capacity. (f) Was the decision and conclusion that there was a need to optimise the SkinEthic™ HCE test method and the solids protocol of the EpiOcular™ EIT test method justified by the data produced in the context of the study? (g) Is the conclusion of the VMG concerning the use of only two instead of three tissue replicates per test chemical and controls scientifically justified

	by the data? 2) **CONCLUSIONS OF STUDY:** The ESAC is requested to assess whether the conclusions, as presented in the material made available to ESAC, are **substantiated by the information generated in the study** and are **plausible with respect to existing information and current views** (e.g. literature). In particular: (a) Are the conclusions on **reproducibility** (WLR and BLR) as well as transferability justified and plausible? (b) Are the conclusions on **predictive capacity** justified and plausible with respect to existing information and with respect to the intended use of the test method within an in vitro test strategy (bottom-up, top-down) (c) Are there **possible gaps between study design and study conclusions** which remain to be addressed in view of the suggested conclusions / use (see also point 3)? (d) Do the data generated with **validation set of chemicals together with possible available existing data** provide sufficient information on the **applicability and possible limitations** of the test method, in particular in view of its potential use within an *in vitro* test strategy (Scott et al, 2010)? (e) Based on the Cosmetics Europe study, performed in the context of the EURL ECVAM/Cosmetics Europe eye irritation validation study, the usefulness of HPLC/UPLC-photometry has been analysed as an alternative endpoint detection system in particular for highly coloured chemicals interfering with the conventional endpoint measurement through standard photometry. Are the conclusions of this report (furnished separately from the VSR) scientifically justified, in particular when considering the study design including the **chemicals assessed** and the **statistics used**? 3) **SUGGESTED USE OF THE TEST METHOD:** The ESAC is requested (a) to evaluate, on the basis of the data summarised in the validation study report, the intended use of the test method within the proposed test strategy (Scott et al., 2010) and keeping in mind that this in vitro test strategy will be part of an integrated approach, using non-testing as well as testing information. (b) to make additional recommendations (as required) on the proper scientific use of the test method within the proposed test strategy (Scott et al., 2010) taking specific aspects of this method into account (e.g. applicability, technical limitations etc.), (c) to identify possible further information required (i.e. are there data gaps or gaps with regard to mechanistic understanding?) to be able to determine the potential use and usefulness of the test method within test strategies and integrated approaches, duly considering the need to test both substances and mixtures.

4.3 TIMELINES

Timelines concerning this request *When does EURL ECVAM require the advice?*	Timeline	Indication
	Finalised ESAC Opinion required by:	**ESAC39**
	Request to be presented to ESAC by written procedure (e.g. due to urgency) prior to the next ESAC	**No**
	Request to be presented to ESAC at ESAC plenary meeting	~~**Either at ESAC meeting in 4Q 2013 or through written procedure.**~~ **At ESAC 39**

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	**YES (already set-up)**
	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

At the time of circulating this ESAC request (21/2/2014), the validation study report has not yet been finalised. The table below lists and describes the available documentation (n=40 documents) in view of discussions on this mandate at ESAC 39 and in view of the upcoming ESAC WG meeting (18-20/3/2014). The documents are logically structured at the highest level in:

1. GENERAL STUDY DOCUMENTS
2. EPIOCULAR SPECIFIC DOCUMENTS
3. SKINETHIC SPECIFIC DOCUMENTS.

Subfolders further structure the available documents.

Count	**Description of document**	**Already available? (YES/NO)**	**File name**
FOLDER: 1. GENERAL STUDY DOCUMENTS			
Subfolder 1.1 Study Management			
1	Project plan for EYE IRRITATION VALIDATION STUDY (EIVS) Describes validation project of the SkinEthic™ HCE SE (short exposure), LE (long exposure) and test strategy (combining EPRA and HCE EIT methods) as well as the EpiOcular™ EIT for the prediction of acute eye irritation	yes	EIVS_VMG_ProjectPlan_V1.pdf
2	Guidance on EIVS conduct and test method performance criteria	yes	EIVS_VMG_PerformanceCriteria_V2.pdf

3	Addendum to the guidance on the EIVS- Instructions for the testing of direct MTT-reducers and/or coloured test chemicals	yes	Addendum - Testing of MTT reducers and colorants-v3.pdf
Subfolder 1.2 Chemicals Selection			
4	Explanatory document regarding the selection of chemicals (test items) for the EIVS	yes	01-Report_EIVS chemicals selection [17Feb14].doc
5	Table of selected test chemicals, including properties and GHS classification	yes	02-Tables_EIVS chemical selection.pdf
6	Figures for the chemicals selection report	yes	03-Figures_EIVS chemicals selection.pdf
7	Appendix 1: summary tables of EIVS chemicals, sorted according to phys./chem. properties, GHS classification	yes	App 1 EIVS_chemicals_summary data.xlsx
8	Appendix 2: condensed information, i.e. graphs, and in vivo Draize eye test results for each test chemical	yes	App 2 EIVS_chemicals_in-vivo master.xlsx
9	Appendix 3: EIVS chemicals and results for OECD QSAR analysis	yes	App 3 EIVS chemicals_OECD QSAR 3.1 analysis.xlsx
10	Appendix 4: EIVS chemicals with representation of their chemical structure	yes	App 4 EIVS_chemicals_molecular structures.pdf
Subfolder 1.3 Status update EIVS study			
11	Status Update of the ECVAM/Cosmetics Europe Eye Irritation Validation Study (as communicated during the 9th meeting of the Validation Management Group (VMG) held on 10-11 May 2012)	yes	EIVS Status Update Communication.pdf
Subfolder 1.4 Validation Study Report			
12	Core document regarding the course ,the outcome and conclusions of the EIVS	yes	EIVS report
FOLDER: 2. SPECIFIC DOCUMENTS EpiOcular EIT			
Subfolder 2.1 Validation Study Report			
13	Report on the ECVAM Quality assurance audit at the production site of MATTEK, 25-26 May 2010	yes	2.1 QA audit of tissue production site.pdf

Subfolder 2.2 Statistics on replicate number			
14	Statistical analyses for using two replicates with the EpiOcular™ test method by a statistical consultant to NICEATM	yes	NICEATM Summary Statistics - EpiOcular Replicates_10May10(VMG).pdf
Subfolder 2.3 Training & Transferability			
15	Report on training and transfer of the test method to the three participating laboratories Beiersdorf, Harlan UK and IIVS	yes	Training EpiOcular Labs MatTek Report VF HK 12.12.13.pdf
Subfolder 2.4 SOP & Data Reporting			
16	Standard Operating Procedure for: Ocular Irritation Assay for Chemicals using EpiOcular™ EIT in the Eye Irritation Validation Study (EIVS), version 6.0 of 9 February 2011	Yes	EpiOcular_SOP_Final Version_6_February_9_2011.pdf
17	Reporting spreadsheets for measurements ad results	Yes	EIVS_lab_protocol_batch_week_number.xls
Subfolder 2.5 Statistical Report EpiOcular			
18	Statistical Analysis and Reporting on the EpiOcular™ EIT in the EIVS, version of 10 February 2014	yes	TNO report_EpiOcular_last final final.pdf
Subfolder 2.6 Optimisation of solids protocol			
19	Standard Operating Procedure for: Ocular Irritation Assay for Chemicals using EpiOcular™ EIT in the Eye Irritation Validation Study (EIVS), version 7.0 as of 11 October 2012. Version 7.0 has been amended to include extended exposure time for solid test substances	Yes	EpiOcular_SOP_Version_7-solids_Oct_11_2012.pdf
20	Update on the optimization of the EpiOcular Solids protocol-presentation by Dr. Yulia Kaluzhny, Mattek.	Yes	EpiOcular-EIT-Solids-MatTek_updated_101212-.pdf
21	Summary of the optimization of the EpiOcular Solids protocol- by Dr. Yulia Kaluzhny, Mattek.	Yes	Solids protocol optimisation Summary by Yulia Kaluzhny.pdf
Subfolder 2.7 Post-optimisation of validation			
Subfolder 2.7.1 SOP final for postopt validation			
22	Standard Operating Procedure for: Ocular Irritation Assay for Chemicals using EpiOcular™ EIT in the Eye Irritation Validation Study (EIVS), final version 8.0 as of 5 March 2013.		EpiOcular_SOP_8_2013.3.5.pdf

Subfolder 2.7.2 Reporting Template			
23	Reporting spreadsheets for measurements ad results, updated version	yes	EIVS_batch week.xls
Subfolder 2.7.3 Statistical report			
24	Statistical Report	yes	EIVS_stat_report.pdf
FOLDER: 3. SPECIFIC DOCUMENTS SkinEthic HCE			
Subfolder 3.1 QA audit of tissue production site			
25	Report on the ECVAM Quality assurance audit at the production site of EpiSkin (Lyon France), 19 March 2010	Yes	ECVAM audit report EpiSkin signed.pdf
Subfolder 3.2 Training & Transferability			
Subfolder 3.2.1 EPRA			
26	Letter from TNO (9 February 2010) containing statement on the transferability study for the Cysteine and Lysine Direct Peptide Reactivity Assays	Yes	TNO - Brief Lammers 09.02.2010.pdf
27	TNO-Report of the training on the Cysteine and Lysine Direct Peptide Reactivity Assays, performed on 3-4 June 2009 (Procter & Gamble, Cincinnati, Ohio, USA).	Yes	TNO - DPRA Training Report 03.11.2009.pdf
28	TNO-Report of the Transferability Study on the Cysteine and Lysine Direct Peptide Reactivity Assays as of 3 November 2009	Yes	TNO - DPRA Transfer Report V8590_02 03.11.2009.pdf
29	Final TNO-Report of the Transferability Study on the Cysteine and Lysine Direct Peptide Reactivity Assays as of 9 February 2010 (signed)	Yes	TNO - DPRA Transfer Report V8590_02 FINAL_signed.pdf
30	ECVAM comments on the TNO Transferability study on the Cysteine and Lysine Direct Peptide Reactivity Assays (16 November 2009)	Yes	TNO - ECVAM Comments ARES(2009)326935.pdf
Subfolder 3.2.2 SkinEthic HCE			
31	Report on the training at L'Oréal of experimental steps with CARDAM representatives, participating in the EIVS, 19-22 January 2010	Yes	EIVS_SkinEthic HCE_Final Training report_Cardam_VMG.pdf

32	Report on the training at L'Oréal of experimental steps with Ceetox representatives, participating in the EIVS, 26-29 January 2010	Yes	EIVS_SkinEthic HCE_Final Training report_Ceetox_VMG.pdf
33	Report on the transfer of the method to participating laboratories (CARAM & Ceetox), participating in the EIVS, 8 February- -9 April 2010	Yes	EIVS_SkinEthic HCE_Final Transfer report.pdf
Subfolder 3.3 SOPs (EPRA & HCE)			
34	Standard operation procedure (version 3.0 of 28 June 2011) for Eye irritation Peptide Reactivity Assay (EPRA) for use in the RhT Test Methods Eye Irritation Validation Study (EIVS)	Yes	EPRA Protocol EIVS v3 28.06.11.pdf
35	Standard operation procedure for eye irritation test with the SkinEthic™ Human Corneal Epithelial Model (version 1.0 of 18 June 2010)	Yes	EIVS_SkinEthic HCE SOP_ V1_ 2010June18.pdf
Subfolder 3.4 Data Reporting Templates			
36	Reporting spreadsheets for measurements and results	Yes	EIVS_LABNAME_SE_BATCHNUMBER_ WEEKNUMBER_v1.0.xls
37	Summary tables for data submissin	Yes	Summary table for data submission v1.xls
Subfolder 3.5 Statistical reports			
Subfolder 3.5.1 Stats replicate number on EIVS data			
38	Analysis of Long Time Exposure Data regarding the impact of reducing the number of samples in a given run from three to two by a statistical consultant to NICEATM	Yes	Analysis of Long Time Exposure Data.pdf
39	Analysis of Short Time Exposure Data regarding the impact of reducing the number of samples in a given run from three to two by a statistical consultant to NICEATM	Yes	Analysis of Short Time Exposure Data.pdf
Subfolder 3.5.2 Statistical report EIVS SkinEthic HCE			
40	Final report from TNO on the statistical analysis and reporting on the SkinEthicTM HCE performance in the EIVS, final version of 14 February 2014	Yes	DRAFT_TNO report_SkinEthic_last final final.pdf

7. TERMS OF REFERENCE OF THE ESAC WORKING GROUP

7.1 ESTABLISHMENT OF THE ESAC WORKING GROUP

During its 39th meeting on 19/19 June 2013, the ESAC plenary unanimously decided to establish an ESAC Working Group "Eye Irritation" charged with the detailed scientific review of a study on two in vitro test methods for eye irritation testing based on reconstructed human tissue (EpiOcular EIT and SkinEthic HCE).

7.2 TITLE OF THE ESAC WORKING GROUP

Full title:
ESAC WG charged with the review of the EpiOcular™ EIT and SkinEthic™ HCE test methods for eye irritation testing.

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 EIVS study concerning the EpiOcular™ test method and, at a later point in time, of the optimised SkinEthic™ HCE test method. The review needs to address the questions put forward to ESAC by EURL ECVAM (section 4.2).

In essence, 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 Validation Management Team (VMT) 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**.** A reporting template has been made available (Appendix 1) intended to facilitate drafting of the 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

The Coordinator has proposed timelines which should be agreed upon during the first Teleconference (Item 1 in the table):

Item	Proposed date/time	Action	Deliverable
1	18-20 March 2014	• Discussion of review material, • Resolution of issues possibly contentious between WG members • Clarification (hearing session) of questions on the study • Drafting of WG report	First draft of WG report
2	End of April / beginning of May	Teleconference to discuss draft report	
3	Further teleconferences as needed in the subsequent weeks	TBD	TBD
4	Early June	Consolidation of WG report and drafting of ESAC opinion	Draft WG report / draft ESAC opinion
5	End of June	Final WG report and draft ESAC opinion to ESAC	
	July	Adoption of WG Report and ESAC opinion by ESAC	Final WG report and ESAC opinion

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC WORKING GROUP

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 study-specific questions outlined in section 4.2.

APPENDIX 1 REPORTING STRUCTURE FOR THE ESAC WORKING GROUP REPORT

The following suggested structure follows 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. A template (ESAC WG consensus report template) has been created and will be made available to the ESAC.

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-v5.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

EURL ECVAM-coordinated Eye Irritation Validation Study (EIVS)

Title page information	
File name	ESAC_WG_REPORT_EIVS&HPLC.doc
Abbreviated title of ESAC request	EURL ECVAM Eye Irritation Validation Study
Relating to ESAC REQUEST No.	2013-03 – v2.01
Request discussed through	ESAC38 (preliminary); ESAC39 (final including review mandate)
Report to be handed over to ESAC Chair and EURL ECVAM Coordinator by	Q2/Q3 2014

Version tracking

Date	Version	Author(s)	Description
2014 07 25	1.1	ESAC WG	Draft
2014 09 10	1.2	ESAC WG	Final draft, containing description of contentious issue and majority / minority position

**) The ESAC WG reached consensus on all review questions but one, which is described in more detail in Appendix 1 to this report and which includes an outline of the majority opinion (five WG members) and the minority opinion (one WG member). The WG was composed of six members.*

TABLE OF CONTENTS

ESAC Working Group

Full title:
ESAC WG charged with the review of the EpiOcular™ EIT and SkinEthic™ HCE test methods for eye irritation testing, and the HPLC/UPLC alternative endpoint measurement.

Abbreviated title: ESAC WG

The 38th ESAC plenary (19 June 2013) unanimously decided to establish an ESAC Working Group "*Eye Irritation*" (referred to below as 'ESAC WG") to undertake a detailed scientific review of an EURL ECVAM/Cosmetics Europe Eye Irritation Validation Study (EIVS) evaluating two *in vitro* test methods for acute eye irritation caused by chemicals, both of which are based on Reconstructed human Tissue (RhT): the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE). The ESAC WG was established according to the ESAC Rules of Procedure and included experts nominated by international partner organisations of EURL ECVAM collaborating under the ICATM (International Collaboration on Alternative Test Methods) framework.

A revised EURL ECVAM Request for ESAC advice was unanimously adopted at the 39th ESAC plenary (11/12 March 2014) to include a scientific review of a related Cosmetics Europe study of HPLC/UPLC as an alternative endpoint detection system, particularly for highly coloured chemicals interfering with the conventional endpoint measurement of formazan by OD-photometry.

The ESAC WG conducted the peer review between 18 March 2014 and 10 September 2014, this Report was endorsed by the ESAC WG on 10 September 2014 apart from the contentious issue described in Appendix 1. The WG met in 18 to 20 March 2014 and further communicated by written procedure and teleconferences (10 April and 21 May 2014).

The ESAC WG members appointed by ESAC were[1]:

- Dr. Jon Richmond (ESAC Member, WG Chair)
- Dr. Annette Kopp-Schneider (ESAC Member)
- Dr. Renate Krätke (ESAC Member)
- Dr. Kristina Kejlova (ESAC Member)
- Dr. Tohru Marunouchi (Fujita-Health University; ICATM nomination by JaCVAM)
- Dr. Jill Merrill (FDA; ICATM nomination by NICEATM/ICCVAM)

ESAC Coordination:

- Dr. Claudius Griesinger (ESAC Coordinator)
- Dr. Michael Schäffer

1. One member of the ESAC WG (JM) had been the ICCVAM liaison to the EIVS VMG during the validation study. She had not been a member of the core Validation Management Group (VMG). Although all decisions during the validation study were taken by the core VMG, JM advised and commented on some elements of the study design. The ESAC WG does not believe that this liaison role constitutes a significant conflict of interest with respect to the work of the ESAC WG.

Abbreviations used in the document

- **BLR** Between-laboratory reproducibility
- **CLP** Classification, Labelling and Packaging regulation (EU CLP)
- **ECHA** European Chemicals Agency
- **ECVAM** European Centre for the Validation of Alternative Methods
- **EIT** Eye Irritation Test
- **EIVS** Eye Irritation Validation Study
- **ESAC** ECVAM Scientific Advisory Committee
- **ESAC WG** ESAC Working Group
- **EURL ECVAM** European Union Reference Laboratory for Alternatives to Animal Testing
- **FDA** Food and Drug Administration
- **GCCP** Good Cell Culture Practice
- **GLP** Good Laboratory Practice
- **HCE** Human Reconstructed Corneal Epithelium
- **HPLC/UPLC** High Performance Liquid Chromatography/Ultra Performance Liquid Chromatography
- **IATA** Integrated Approach to Testing and Assessment
- **ICATM** International Collaboration on Alternative Test Methods
- **MTT** 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
- **OD** Optical density
- **OECD** Organisation for Economic Co-operation and Development
- **PC** Positive Control
- **RhT** Reconstructed Human Tissue
- **SCCS** European Commission Directorate General for Health and Consumer Protection: Scientific Committee on Consumer Safety
- **SOP** Standard Operating Procedure
- **UN GHS** United Nations Globally Harmonized System for the Classification and Labelling of Chemicals.
- **VC** Vehicle Control
- **VMG** Validation Management Group
- **VSR** Validation Study Report
- **VMT** Validation Management Team
- **WLR** Within-laboratory reproducibility

Executive summary

The ESAC WG was established by ESAC in response to a formal request from EURL ECVAM for a peer review of, and scientific advice on:

1. An Eye Irritation Validation Study (EIVS) co-sponsored by EURL ECVAM and Cosmetics Europe evaluating two Reconstructed human Tissue (RhT) *in vitro* test methods (the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE)) for acute eye irritation caused by chemicals: and
2. A related Cosmetics Europe study of HPLC/UPLC as an alternative endpoint detection system, particularly for highly coloured chemicals interfering with the conventional endpoint measurement of formazan by OD-photometry.

The ESAC WG was charged with conducting a detailed scientific peer review of these studies and advise on:

1. 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 **reliability** (transferability; within and between laboratory reproducibility) of the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE).

2. The **usefulness** of HPLC/UPLC-photometry as an alternative endpoint detection system in particular for highly coloured chemicals interfering with the conventional endpoint measurement by OD-photometry.

The analysis and conclusions of the ESAC WG are based primarily on the:

1. March 2014 Validation Study Report on "THE EURL ECVAM - COSMETICS EUROPE PROSPECTIVE VALIDATION STUDY OF RECONSTRUCTED HUMAN TISSUE-BASED TEST METHODS FOR IDENTIFYING CHEMICALS NOT REQUIRING CLASSIFICATION FOR SERIOUS EYE DAMAGE/EYE IRRITATION TESTING", the study ANNEXES listed with the ESAC request, supplementary information obtained during the preparation of this report; and

2. The Cosmetics Europe Task Force Eye Irritation (TF-EI) "Use of HPLC/UPLC for Endpoint Detection in *In Vitro* Reconstructed human Tissue (RhT) Test Methods to Expand Applicability to Strongly Coloured Test Substances" report, dated 1 March 2014, and other supporting documents supplied to ESAC WG.

Eye Irritation Validation Study (EIVS)

After the initial EIVS ring-trial of chemical testing neither of the RhT test methods satisfied the acceptance criteria set by the VMG.

1. SkinEthic™ HCE SE and LE protocols, and their use in combination in a test strategy with the Eye irritation Peptide Reactivity Assay (EPRA), failed to meet the predetermined acceptability values set by the VMG for sensitivity, specificity or overall accuracy. The SE and LE protocols were, however, found to be highly reproducible. The VMG advised, and the manufacture agreed, to undertake test method optimisation outside the EIVS.

   1.1. The ESAC WG believes this was a sound decision.
   1.2. The ESAC WG supports the VMG view that the supplementary EPRA testing did not add value to the use of the SkinEthic™ HCE protocols used in this validation study.

2. With the original EpiOcular™ EIT protocol (SOP V 6.0) some acceptance predictive capacity criteria were not met for solid chemicals. The VMG allowed further test optimisation within the EIVS. The findings with the revised solid chemicals protocol (SOP V 8.0) met the acceptance criteria established by the VMG. These post-optimisation findings are incorporated and analysed in the VSR, and form the basis of the conclusions and recommendations of the ESAC WG.

Having taken account of supplementary data analysis commissioned by the ESAC WG and referenced elsewhere in this report, ESAC WG supports this test method now being considered for regulatory use.

HPLC/UPLC

1. On the basis of Cosmetics Europe Report, and supplementary analysis undertaken during the WG Review, the ESAC WG believes that suitably qualified HPLC/UPLC assay systems can be used as an alternative endpoint detection system in particular for highly coloured chemicals interfering with the conventional endpoint measurement by OD-photometry.
2. However, the HPLC/UPLC provided marginally but systematically lower values for formazan levels than OD-photometry. This may impact on some of the prediction models which currently rely on formazan quantification by OD-photometry.

1. Study objective and design

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

NOTE: (a) please summarise briefly in your own words the study objective as outlined in the VSR and (b) provide an appraisal as to whether the study objective is clearly and comprehensibly defined in the VSR.

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

EIVS

The EIVS Validation Study Report (VSR) title serves as a concise, simplified summary of the main study objective: "PROSPECTIVE VALIDATION STUDY OF RECONSTRUCTED HUMAN TISSUE-BASED TEST METHODS FOR IDENTIFYING CHEMICALS NOT REQUIRING CLASSIFICATION FOR SERIOUS EYE DAMAGE/EYE IRRITATION TESTING".

Section 2.1 of the VSR fully defines the study objectives. These were to evaluate the relevance (predictive capacity) and reliability (transferability; and within and between laboratory reproducibility) of two RhT test methods, the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE), to discriminate chemicals (substances and chemical mixtures) 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) and as implemented by the EU CLP regulation (EU CLP). The test methods were not designed, intended, or evaluated to differentiate between GHS Category 1 (irreversible effects) and 2A-B (reversible effects).

The relevance and reliability of the two RhT test methods, the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE), were evaluated within a ring trial testing 107 chemicals (substances including chemical mixtures) with a view to consideration of their use in future *in vitro* tiered test strategies (Scott et al, 2010) within an integrated, tiered approach using strategic combinations of alternative test methods to replace the Draize eye test, either as the initial step in a Bottom-Up Approach or a second or subsequent step in a Top-Down Approach.

The Cosmetics Europe Task Force Eye Irritation Report

The Cosmetics Europe Task Force Eye Irritation report entitled "Use of HPLC/UPLC for Endpoint Detection in *In Vitro* Reconstructed human Tissue (RhT) Test Methods to Expand Applicability to Strongly Coloured Test Substances" (referred to below as the Cosmetics Europe Report) is the essential reference document for the HPLC/UPLC endpoint detection system relied upon by the ESAC WG.

The Cosmetics Europe Report is not a Validation Study Report (VSR): the study evaluated the performance of a bio-analytical endpoint measurement system (which is not itself a test method).

The Cosmetics Europe Report confirms that the aim of the study was to determine the relevance (reliability and reproducibility) of suitably qualified analytical HPLC/UPLC systems to measure the MTT reduction product formazan generated by the use of three different RhT test methods which currently rely on the photometric detection and quantification by photometric measurement of optical density (OD) of formazan as the endpoint measurement for cytotoxicity. Each of the three

RhT test methods used within the study was relevant to one of three discrete toxicity endpoints (skin corrosion, skin irritation and eye irritation).

The starting point for the Cosmetics Europe study is the need to devise RhT/MTT-reduction product measurement systems that can be used with highly coloured chemicals which interfere with the OD-photometric measurement of the formazan endpoint. The study also evaluated the use of the HPLC/UPLC endpoint measurement system for chemicals that do not have this property.

The HPLC/UPLC study was conducted with the intention of using the findings to:

- Extend the applicability of *in vitro* RhT test methods for skin corrosion, skin irritation and eye irritation to include strongly coloured chemicals which interfere with current endpoint measurement - the OD-photometric detection and quantification of the MTT reduction product formazan.
- Establishing the suitability of qualified HPLC/UPLC systems for consideration as an alternative endpoint detection system for these and other RhT/MTT-based *in vitro* test methods and other chemicals, and demonstrating the Within Laboratory Reproducibility (WLR) and Between Laboratory Reproducibility (BLR) of HPLC/UPLC for this purpose – with different suitably qualified HPLC/UPLC systems used in the different laboratories participating in the study.

(b) Appraisal of clarity of study objectives as outlined in the VSR

EIVS

With respect to EIVS, the ESAC WG believes that the study objectives were sufficiently clear, and reflected in the way the study was designed, conducted, analysed, and reported in the VSR.

The Cosmetics Europe Task Force Eye Irritation Report

The Cosmetics Europe Report has no “Objectives” section. The information summarised above was compiled from various sections of the report.

The ESAC WG believes that the full text of The Cosmetic Europe Report clearly and sufficiently defines the study objectives.

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

NOTE: What is, according to the VSR, the overall purpose of the test method? Examples are a) scientific use (e.g. basic/applied research, b) screening for product development c) regulatory testing etc.

Both the EIVS and the Cosmetics Europe HPLC/UPLC reports describe the intended application of the RhT test methods and suitably qualified HPLC/UPLC bio-analytical measurement systems as being for regulatory testing.

EIVS

The study reported in the VSR was undertaken with a view to determining if the two RhT test methods, the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE), are ready to be considered for regulatory use.

The EIVS study and the VSR conclusions and recommendations are relevant to the possible future use of these RhT test methods in a future *in vitro* tiered test strategy (Scott et al, 2010) used within an integrated approach using strategic combinations of alternative test methods to replace the Draize eye test to 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 United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) and as implemented by the EU Classification, Labelling and Packaging regulation (EU CLP).

The Cosmetics Europe Task Force Eye Irritation Report

A number of RhT test methods are already validated, accepted and used by manufacturers, test laboratories, and regulatory authorities for evaluation of skin corrosion (OECD Test Guideline (TG) 431 (OECD, 2013)) and skin irritation ((OECD TG 439 (OECD, 2013)). Other RhT test methods are currently undergoing validation for use as part of integrated testing strategies (Scott et al, 2010) for eye irritation.

The Scientific Committee on Consumer Safety (SCCS) of the European Commission Directorate General for Health and Consumers previously advised that *"...for coloured substances, a different endpoint, not involving Optical Density (OD) quantification, should be envisaged. Analytical methods such as HPLC/UPLC might be more appropriate to detect formazan in the in vitro assay..."* (McNamee *et al*. 2009). The Cosmetics Europe study proposes the use of suitably qualified HPLC/UPLC endpoint measurement in this context.

The use of suitably qualified HPLC/UPLC endpoint measurement is proposed for consideration for regulatory use as an alternative endpoint detection measurement for *in vitro* RhT test methods which currently rely on the detection and quantification by OD-photometry of the MTT reduction product formazan as a measure of cytotoxicity – specifically to extend the range of chemicals which such test methods can evaluate to include highly coloured chemicals which interfere with the conventional endpoint measurement by OD-photometry. An alternative endpoint detection measurement for RhT/MTT reduction product test methods that overcomes this current limitation would extend the applicability domain of these RhT tests.

The ESAC WG believes The Cosmetics Europe Report describes a bio-analytical that has the potential to overcome a current limitation on the use of these test methods.

(a) Analysis of the scientific rationale provided in the VSR

NOTE: Is the scientific rationale of the test method AND (consequently) for conducting the study clearly explained? Consider how the test method may contribute

(a) to the scientific understanding of the specified health/environmental effect or aspects of it?, i.e. does it provide relevant mechanistic information such as physiological pathways relevant for toxicity ("toxicity pathways") or other key physiological events leading to toxicity?

(b) to the prediction of the specified downstream health/environmental effect or aspects of it?

Moreover, does the VSR make sufficient reference to the relevant body of scientific literature?

EIVS

The EIVS VSR describes the scientific rationale of the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE) partly in terms of other validated and accepted RhT models' proven ability to measure cytotoxicity by the OD-photometric quantification of MTT reduction products (Mosmann, 1983) as a reliable surrogate measurement of a range of toxicological endpoints after the controlled exposure of an RhT air-tissue interface to solid or liquid test materials (substances , including chemical mixtures).

The VSR also provides a scientific rationale and justification in the specific context of eye irritation, reasoning that RhT test system cytotoxicity/tissue viability is a potentially plausible surrogate measure of eye irritancy capable of discriminating between UN GHS eye irritants (CLP Category 1 and Category 2) and non-irritants produced by a variety of relevant, known, *in vivo* mechanisms of ocular toxicity. The ESAC WG also notes that in other contexts (for example skin irritation) cytotoxicity/cell viability following chemical exposure is a recognised surrogate measure for chemically induced *in vivo* toxicity: with the *in vivo* cytotoxicity stimulating an inflammatory response involving the innate immune system in proportion to the degree of cytotoxicity produced by the chemical.

The VSR concedes that not all of the relevant *in vivo* mechanisms of ocular toxicity are addressed by the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ HCE test methods, and the ESAC WG notes that no histopathological evaluation is performed as part of the test methods. If these RhT test methods are to be used to identify non-irritants within an integrated testing strategy then the ESAC WG believes other test methods will be needed to assess mechanisms of toxicity where the relevant adverse outcome pathway has a vascular component or is marked by changes other than cell death.

The arguments advanced in the VSR in support of the test methods take account of the findings of previous pre-validation studies and other peer reviewed publications. The ESAC WG notes that some of the peer reviewed papers cited in the VSR narrative are not listed in the VSR bibliography.

The VSR provides plausible evidence and reasoned arguments in favour of these RhT test methods having the potential to be at least as relevant and reliable as the current *in vivo* test system for the identification of non-irritants when combined with other *in vitro* methods within an integrated, tiered approach using strategic combinations of alternative test methods within future (tiered) testing strategies to replace the Draize eye test (Scott et al, 2010).

The VSR acknowledges that, for the foreseeable future, the *in vivo* Draize rabbit eye test is unlikely to be replaced by a single *in vitro* test (Eskes et al, 2005). Instead the VMG proposes that the intended use of these RhT test methods would be within a larger testing strategy designed to replace or reduce animal testing for determining chemical-induced eye irritation potential for regulatory purposes (Scott *et al.*, 2010).

It is proposed by Scott et al (2010) that within such future testing strategies, based on the expected ocular toxicity profile of the test chemical, and using one of two tiered testing approaches (Bottom-Up or Top-Down), the systematic and sequential application of one or more *in vitro* test method would be used to determine chemically induced eye irritation potential. The full details of the appropriate integrated testing strategies and test methods have yet to be defined.

The validation study was designed and conducted to provide evidence of whether the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Reconstructed Corneal Epithelium (HCE) test methods are sufficiently relevant and reliable to identify chemicals not requiring classification, i.e. chemicals that are not UN GHS CLP Category 1 (serious eye damage) or Category 2 (eye irritant). The results of the study (i.e. prediction of non-classified chemicals) are to be considered for use within a testing strategy (Scott et al, 2010).

The Cosmetics Europe Task Force Eye Irritation Report

The Cosmetics Europe Report describes an HPLC/UPLC endpoint detection system, not a test method.

In the view of the ESAC WG The Cosmetics Europe Report adequately describes and explains a known limitation of the currently used OD-photometric endpoint detection system - its inability to reliably assess the cytotoxic effects of chemicals which interfere with the OD-photometric endpoint measurement - and provides the rationale for the development and use of a suitably qualified HPLC/UPLC endpoint detection system as an alternative endpoint measure.

The Cosmetics Europe Report summarises relevant peer-reviewed literature and other reference material explaining the relevance and limitations of RhT test methods which rely on the OD-photometric quantification of the MTT reduction product formazan as a means of estimating cytotoxicity as a surrogate measure of skin corrosion, skin irritation, and eye irritation, and other toxicological endpoints. It also relies on references explaining the likely relevance and reliability of suitably qualified HPLC/UPLC measurement systems for MTT reduction products for the same purposes.

In the view of the ESAC WG, the breadth and quality of the reference materials cited in the Cosmetics Europe Report is both informative and supports the claims made by the authors of the Cosmetics Europe Report with respect to the underlying science and the need for and potential benefits of improved RhT formazan endpoint detection systems to accurately estimate cytotoxicity when cytotoxicity measurements cannot be reliably derived by OD-photometric methods.

(b) Analysis of the regulatory rationale provided in the VSR

NOTE: Is a regulatory rationale specified, i.e. a specific application of the test method for purposes of generating data with respect to regulatory requirements as specified in legislation or internationally agreed guidelines etc.? If so, how does the study and its objective and design relate to this regulatory rationale? Are the relevant regulatory documents appropriately referenced?

Both the EIVS VSR and the Cosmetics Europe HPLC/UPLC report identify 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 RhT-based methods in the context of the regulatory requirements are adequately specified.

EIVS

If validated for eye irritancy determination, RhT test methods such as the EpiOcular™ EIT and the SkinEthic™ HCE would have the potential to reduce animal testing by reliably identifying chemicals not requiring classification when used within an appropriate non-animal testing strategy (Scott et al, 2010).

The VMG makes the point that considering the small prevalence of chemicals inducing serious eye damage (Adriaens et al, 2014) RhT test methods validated for this purpose would significantly reduce animal testing by identifying the much larger number of chemicals not requiring classification.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC WG believes the regulatory rationale is clearly set out in the Cosmetics Europe Report.

RhT test methods relying on the detection and quantification of formazan as the endpoint measurement are already validated, accepted and used by manufacturers, test laboratories , and regulatory authorities for evaluation of skin corrosion (OECD Test Guideline (TG) 431 (OECD, 2013a)) and skin irritation ((OECD TG 439 (OECD, 2013b)). Other RhT test methods are currently undergoing validation for use as part of a proposed testing strategy (Scott et al, 2010) for eye irritation (EIVS

Report). One of the major limitations of the current MTT OD-photometric assay is interference with the endpoint measurement by chemicals with absorbance measurements at or close to that of formazan.

The SCCS previously advised that *"...for coloured substances, a different endpoint, not involving Optical Density (OD) quantification, should be envisaged. Analytical methods such as HPLC/UPLC might be more appropriate to detect formazan in the in vitro assay..."* (McNamee *et al*. 2009).

An alternative endpoint detection measurement for RhT/MTT reduction product formazan that overcomes the current limitation on their use for chemicals which interfere with the OD-photometric detection and quantification of formazan would extend the applicability domain of RhT test methods that rely on the measurement of formazan.

The Cosmetics Europe Report also references the EIVS study, pointing out that two test chemicals used within EIVS demonstrated strong colour interference with the measurement by OD-photometric absorbance of formazan.

1.3 Appraisal of the appropriateness of the study design

NOTE: Is the study design appropriate in view of the stated objective of the study? This includes an analysis of the number of laboratories involved in the study, the organisation of study management including chemical selection, quality check of data, and independence of statistical analysis, i.e. was the statistician independent from the test method submitter/developer and, depending on the study, from the VMG. More technical aspects can also be considered such as an appraisal of the nature and number of test items used (details however to be provided in section 6, test materials), retesting in case of unqualified tests, pre-defined test acceptance criteria etc.

EIVS

The VSR describes a validation study of two RhT test methods, the EpiOcular™ EIT and the SkinEthic™ HCE, and provides a detailed and reasoned rationale for the validation study design. An Annex to the VSR, "Eye Irritation Validation Study (EIVS), Guidance on the Eye Irritation Study (EIVS) Conduct for the Reconstructed Human Tissue (RhT) Assays and Performance Criteria to Assess the Scientific Validity of SkinEthic™ HCE and the EpiOcular™ EIT", describes in detail how the study was planned and conducted.

EIVS complies 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 SkinEthic™ HCE (in combination with EPRA) and the EpiOcular™ EIT (with separate protocols for liquids and solids) were assessed for their usefulness as test methods to identify chemicals not requiring classification for serious eye damage/eye irritation (UN GHS / EU CLP No Category; "non-irritants") and their reliable discrimination from all classes of classified chemicals with a view to their forming an initial step of a Bottom-Up approach or a later stage of a Top-Down approach (Scott *et al.*, 2010).

The relevance and reliability of the two RhT test methods, the EpiOcular™ EIT and the SkinEthic™ HCE, were assessed by a ring trial using a number of coded test chemicals (substances and chemical mixtures, N=107) as determined by the statistician to ensure a study of appropriate power, and

supported by the best available quality-assured *in vivo* Draize eye irritation reference data for the comparative evaluation of the accuracy of the results.

The EIVS VMG established test method performance acceptance criteria to evaluate the performance and validity of the SkinEthic™ HCE SE and LE test methods, the SkinEthic™ HCE/EPRA test strategy, and the EpiOcular™ EIT as stand-alone test methods for the identification of chemicals not classified as eye irritants in the framework of Bottom-Up/Top-Down test strategies. In setting those criteria the VMG took into account of:

1. The background and specific objectives of the validation study;
2. The requirements of regulatory authorities and industry when testing and classifying chemicals for eye irritation;
3. The within test variability inherent in the *in vivo* Draize eye irritation data and the manner in which these data are currently used for classifying eye irritants according to UN GHS / EU CLP;
4. The standards of performance which are expected from *in vitro* test methods;
5. The way in which the *in vitro* tests are to be used (as one test within a tiered test strategy); and
6. The statistical power of the validation study.

In its scientific review of the study, the study findings, and the VMG conclusions the ESAC WG concludes that, subject to specific qualifications set out below, the study design was generally appropriate and robust, the acceptance criteria appropriate at the time, and that the study findings have provided the evidence and analysis required to satisfy the study objectives.

SkinEthic™ HCE

A total of 107 coded chemicals were tested in two different protocols in three runs with three replicate tissues per run, in three laboratories. The same chemicals were also tested in Eye Irritation Peptide Reactivity Assay (EPRA) in one run with three replicate measurements in one laboratory (TNO).

The RhT testing was performed by three laboratories (CARDAM, CeeTox, and L'Oréal). As the lead laboratory, L'Oréal provided training and support to the two naïve laboratories during the training and transfer phase of the study.

Two different protocols of SkinEthic™ HCE, a short-time exposure protocol (SE), and a long-time exposure protocol (LE) were evaluated. A test strategy combining, in a sequential manner, the Eye Irritation Peptide Reactivity Assay (EPRA) with both the SE and LE variants of the SkinEthic HCE test method was also assessed.

The SkinEthic™ HCE SE and LE protocols were developed in the expectation that the SE protocol would be better suited for use with EPRA reactive chemicals, and the LE better suited for use with EPRA non-reactive chemicals. The EPRA data were generated before the chemical selection was finalised to ensure the chemicals set selected for the main study had a balanced set of EPRA-reactive and EPRA non-reactive chemicals. All chemicals were evaluated using both the SkinEthic™ HCE SE and LE protocols.

The VMG set acceptance criteria for the maximum number of non-acceptable/non-qualifying test run sequences that any laboratory could make during the study. The study documents made provision for reporting if certain threshold numbers of non-acceptable/non-qualifying test run sequences were reached, in which case the lead laboratory would seek to identify and remedy the cause of the

problem. This was the only inter-laboratory contact permitted during the testing phase of the study: it did not take place during the testing phase of the study.

However, early in the transfer phase, a series of non-qualified runs was reported by one laboratory (CeeTox, USA) on the basis of unacceptable Positive Control values. The plausible explanation offered in the VSR is that the tissue component of the test kits had been compromised whilst in transit: the transport arrangements were reviewed and improved, and the SOP amended to include an assessment of Positive Control values as confirmation of tissue viability. This problem was not seen again after these changes were implemented.

EpiOcular™ EIT

In the ring trial a total of 107 coded chemicals were tested (three runs with two replicate tissues per run, in three laboratories) using EpiOcular™ EIT SOP (V 6.0), using separate protocols for liquid and solid chemicals, and two alternate cut-off points (50% and 60% cell viability).

The VMG set acceptance criteria for the maximum number of non-acceptable/non-qualifying test run sequences that any laboratory could make during the study. The study documents made provision for reporting if certain threshold numbers of non-acceptable/non-qualifying test run sequences were reached, in which case the lead laboratory would seek to identify and remedy the cause of the problem. This was the only inter-laboratory contact permitted during the testing phase of the study: in the event, there was no inter-laboratory contact during the testing phase of the study.

The RhT testing was performed by three laboratories (Beiersdorf, Germany (the lead laboratory); Harlan, UK; and IIVS, USA). The manufacturer and lead laboratory provided training and support to the other laboratories during the training and transfer phase of the study.

The original EpiOcular™ EIT SOP V6.0 protocol for liquid chemicals produced results satisfying all of the VMG acceptance criteria for sensitivity, specificity and overall accuracy. The 60% cut-off was preferred as it resulted in a better sensitivity and generated no false negatives based on the mode of all predictions (the 50% cut-off generated one false negative for a Category 2B chemical).

With the original EpiOcular™ EIT protocol (SOP V6.0) some of the VMG acceptance criteria were not met for solid chemicals. Six chemicals were under-predicted with the 60% cut-off based on the mode of all predictions, one of which (which was misclassified by two of the laboratories) is classified *in vivo* as Category 1.

The VMG considered that there was scope for improved test method performance if a balanced increase in sensitivity offset by some decrease in specificity could attain a compromise sensitivity ≥ 90% with specificity maintained ≥ 60%. Further optimisation was therefore recommended for the EpiOcular™ EIT protocol for solid chemicals. The VMG allowed test optimisation to be undertaken within EIVS by the test developer in an attempt to improve sensitivity (lowering the false negative rate) without unduly compromising specificity (misclassifying true negatives).

The VSR describes the test optimisation process. This was undertaken using 11 of the most challenging EIVS solid chemicals: the six EIVS solid chemicals that had been under-predicted (false negatives), and five EIVS chemicals that had correctly been predicted as non-irritants but which had given borderline results. Having considered the new results and an amended SOP (V 8.0), the VMG commissioned a post-optimisation validation study of the EpiOcular™ EIT optimised solid chemicals protocol, undertaken in one laboratory with a total of 60 solid chemicals (the original 52 EIVS solid chemicals plus eight additional chemicals to compensate for the 11 EIVS chemicals used during the optimisation of the protocol) - effectively increasing the EIVS chemicals set to N=115.

The post-optimisation testing was undertaken only in the lead laboratory (Beiersdorf). The VMG reasoned that, in view of the high reproducibility of the test method during the original ring-trial, supplementary testing in one laboratory would be sufficient to determine the predictive capacity of the revised EpiOcular™ EIT optimised solid chemicals protocol.

The ESAC WG considered whether the choice of the one laboratory that conducted the supplementary testing using the protocol optimised for the testing of solid chemicals might have unintentionally introduced a source of bias – for example had the chosen laboratory produced the best values with the original protocol. The ESAC WG believes the choice of this laboratory did not introduce any such bias. The performance values obtained at the chosen laboratory when it used the protocol not optimised for solid chemicals were comparable to the values obtained by all three laboratories that participated in the validation of the original protocol (SOP V6.0).

The post-optimisation findings are incorporated and analysed in the VSR, and form the basis of the VMG conclusions on the performance of the EpiOcular™ EIT.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC WG believes the Cosmetics Europe Report provides a detailed and reasoned rationale for the HPLC/UPLC study design.

With respect to the selection of test laboratories: Cosmetics Europe as a co-sponsor of the study determined general and specific requirements that any participating laboratory must fulfil, and then nominated the laboratories which undertook the testing. There was no tendering and selection process. This is not best practice, but the ESAC WG found no evidence of this having compromised or biased the study findings.

The ESAC WG noted that as this was not a test method validation study, rather it was the evaluation of an analytical method, and there was no SOP for the qualified HPLC/UPLC measurements. The Cosmetics Europe Report states that the suitably qualified HPLC/UPLC systems were used in accordance with the criteria required for qualification.

Two study phases are described in the Cosmetics Europe Report.

- Phase 1: qualifying the HPLC/UPLC test systems; and
- Phase 2: the use of suitably qualified HPLC/UPLC for endpoint measurement for formazan detection and quantification when a range of chemicals were evaluated with three RhT test methods - each used for one of three different toxicological endpoints.

The ESAC WG considered the study design for both phases of the reported study:

Phase 1: Qualification

The Cosmetics Europe Report describes the qualification of the HPLC/UPLC systems in detail.

The specific HPLC/UPLC systems at each of the three laboratories were qualified according to the May 2001 US Food and Drug Administration (FDA) guidance for industry “Bioanalytical Method Validation”. In this context “qualification” is a technical term for the validation of bio-analytical methods, encompassing all of the procedures required to demonstrate that a particular bio-analytical method used for quantitative measurement of specified analytes in a given biological matrix is both reliable and reproducible for its intended use. The qualification process involves the use of specific laboratory investigations documenting and confirming that the performance characteristics of the bio-analytical method are suitable and reliable for its intended analytical use.

Based on the information contained in the Cosmetics Europe Report (including detailed technical specifications, conditions, and performance data on the HPLC/UPLC systems; information on sample preparation; and analysis of key parameters) the ESAC WG is satisfied that:

1. Determination of the reliability and reproducibility of HPLC/UPLC as an alternative endpoint for detection of formazan in the MTT assay in *in vitro* RhT test methods falls within the scope of the FDA guidance document.
2. The fundamental parameters (in this case selectivity, precision and accuracy, matrix effect, carryover, reproducibility, and stability) for the qualification of HPLC/UPLC for measurement of formazan in RhT extracts were relevant and appropriate.

Phase 2:

The ESAC WG is satisfied that for the purposes of this study a sufficient number of RhT MTT-reduction product test methods, chemicals, and laboratories were used.

The ESAC WG analysis of some elements of study design (e.g. the adequacy of the number of laboratories involved in the study, the organisation of study management, the statistical analysis, and the nature and number of test items used) is provided elsewhere in this report.

The WG considers that the study design described in The Cosmetics Europe Report made proper provision to confirm that:

1. Formazan extracts are stable during the operational conditions under which HPLC/UPLC was performed, and for the conditions and duration of sample storage for the purposes of this study.
2. Chemicals that compatible with the RhT MTT-reduction product test methods give comparable results when the endpoint in the same samples is measured using both OD-photometric measurement and the suitably qualified HPLC/UPLC system.
3. For chemicals that are incompatible with RhT test method because they interfere with MTT OD-photometric assay, when the endpoint is measured by the suitably qualified HPLC/UPLC system the application of the RhT test method Prediction Models resulted in a generally appropriate classification (based on the RhT reference data) of the chemicals with respect to the relevant toxicological endpoint. The Cosmetics Europe Report also provides the classifications that would have resulted had *in vivo* reference data been used.

Three laboratories (L'Oréal R&I, Pierre Fabre Laboratories, and VITO) each tested up to 26 chemicals. The three laboratories each used a different RhT test method measuring different toxicological endpoints (EpiOcular™ EIT as per the EIVS protocol v6.0 for eye irritation (the version of the SOP which pre-dated the optimisation of the test method for solid chemicals within EIVS) was performed by Pierre Fabre Laboratories - and the SkinEthic™ RHE as per OECD TG 431 for skin corrosion, and the EpiSkin™ as per OECD TG 439 for skin irritation were both performed at L'Oréal R&I). Where the RhT test method protocols made provision for the identification of MTT-reducers and chemicals incompatible with the OD-photometric endpoint measurement systems, the relevant SOP provisions were followed.

The laboratories performing the RhT test methods also undertook the formazan extraction procedures (with the same samples then used for both the OD-photometry and the HPLC/UPLC measurements) and the OD-photometric endpoint measurement steps specified in the test method SOPs.

Three laboratories (L'Oréal R&I, Pierre Fabre Laboratories, and VITO) then used the suitably qualified HPLC/UPLC system for endpoint measurements of all of the samples, and applied the Prediction Models using both endpoint measurements.

To evaluate intra-laboratory performance the Pierre Fabre Laboratories undertook two separate HPLC/UPLC endpoint measurements of the same EpiOcular™ eye irritation formazan samples.

1.4 Appropriateness of the statistical evaluation

NOTE: Consider whether the statistical approaches chosen are appropriate. This includes statistical calculations performed ex-ante such as sample size calculations as well as ex post statistical analysis of the data (e.g. for purposes of variability and predictive capacity). Is the choice of methods sufficiently justified?

EIVS

The statistical reports annexed to the EIVS VSR are detailed and informative. However, see below, the ESAC WG takes a different view of the way predictive capacity should be calculated in the context of this study: in the VSR the VMG relies on calculations based on all individual test runs rather than the number of test chemicals; the ESAC WG believes the VMG approach artificially increases the sample size, produces inappropriately narrow confidence limits, and results in an inappropriate degree of certainty with respect to the point estimates of sensitivity and specificity.

The VSR records that having taken independent statistical advice *"…statistical tools that were used for data analysis included in the final biostatistics reports were described in the Guidance document on the conduct of the EIVS and in a Statistical Analysis and Reporting Plan. The biostatistics analyses procedures reported in the Statistical Analysis and Reporting Plan were developed by the ECVAM and TNO biostatisticians before completion of the experimental phase of the study and were approved by the VMG before the biostatistics analyses began"*. The study results were analysed by an independent statistician.

The ESAC WG is satisfied on the basis of the documentation provided that these procedures were followed during the planning, conduct, analysis and reporting of the study.

For statistical purposes four protocols were evaluated: two RhT *in vitro* test methods, including four alternative time combinations for exposure and incubation: with EpiOcular™ EIT separating liquids from solids; and SkinEthic™ HCE differentiating EPRA reactive from non-reactive chemicals. This required a balanced chemical selection of: (i) classified versus non-classified chemicals; (ii) solids versus liquids; and (iii) EPRA reactivity versus non-reactivity.

Non-qualifying test runs were not included in the data used to calculate final BLR or Predictive Capacity values for the EpiOcular™ EIT and SkinEthic™ HCE test methods.

For the SkinEthic™ HCE test method no test chemical used for the testing phase was considered incompatible with the test method by any of the three laboratories, with either the SE or the LE protocol. All chemicals were thus included in all of the statistical analyses of this test method.

For the EpiOcular™ EIT in one laboratory one of the original chemicals (#33) proved incompatible with the test system due to colour interference with the OD-photometry measurement of formazan: the same was true of one of the non-EIVS chemicals (#98) used for retesting after protocol optimisation for solid test chemicals.

The ESAC WG gave careful consideration to the means by which the data generated could be used to best reflect the predictive performance of the methods without overestimating accuracy values. The ESAC WG believes the relevant confidence limits for the predictive capacity must be based on the number of chemicals tested, rather than the number of test runs performed as in the context of this study the chemical is the sampling unit and the test runs are technical replicates of each sampling unit. As a result, two separate classes of variability have to be considered: variability between runs for each chemical and variability between chemicals. An evaluation of predictive capacity on the basis of the number of test runs, and disregarding the fact that there were multiple runs with each chemical, does not take proper account of the two-level nature and structure of the data. The length of the confidence interval decreases with sample size, and artificially increasing sample size in this way results in confidence limits that are too narrow. The ESAG WG believes the confidence limits based on the number of test runs relied upon in the VMG Report are likely to be optimistic. The ESAC WG therefore requested additional analysis with confidence limits derived from the number of chemicals tested rather than the number of test runs. See Section 10.1.

Furthermore, the ESAC WG considers that the predictive capacity calculation should reflect the way decisions will be made in practice. Specifically, as chemical classification will be based upon one test run, the analysis of the study results should allow for an accuracy range to be determined based upon the best and worst case based on data from single runs (i.e. using each single run per chemical and selecting the runs in view of obtaining (a) the worst possible results or (b) best possible results in terms of concordance of predictions with the reference data) rather than aggregating the data from different test runs with the same chemical. There was agreement within the ESAC WG that values calculated on the basis of best-case/worst-case scenarios would provide sufficiently robust performance values to better determine whether the performance levels envisaged by the VMG had been met. The ESAC WG commissioned additional data analysis from EURL ECVAM based on this approach, and the result were discussed at the May 2014 ESAC WG teleconference. See Section 10.1.

Although the ESAG WG settled on the best-case/worst-case option, consideration was also given to bootstrapping. The ESAC WG agreed to recommend to EURL ECVAM that, in cases where data matrices of all performed runs are used as a basis for the assessment of the predictive capacity of assays, bootstrap calculations should be performed as the best practice approach in order to derive accurate and appropriate figures.

SkinEthic™ HCE

All of the chemicals selected were tested in both the SkinEthic™ HCE SE and the LE protocols, and the results obtained with the SE and LE protocols were then independently assessed with respect to their reproducibility and predictive capacity.

The combined EPRA/SE/LE model as described in the SOP was evaluated for predictive capacity only.

Within-laboratory reproducibility

For each laboratory, concordance of classifications and overall Standard Deviation (SD) were calculated based only on qualified tests for test chemicals for which at least two qualified tests were available.

Standard Deviation associated with each laboratory was calculated using all available test sequences, i.e., including both qualified and non-qualified tests.

Between-laboratory reproducibility

The reported BLR values are based on the final classification for each test chemical in each participating laboratory, based on the arithmetic mean value obtained from three acceptable test run sequences.

Concordance of classifications between laboratories and overall Standard Deviations were calculated based only on qualified tests for chemicals for which at least one qualified test per laboratory was available. The overall Standard Deviation for the study was calculated using all available test sequences.

Predictive capacity

All qualified tests for each test chemical were used to calculate predictive capacity values. The calculations were based on the individual final prediction of each qualified test in each laboratory.

See EIVS comments immediately above.

EpiOcular™ EIT

Although separate protocols are followed for liquid and solid chemicals, the reproducibility and predictive capacity of EpiOcular™ EIT were evaluated for the whole test method (the testing of liquids and solids).

EpiOcular™ EIT was originally developed with, and the submission of the test method to EURL ECVAM for validation only referenced, a single threshold of 60% mean tissue viability in the prediction model. However, before training and transferability took place, out of necessity the manufacturer had to substitute the original test kit insert-membrane with another form of membrane insert.

This change produced minor changes in the results obtained with the test kit. With the new membrane the manufacturer demonstrated that a sensitivity > 90% could potentially still be achieved by using a 50% mean tissue viability cut-off instead of 60%, with a significant gain in specificity. The VMG therefore decided that within EIVS the EpiOcular™ EIT test method performance would be evaluated using two alternate prediction models:

1. The original cut-off at 60% mean tissue viability as in the submission to EURL ECVAM; and
2. A cut-off at 50% mean tissue viability.

The VMG conclusions and recommendations with respect to EpiOcular™ EIT are based upon an analysis of the initial round of testing in three laboratories, and the post-optimisation for solids testing data from one laboratory.

The ESAC WG's concerns about the basis of some of statistics set out in the VSR and relied upon by the VMG are set out above, and are discussed further at Sections 10.1 and 13.1.

Within-laboratory reproducibility

For each laboratory, concordance of classifications and overall Standard Deviation (SD) were calculated based only on qualified tests for test chemicals for which at least two qualified runs were available.

Standard Deviations associated with each laboratory were calculated using all available test sequences, i.e., including both qualified and non-qualified tests were also calculated.

Between-laboratory reproducibility

The BLR values reported in the VSR are based on the final classification for each test chemical in each participating laboratory, based on the arithmetic mean value obtained from three acceptable test run sequences.

Concordance of classifications between laboratories and overall Standard Deviations were calculated based only on qualified tests for chemicals for which at least one qualified test per laboratory was available.

The overall Standard Deviations for the study were calculated using all available test sequences.

As mentioned above, only one laboratory reported test results for solid chemicals after protocol optimisation for solid chemicals: therefore there are no BLR values for the use of the optimised protocol (SOP V 8.0).

Predictive capacity

All qualified tests for each test chemical were used to calculate predictive capacity values.

The calculations were based on the individual final prediction of each qualified test in each laboratory.

The VMG conclusions and recommendations with respect to EpiOcular™ EIT are based upon an analysis of the initial round of testing, the post-optimisation testing data, and the additional analysis referred to at Section 1.4 and 10.1.

The Cosmetics Europe Task Force Eye Irritation Report

Phase 1: Qualification

The Cosmetics Europe Report details out how the acceptance criteria for, and statistical evaluation of, the parameters (selectivity, precision, accuracy, carryover, matrix effect (RhT insert effect), stability and reproducibility) were derived and performed.

Phase 2: Testing

The ESAC WG's analysis takes account of this not being a formal test method validation study.

Section 2.8 of the Cosmetics Europe Report, "Statistical analysis", provides only limited and basic information, and the ESAC WG comments are therefore based on the full text of the Cosmetics Europe Report and additional information provided during and after a presentation of details of the study given to ESAC WG at its meeting at ECVAM in March 2014.

Regression analyses of all MTT formazan calibration curves were consistent with high intra- and inter-day reproducibility.

For the main study HPLC/UPLC measurements, which were conducted in three laboratories, the low SD values obtained tend to confirm that the viabilities measured by the HPLC/UPLC systems qualified by the different laboratories were comparable in all of the laboratories. These SD values were well

below the SD values previously reported for RhT test method validation studies (Pfannenbecker et al, 2013: Zuang et al, 2002; Kandárová et al, 2006: Spielmann et al, 2007) using OD-photometric determination of the endpoint. The Cosmetics Europe Report concludes the study results demonstrate a very high level of reproducibility between the participating laboratories measuring formazan in solvent extracts using HPLC/UPLC as an endpoint detection system.

For consistency checking of the HPLC/UPLC measurements over time, the formazan extract samples from the *in vitro* eye irritation RhT test methods were analysed twice in one participating laboratory (Pierre Fabre Laboratories) and showed 100% concordance.

To determine whether the tissue viability values obtained by the OD-photometric and HPLC/UPLC endpoint measurement systems were consistent, the difference between the values obtained by HPLC/UPLC analysis and in a repeated HPLC measurement were divided by 2 (See the Cosmetics Europe Report, Table 15). The values thus derived (0.0-2.55) also tended to confirm a high level of consistency with qualified HPLC system measurements separated by time. The HPC/UPLC values were consistently and marginally lower than those obtained by OD-photometry.

The ESAC WG accepts that variations on these methods of assessing agreement between two test methods are commonly used. However the use of correlation for this purpose can be misleading, particularly when the true value of the variable of interest is unknown (Bland and Altman, 1986). The ESAC WG has taken account of an alternative approach developed for use in these situations (see Section 7.1).

2. Collection of existing data

NOTE: Validation studies typically make use of existing data, e.g. either as reference data (prospective studies) OR as reference data and testing data (retrospective study). Moreover, validation studies may use other information such as data in the literature, data banks etc.

2.1 Existing data used as reference data

Which data sources were used for compiling the reference data associated with the test chemicals?

EIVS

The VSR and annexes provide detailed information on this point. The VSR references and relies on an extensive collection of information generated before and during the development and pre-validation of the RhT test methods.

The same chemicals were selected and used for the evaluation of the two RhT test methods - the EpiOcular™ EIT and SkinEthic™ HCE test methods for eye irritation testing. Eight additional solid chemicals were also tested in one laboratory using the EpiOcular™ EIT post-optimisation protocol for solid chemicals testing.

An essential requirement for chemical selection was the availability of complete and quality assured supporting *in vivo* data to allow comparative evaluation of the predictive capacity of the RhT test methods as measured against *in vivo* (Draize eye test) reference method.

Toxicological and physicochemical properties were important secondary considerations.

Attempts were made to identify and avoid chemicals known to have been used to develop and optimise the test methods.

The databases consulted included:

1. ECETOC database of eye irritation reference chemicals.
2. EC (DG-SANCO) Cosmetics Ingredients (CosIng) database.
3. EC New Chemicals Database (NCD) of notified substances.
4. ICCVAM (NICEATM) database of eye irritation reference chemicals.
5. US EPA database of pesticide actives.

The ESAC WG believes that, although there is some overlapping content, these are appropriate and authoritative sources; that is was not necessary to commission additional *in vivo* testing; and that there are insufficient high quality human data that might have been used as reference data.

There are only limited reference data available for the identification of mixtures of chemicals that might have served as test chemicals for this study.

The Cosmetics Europe Task Force Eye Irritation Report

Two primary sources were used for selection of test substances:

1. The chemicals repository of the EURL ECVAM / Cosmetics Europe EIVS.
2. The Scientific Committee for Consumer Safety (SCCS) Memorandum (addendum) on the *in vitro* test EPISKIN™ for skin irritation testing (SCCS, 2010).

These sources provided data on potential test substances including nomenclature, physical characteristics, source and *in vivo* classification of irritant/corrosive effects.

In addition, to ensure the identification of a sufficient number of suitable chemicals, similar information was sought on proprietary chemicals held by Cosmetics Europe member companies.

The quality controls on the information contained in the two primary sources are known: the quality controls applied to the reference data obtained for the Cosmetics Europe member companies are not explained in the Cosmetics Europe Report.

2.2 Existing data used as testing data

Point 2.2 only concerns retrospective validation studies or modular studies combining existing and newly generated data to assess an assay. Which data sources were used to collect existing testing data?

EIVS

Not applicable.

The Cosmetics Europe Task Force Eye Irritation Report

Not applicable.

2.3 Search strategy for retrieving existing data

*NOTE: Please describe and evaluate whether and how the search for existing data was planned, organised and conducted. In particular: has a **search strategy** been described and consistently applied?*

EIVS

See Section 2.1 above.

The reference data were used to identify a balanced chemical selection based on classified versus non-classified chemicals, solids versus liquids, and EPRA reactivity versus non-reactivity.

An operational master-list was generated short-listing 160 potentially eligible and available chemicals. The selection criteria were then refined to select compounds for the study.

The subsequent chemical selection was managed in two stages, first determining eligible and available substances for preliminary EPRA assessment (N=135); followed by a definitive short-list (N=107) for evaluation in the study taking the EPRA results into account.

Statistical power analysis (sample size calculation) stipulated a minimum requirement of 26 classified chemicals and 26 non-classified chemicals for each protocol.

The selection criteria then devised to produce the best balanced and most proportionate set of chemicals for the study were:

1. Classified versus non-classified chemicals was set at 50±5%.
2. A 50/50 weighting of category 1 and category 2.
3. Adequate representation of sub-categories 2A and 2B.
4. For physical state, liquids versus solids, 50±10%.
5. The division of reactive versus non-reactive was set with a wider margin at 50±15%.

After a set of 104 chemicals were selected to meet the above statistical provisions a further two chemicals, both known to cause permanent ocular colouration *in vivo*, were added to give some insight into whether such chemicals could be evaluated by the RhT test methods and the measurement endpoints being applied. These 106 chemicals were coded and distributed to the laboratories involved in the ring trial.

A known MTT reducer, potentially incompatible with the test method endpoint measure system, was then identified amongst these 106 chemicals. By the time a substitute chemical had been identified and distributed to the laboratories, testing of the MTT reducer had already been done. Therefore this additional chemical, original intended to be substituted for one of the 106 chemicals, brought the total number of EIVS chemicals tested in the ring trial to 107.

The majority of the EIVS test chemicals were pure single constituent substances, each represented by a discrete molecular structure. However, the selection included eight polymers (three homo-polymers, five co-polymers), four occurring in aqueous medium. The EIVS set also included 10 quasi-polymers (eight occurring as aqueous liquids)

There was an absence of other types of chemical mixtures.

Following the ring-trial *in vitro* testing of the 107 chemicals, and after statistical evaluation of the results, the EpiOcular™ EIT protocol for solid chemicals was subject to further optimisation. The EpiOcular™ EIT protocol optimised for solids (SOP V 8.0) was then subject to post-optimisation evaluation in one laboratory, with repeat testing of all 52 EIVS solid chemicals, and an additional eight solid chemicals, increasing the total number of chemicals used within the EIVS to 115.The eight supplementary solids comprised two GHS category 1, three category 2A, one category 2B and two GHS unclassified chemicals.

An unexpected problem arose during the study. Each test laboratory was to decide whether each test chemical was a liquid or a solid: materials which could not be pipetted at 37°C were to be deemed to be, and were to be tested as, solids. Chemical #37 was classified and tested as a liquid using the EpiOcular™ EIT during the ring-trial (the main part of EIVS) but as a solid during the validation of the EpiOcular™ EIT optimised solid chemicals protocol. This is believed to be due to seasonal factors and the way the chemical was handled between being taken from storage and tested. The VMG recommended that the SOP be amended to better define the protocol to be used for test chemicals with an unclear physical state, specifically that all viscous, waxy and gel-like chemicals are placed in a water bath for 15 minutes at 37°C before deciding if they should be tested with the liquids or the solids protocol; and that the test chemical should be applied directly from the water bath and should not be brought to room temperature before testing.

The ESAC WG would offer one additional comment. In places the RhT SOPs refer to elements of the test methods being undertaken at “room temperature”: this is an imprecise term, we believe that an appropriate temperature range should be specified instead.

The Cosmetics Europe Task Force Eye Irritation Report

The sources described at 2.1 above were used to identify data with respect to test substances meeting the pre-defined primary selection criteria (see below).

ESAC WG notes that no skin corrosive chemicals were included in the chemicals selected. This is discussed at Section 6.2 below.

2.4 Selection criteria applied to existing data

NOTE: Have consistent evaluation and decision criteria been pre-defined and applied in order to select the data and has the selection of data been explained in a transparent manner?

EIVS

See above

The Cosmetics Europe Task Force Eye Irritation Report

This section is not applicable to this study. The criteria applied to select the most appropriate chemicals are discussed elsewhere in this report.

3. Quality aspects relating to data generated during the study

3.1 Quality assurance systems used when generating the data

NOTE: Have quality assurance systems such as GLP (Good Laboratory Practice) or GCCP (Good Cell Culture Practice) been used when generating the data?

EIVS

Although five of the six laboratories that undertook the RhT testing are GLP accredited, all worked to the "spirit of GLP", with no laboratory conducting the testing in compliance with the full GLP requirements.

The VMG determined the minimum essential quality assurance requirements to be applied by all participating laboratories based on Balls et al, 1995.

The Cosmetics Europe Task Force Eye Irritation Report

Neither the RhT testing, nor the HPLC/UPLC measurements were performed to full GLP requirements

3.2 Quality check of the generated data prior to analysis

NOTE: Have the generated data been checked for quality including correct formatting (-> data reporting) prior to analysis. Has the quality check been performed by a staff member independent from the laboratory staff generating the data?

EIVS

Details of the data management procedures and statistical tools used for data analysis are included in the final biostatistics reports and the Guidance document on the conduct of the EIVS and in a Statistical Analysis and Reporting Plan.

The biostatistics analysis procedures reported in the EIVS Statistical Analysis and Reporting Plan were developed by EURL-ECVAM and TNO biostatisticians before completion of the ring-test phase of the study, and approved by the VMG before the biostatistical analysis began.

All laboratories conducting the RhT test methods used standard data-capture templates developed by the lead laboratory and approved by the VMG. The results of each run were entered by the technicians: these were checked against the raw data by the study director before the templates were forwarded to an independent statistician who compiled the results in an evaluation template: the evaluation template was then reviewed by the study directors.

An independent statistician calculated the reliability and predictive capacity values using the error checked evaluation template, and these calculations were then separately reviewed by an EURL-ECVAM biostatistician.

All statistical reports were then quality checked by the VMG.

The Cosmetics Europe Task Force Eye Irritation Report

Similar systems were applied to both Phases of the study, using data-capture templates developed by the Cosmetics Europe Eye Irritation Task Force.

Raw data were entered by the technician performing the test/analysis. These data were error checked by the local study director prior to the data being submitted to the study organiser before being entered into an evaluation template which was then error checked before the final evaluation was performed.

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)

*NOTE: Please describe the quality of the **testing data**. This may concern data newly generated in the context of the study and/or existing data (e.g. in case of retrospective or modular studies).*

EIVS

SkinETHIC™ HCE

For SkinETHIC™ HCE the number of acceptable test run sequences was high (100% for the SE protocol, 99.7% for the LE protocol). The minimum acceptance criterion set by the VMG for this characteristic was fulfilled (≥ 85%).

The ESAC WG therefore concludes that the reported performance of the SkinEthic™ HCE was based on high-quality data.

EpiOcular™ EIT

99.7% of test runs met the validation study acceptance criteria (≥ 85%), therefore the ESAC WG concludes that the validation of the EpiOcular™ EIT was based on high-quality data.

The only data gap is for chemical #33 which one laboratory considered to be incompatible with the test method due to colour interference with the MTT assay: that chemical was therefore excluded from the statistical analysis for that laboratory.

The Cosmetics Europe Task Force Eye Irritation Report

The testing data for both phases of the study is almost complete.

ESAC WG believes the data are of high quality.

4.2 Quality of the reference data for evaluating relevance[2]

*NOTE: What is the quality of the **reference data** used? Are the data and their quality sufficient in view of the study objective? To which extent has the quality of the reference data impacted on the conclusions drawn reg. performance of the assay studied?*

EIVS

See Section 4.1 for *in vitro* data and Section 2 for the *in vivo* reference data. The data used is considered to be the best available Draize eye test reference data for the chemicals tested by both RhT test methods. The limitations of DET data are discussed at Section 14.

The Cosmetics Europe Task Force Eye Irritation Report

Reference data was used only to identify and characterise the chemicals selected for use in the study.

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

NOTE: Having considered the quality of the testing data (section 4.1) and reference data (section 4.2), consider here whether the quality of the entire data set was sufficient to draw robust conclusions?

EIVS

SkinETHIC HCE

For SkinEthic™ HCE the number of acceptable test run sequences was high (100% in total for the SE protocol, 99.7% in total for the LE protocol).

ESAC WG concludes that the reported performance of the SkinEthic™ HCE to date has been based on high-quality data. The acceptance criterion set by the VMG for this characteristic was fulfilled (≥ 85%).

EpiOcular™ EIT

99.7% of test runs met the validation study acceptance criteria (≥ 85%), therefore the ESAC WG concludes that the validation of the EpiOcular™ EIT was based on high-quality data.

However, the ESAC WG believes (see Section 9.1) it would have been better to conduct a formal three-laboratory study to evaluate the performance of the EpiOcular™ EIT optimised for solids protocol (SOP V 8.0).

The Cosmetics Europe Task Force Eye Irritation Report

ESAC WG believes all of the reference and new data relied upon in the study report are of sufficient quality and quantity bearing in mind the nature and objectives of the study.

HPLC/UPLC data obtained for 24 of the 26 chemicals were evaluated. The two exceptions were for technical reasons: for chemical #7 insufficient test material was available (the Cosmetics Europe

[2] 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."

Report provides no explanation for this) and for chemical #26 there was a technical problem with the HPLC/UPLC column.

Eye Irritation

Chemicals #10 and #21 were found, using the acceptance criteria set out in the RhT SOPs, to be incompatible with the measurement of formazan by OD-photometry in the RhT test method. HPLC/UPLC data were obtained for these chemicals.

Skin Corrosion

Chemicals #21 and #26 were found, using the acceptance criteria set out in the SOP, to be incompatible with the measurement of formazan by OD-photometry in the RhT test method: cell viability was measured by the HPLC/UPLC system.

Skin Irritation

With respect to the HPLC/UPLC full data are available for 25 of the 26 chemicals: the one exception, chemical #26, was due to a technical problem with the HPLC/UPLC column.

Chemicals #10, #11, #21 and #22 were found, using the acceptance criteria set out in the SOP, to be incompatible with the measurement of formazan by OD-photometry in the RhT test method. HPLC/UPLC data were obtained.

5. Test definition (Module 1)

5.1 Quality and completeness of the overall test definition

NOTE: This included an analysis of the description of the (a) test system, (b) the protocol, (c) test acceptance criteria, (d) prediction models, (e) biological and/or mechanistic relevance of the test method for the target organ/species/system etc.

EIVS

SkinEthic™ HCE Test

The SkinEthic™ HCE test method is based on RhT technology; the test material (solid or liquid chemicals) is applied directly to the air-tissue interface, and the eye irritation potential of the chemical is derived from a prediction model based on percentage tissue viability (adjusted against a negative control) as estimated by OD-photometry of the MTT reduction product formazan.

The tissue component of the test kit is manufactured by controlled culture of immortalised human corneal epithelial cells in a chemically defined medium to produce a multi-layered epithelium similar to *in vivo* corneal epithelium with columnar basal cells, wing cells and squamous cells. The model is also characterised by the presence of relevant ultra-structural features (such as intermediate filaments, mature hemi-desmosomes and desmosomes) that characterise the corneal epithelium *in situ*. Specific cytokeratins 64kD (K.3) have also been described (Nguyen D.H. et al, 2003).

The protocol for the SkinEthic™ HCE test method described the following testing paradigms**:**

(1) The first step was an "Eye irritation Peptide Reactivity Assay" (EPRA). This was done in laboratory before chemical selection was finalised. Depending on the chemical reactivity measured in this initial EPRA, the next step would be:

(2) For EPRA reactive chemicals the SkinEthic™ HCE Short-time Exposure protocol (HCE SE).

(3) For EPRA non-reactive chemicals the SkinEthic™ Long-time Exposure protocol (HCE LE).

However, for the purposes of the EIVS study all chemicals were tested using both the HCE SE and HCE LE protocols.

The essential differences between the SkinEthic™ HCE Short-time Exposure protocol (HCE SE) and the SkinEthic™ Long-time Exposure protocol (HCE LE) were the length of time the air-tissue interface is in contact with the test material (the exposure time) and the presence or absence of a post-exposure incubation phase. The HCE SE used a 10 min exposure without post-treatment incubation, while the HCE LE was based on 1 h exposure followed by 16 h post-treatment incubation.

Following treatment with a test chemical the relative tissue viability was determined against the negative control value by the reduction of the vital dye MTT to formazan as measured by OD-photometry. Tissues treated with chemicals inducing eye irritation or serious eye damage (UN GHS/EU CLP Category 2 or Category 1, respectively) were expected to show a decrease in tissue viability below a specified threshold in comparison to the negative control.

The histological and ultra-histological features of the test system, and the use of similar endpoints and endpoint detection systems within RhT systems validated for other toxicological endpoints, made it plausible that this test RhT system may provide insights into the likely eye irritancy potential of chemicals.

Although derived from cultured human corneal cells, ESAC WG notes that the cells used have been transformed and immortalised. The ESAC WG offers no opinion on whether these cells are inherently more biologically relevant in this context than other epithelial cells (e.g. RhT models derived from normal untransformed human keratinocytes).

EpiOcular™ EIT

The EpiOcular™ tissue construct is prepared from normal human-derived epidermal keratinocytes grown under defined conditions. The VSR does not state if this is done in compliance with Good Cell Culture Practice.

The resulting 3D tissue is a non-keratinized multi-layered (5-8 cell layers) and stratified (but non-cornified) epithelium intended to model properties of human corneal epithelium.

Test materials (solid and liquid chemicals) are applied directly to the epithelial surface, which is itself in direct contact with the air. Applying chemicals directly to the air-interface of the epithelial surface mimics *in vivo* exposure of the corneal epithelium.

In the EpiOcular™ EIT, liquids and solids were applied using protocols specifying different exposure and post-exposure incubation times:

1. Liquid chemicals - 30 min exposure followed by 120 min post-treatment incubation.

2. Solid chemicals - in the original protocol (SOP V6.0) for solids - 90 minute exposure followed by 18 hours post-treatment incubation.

3. For the optimised protocol solids (SOP V 8.0) - the exposure time for solid chemicals was six hours with the post-treatment time unchanged.

The Cosmetics Europe Task Force Eye Irritation Report

Although this was not a validation study of the RhT test methods used, the Cosmetics Europe Report provides summary details of the RhT test methods.

The Report provides a detailed description of how HPLC/UPLC was suitably-validated (FDA, 2001) as an endpoint measurement for the quantification of formazan generated by RhT test methods reliant on the quantification of formazan as the endpoint measurement.

The ESAC WG notes that as yet there has been no paper on this study published in the peer-reviewed scientific literature.

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

NOTE: Are the SOPs sufficiently detailed and complete? Are the prediction models sufficiently well explained to be applied in the correct manner?

EIVS

The lead laboratories were responsible for preparing detailed SOPs for the EpiOcular™ EIT, SkinEthic™ HCE SE/LE and EPRA in collaboration with the test method developers.

The ESAC WG notes that the EpiOcular™ EIT and The SkinEthic™ SOPs require the use of different extractants to obtain the samples used for formazan quantification. The reasons for this are not clear from the study documents.

The SkinEthic™

The SOPs used during the practical testing phase of EIVS were approved by the VMG.

The SOPs were clearly written, and the testing was performed by the three laboratories without difficulty.

The one reported problem with obtaining acceptable test run sequences occurred in the training and transfer phase and was believed to relate to damage to test kits in transit.

The Prediction Models are clearly described in the VSR. Within the study the Prediction Model was applied by a statistician using test data generated by the participating laboratories.

EpiOcular™ EIT

All of the protocols and SOPs used within EIVS were clear and well-written.

The protocol changes are explained, justified and documented in the VSR and supporting documents.

The Cosmetics Europe Task Force Eye Irritation Report

This was not a formal validation study, rather it was an assessment of the performance of a qualified bio-analytical system. SOPs for the use of the HPLC/UPLC systems were not prepared or required for the study.

The SOPs used for the RhT test methods were as follows:

1. EpiOcular™ EIT as per OECD the EIVS protocol v6.0 for eye irritation (a version of the SOP which pre-dated the optimisation of the test method for solid chemicals within EIVS).
2. SkinEthic™ RHE as per OECD TG 431
3. EpiSkin™ as per OECD TG 439

6. Test materials

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

NOTE: Is the number of test items sufficient in order to draw conclusions with respect to the objective of the study? If not, are there reasons for deviations and are these explained and justified?

EIVS

See 2.3 above.

The statistical basis for the number of chemicals used is explained in the VSR and related annexes.

Almost complete data sets were obtained in all laboratories, for all test methods, and with all chemicals.

The ESAC WG believes the number of test chemicals was appropriate with respect to the study objectives, and the conclusions drawn by the VMG.

EpiOcular EIT

A total of 19 of the chemicals used for the ring trial and supplementary testing using the optimised for solids protocol had been used for the development and optimisation of the EpiOcular EIT using earlier test method SOPs (all of which pre-dated the optimisation of the protocol SOP V 6.0 used for the testing of solid chemicals within EIVS). For 17 of these 19 chemicals the results obtained in the EIVS study were consistent with the findings obtained by the manufacturer during test method development and optimisation.

Based on the results of the initial ring trial with SOP V 6.0, the 11 chemicals used within EIVS for optimising the EpiOcular EIT protocol for solid chemicals were challenging.

All 52 EIVS solid chemicals, and an additional eight solid chemicals, were retested in one laboratory using only the EpiOcular EIT optimised for solids protocol.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC WG believes that the 26 chemicals selected were adequate for this study, but notes that no skin corrosives were included in the test chemicals.

In coming to this conclusion the ESAC WG has taken account of:

- The primary and secondary criteria applied to ensure the chemicals selected were balanced and proportionate with respect to the aims of the study.
- Some of the chemicals would not be testable in some or all of the chosen RhT test methods. These chemicals are discussed elsewhere in this report.

Using the primary and secondary selection criteria for chemical selection 24 different chemicals were identified using the sources listed at 2.1 above. Two of these chemicals were identified, on the basis that they are known to interfere with the OD-photometric determination of formazan, and tested both neat and diluted to 1% (w/v) aqueous. These chemicals and dilutions comprised the 26 test materials used for the study.

Seventeen of the chemicals were from the EIVS chemicals repository, two from the SCCS Memorandum (addendum) on the *in vitro* test EpiSkin™ for skin irritation testing (SCCS, 2010), four from proprietary sources (excluding EIVS proprietary sources), and one from a general source.

6.2 Representativeness of the test items with respect to applicability

NOTE: Describe how suitable the selected test items are in order to gain – through empirical testing during the study – insight into the applicability domain / limitations of the test method.

EIVS

The EIVS chemicals were intended to be representative of the full range of chemicals for which *in vivo* test data are still required. The selection process and criteria are described above.

The ESAC WG notes the difficulty of finding chemical mixtures (other than polymers, co-polymers, and quasi-polymers) as test chemicals for validation studies. Confidence in the performance of the test methods would have been enhanced had a wider range of chemical mixtures been available and used – for example chemical mixtures which had already been assessed and labelled as eye irritants or non-irritants, e.g. by using the additivity method in accordance with GHS or CLP.

The Cosmetics Europe Task Force Eye Irritation Report

Details of the test substances including nomenclature, physical characteristics, source and *in vivo* identification of irritant/corrosive effects are shown in Table 2 of the Cosmetics Europe Report. Table 9 confirms that these test substances were balanced according to the pre-defined selection criteria.

In line with the objectives of the study, and consistent with other elements of the study design, the primary classes of chemicals were:

1. Coloured chemicals anticipated to produce colour interference of formazan detection using OD-photometry.
2. Coloured chemicals not anticipated to produce colour interference of formazan detection using OD-photometry.
3. Non-coloured chemicals.

Physical form was not a primary selection criterion, but attempts were made to ensure that the chemical set was balanced and representative with respect to physical form (solid and liquid chemicals).

The ESAC WG notes that with respect to the toxicological endpoints relevant to the RhT test methods used for this study:

1. The chemicals selected were intended to cover entire viability range (0-100%) for the *in vitro* eye irritation RhT test method.

2. There were no skin corrosives included in the chemicals set. The Cosmetics Europe Report offers a plausible explanation for this. Recent OECD analyses (Alépée et al, 2014a and 2014b) identified no *in vivo* corrosive chemicals which were coloured chemicals.

   2.1. The ESAC WG believes that the study would have been more robust had the test chemicals set also included one or more non-coloured skin corrosive chemical.

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

7.1 Assessment of repeatability and reproducibility in the same laboratory

*NOTE: How was variability and reproducibility within laboratories assessed? Possible parameters to study are (a) intrinsic data variability e.g. between replicates or runs; (b) concordance in predictions between replicates or runs. Regarding point (b), consider whether reproducibility and repeatability have been assessed separately. [**repeatability** = agreement of test results (same substance, identical conditions, e.g. equipment, operator etc.) while **reproducibility** = agreement of test results (same substance, same protocol, but not under identical conditions, e.g. different operator).*

EIVS

SkinEthic

The results presented in the VSR, based on the results of successfully completed test sequences (100% for the SE protocol, 99.7% for the LE protocol) for the SkinEthic™ HCE SE and LE protocols, tend to confirm high reproducibility.
The reported within-laboratory reproducibility was 93.9% concordance of classification for the SE protocol, and 95.5% for the LE protocol.
The SkinEthic™ HCE test method was found to be highly reproducible. The within-laboratory reproducibility (WLR) (93.9% and 95.5% concordance of classifications for the SE and LE, respectively) and the between-laboratory reproducibility (BLR) (92.3% concordance of classifications for both the SE and LE protocols) were significantly above the acceptance criteria set by the VMG (WLR ≥ 85% and BLR ≥ 80%).

EpiOcular™ EIT

Based on the results for the complete test sequences (99.7%), and using the SOP optimised for solid chemicals (V 8.0), the EpiOcular™ EIT test method was highly reproducible. Using the method of analysis set out in the VSR the WLR was 93.6% (50% cut off) and 95.2% (60% cut off).

The basis, and possible limitations, of these calculations are described elsewhere in this report (for example, Section 1.4). The ESAC WG, having analysed the data set out in the statistical reports estimates the 95% Confidence Interval (95%-CI) for the WLR using the 60% cut off to be 89.0% to 94.4%.

The WLR of the EpiOcular™ EIT optimised for solid chemicals protocol (SOP V8.0) at the one laboratory that conducted this supplementary testing improved when using the protocol optimised for the testing of 60 solid chemicals, with a reported WLR of 93.6% concordance of classifications for the 50% cut-off Prediction Model, and 95.2% concordance of classifications for the 60% cut-off Prediction Model.

All of these values are significantly above the acceptance criterion set by the VMG (WLR ≥ 85%).

The VSR gave separate consideration to the two sets of data generated by the one laboratory that tested the solid chemicals using both the original (SOP V6.0) and the "optimised-for-solids" protocol (SOP V8.0). Forty nine (49) chemicals were common to the two datasets. Using these to calculate the concordance of classifications obtained by the laboratory, the WLR values derived are 91.8% (with the 50% cut-off) and 95.9% (with the 60% cut-off) for the optimised protocol SOP V8.0), and 91.8% (with 50% cut-off) and 93.9% (with the 60% cut-off) for the original protocol (SOP V6.0).

Again, all of these values are significantly above the acceptance criteria set by the VMG (WLR ≥ 85%).

The Cosmetics Europe Task Force Eye Irritation Report

During the Phase 1 (qualification) reproducibility was addressed with both intra- and inter-day evaluations of the formazan calibration curve (reference materials having been prepared by serial dilution): three times on the same day with different stock solutions for intra-day reproducibility and repeating the calibration curve. Very high intra- and inter-day reproducibility was confirmed by regression analysis.

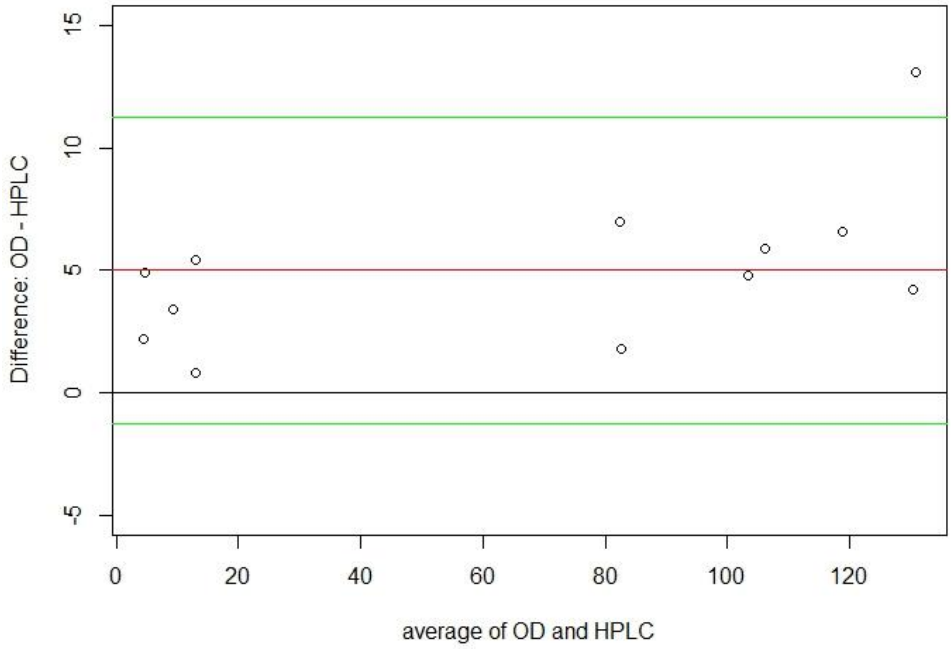
To evaluate intra-laboratory performance during Phase 2 (testing) the Pierre Fabre Laboratories undertook two separate HPLC/UPLC endpoint measurements of the same EpiOcular™ eye irritation formazan samples. The datasets confirm the results were 100% concordant.

Whilst the ESAC WG accepts that variations on these means of assessing agreement between two test methods are commonly used, the use of correlation for this purpose can be misleading and make it difficult to recognise differences, and the significance of the differences, between the findings with the different measurement systems, particularly when the true value of the variable of interest is unknown (Bland and Altman, 1986).

ESAC WG has taken account of an alternative approach (Bland and Altman, 1986) developed for use in these situations (see Section 1.4) and has re-analysed the data for the results with 12 chemicals obtained by the Pierre Fabre Laboratory using the EpiOcular™ EIT test method (a subset of VSR Table 15). The Bland-Altman Plot reproduced below plots for each chemical the difference between the OD-photometry and HPLC/UPLV values (Y-axis), against the average (arithmetical mean) of the OD-photometry and HPLC/UPLC value (X-axis).

Assuming that this sample is representative of the three RhT methods and the chemicals tested the plot also tends to confirm that the limits of agreement are good.

The plot also shows that the HPLC/UPLC values values are systematically, marginally lower than the OD-photometry: this suggests that HPLC/UPLC measurements, when judged against formazan levels estimated by OD-photometry, may slightly under-estimate the level of cytotoxicity.



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

NOTE: Are the conclusions on within-laboratory variability and repeatability reproducibility justified by the data as evaluated?

EIVS

The SkinEthic™ HCE

The statistical methods and findings are summarised above.

The SkinEthic™ HCE test method was found to be highly reproducible within laboratories.

The acceptance criterion for within-laboratory reproducibility (WLR ≥ 85%) set by the VMG was met, with the values obtained being significantly above the acceptance criteria set by the VMG.

EpiOcular™ EIT

The EpiOcular™ EIT test method was found to be highly reproducible within laboratories, with all of the protocols used, and endpoints applied, in the course of the study reported in the VSR.

All of the BLR values discussed above are significantly better than the minimum acceptance (WLR ≥ 85%) criterion set by the VMG.

The Cosmetics Europe Task Force Eye Irritation Report

The ESAC WG believes the reported within-laboratory results are consistent with the findings expected from an evaluation of a suitably qualified bio-analytical endpoint measurement system.

8. Transferability (Module 3)

8.1 Quality of design and analysis of the transfer phase

NOTE: Was the transfer phase appropriately planned, e.g. were there transfer instructions, training, minimum requirements, training SOP (if appropriate)? Were evaluation / decision criteria established beforehand defining successful transfer? If so, were these consistently applied during the analysis?

EIVS

SkinEthic™ HCE

The two naïve laboratories participating in the validation of SkinEthic™ HCE were trained by the lead laboratory (L'Oréal). Suggestions for improvement and clarification of the SOP were implemented by L'Oréal in the final version of the SOP used in the ring trial of the validation study.

Training and proficiency testing was done with 14 coded test chemicals: these included colourants and direct MTT reducers.

The variability obtained with both the SE and LE protocols at the three laboratories was low, and concordance between results of the three laboratories was high.

ESAC WG agrees with the VMG that the participating laboratories demonstrated adequate proficiency in performing the SkinEthic™ HCE and readiness to enter the formal validation study.

EpiOcular™ EIT

The three participating laboratories were trained by the manufacturer (MATTEK Corporation) to assure optimal transfer of the test protocol into their facilities and to ensure that the SOP did not allow for differing local-interpretations of the experimental steps.

After training, the participating laboratories were required to demonstrate proficiency in performing the EpiOcular™ EIT and readiness to enter the formal validation study. The technicians undertaking the testing worked with eight coded proficiency chemicals. All test chemicals were consistently predicted by the three laboratories. The variability in the findings produced by individual operators was low: using the 60% cut-off in the prediction model, one liquid chemical was predicted differently by one operator in one laboratory.

Suggestions for improvement and clarification of the SOP were implemented in the SOP (V6.0) that was used in the ring trial of the validation study.

The Cosmetics Europe Task Force Eye Irritation Report

No transfer phase was necessary.

Each laboratory undertaking the HPLC/UPLC endpoint measurement used its own suitably-qualified HPLC/UPLC system, and also had experience using the equipment for other purposes.

Each laboratory performing the RhT test methods was already experienced in their use.

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

NOTE: Are the conclusions justified by the data generated? Have critical issues that may impact on transferability been identified?

EIVS

SkinEthic™ HCE SE and LE

Based on the reported BLR and WLR findings, the ESAC WG agrees with the conclusions of the VMG that the SkinEthic™ HCE SE and LE protocols proved to be easily transferable between properly equipped and staffed laboratories, including those having no prior experience in performance of similar test methods (i.e. the two naïve laboratories that participated in the study).

Experienced personnel can readily be trained in the test method, and the necessary equipment and supplies can be readily obtained.

The ESAC WG believes that transferability should be re-confirmed if an optimised SkinEthic™ HCE model and protocol are validated.

EpiOcular™ EIT

The ESAC WG agrees with the VMG that the highly reproducible and accurate results obtained between operators and laboratories in the EpiOcular™ EIT transfer study demonstrated the transferability of the test method to naïve laboratories.

The Cosmetics Europe Task Force Eye Irritation Report

Not applicable to this study.

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

9.1 Assessment of reproducibility in different laboratories

NOTE: How was variability and reproducibility between laboratories assessed? Possible parameters to study are (a) intrinsic data variability; (b) concordance in predictions between laboratories.

EIVS

SkinEthic™ HCE SE and LE protocols

The BLR figures reported in the VSR, derived from the results for the fraction of complete test sequences (100% for the SE protocol, 99.7% for the LE protocol), tend to confirm that the SkinEthic™ HCE SE and LE protocols are highly reproducible between laboratories.
The BLR reported 92.3% concordance of classification for both SE and LE protocols. These values are significantly above the acceptance criteria set by the VMG (BLR ≥ 80%).

EpiOcular™ EIT

As testing using the EpiOcular™ EIT protocol optimised for the testing of solid chemicals (SOP V8.0) was only undertaken by one laboratory, a three laboratory BLR calculation is only available for the use of the original protocol and SOP (V6.0) and the original 107 test chemicals.

On that basis the EpiOcular between-laboratory reproducibility (BLR) was reported as 91.3% and 93.3% concordance of classifications for the 50% and 60% cut-offs analysed in this study.

These values are significantly above the acceptance criteria set by the VMG (BLR ≥ 80%).

There was no between laboratory study conducted using the optimised protocol for solids (SOP V8.0), the only BLR data provides is based on data generated using (SOP V6.0). The ESAC WG believes confidence in the revised protocol would have been strengthened had such a three laboratory study been conducted.

The Cosmetics Europe Task Force Eye Irritation Report

This was not a conventional BLR study. Each laboratory measured formazan levels by suitably qualified HPLC/UPLC systems – but the technical specification of the HPLC/UPLC systems varied between laboratories.

Three laboratories each analysed all the formazan extracts prepared from the RhT test methods used to test the chemicals within this study, having suitably qualified their HPLC/UPLC systems. The laboratories that conducted the RhT tests measured formazan levels using the OD-photometric endpoint described in the SOP for the RhT test methods.

Eye Irritation

The low SD values calculated from the HPLC/UPLC results obtained by the three laboratories demonstrate that the cell viabilities derived from the HPLC/UPLC measurements in each of the participating laboratories were comparable. The SD values were lower than the values normally associated with endpoint measurement by OD-photometry using the same RhT test methods.

In the view of ESAC WG the results detailed in the Cosmetics Europe Report demonstrate a very high level of reproducibility between the participating laboratories when measuring formazan solvent extracts from the EpiOcular™ EIT test method using HPLC/UPLC as an endpoint detection system.

However, the OD-photometry values are systematically, marginally higher than the HPLC/UPLC values: this suggests that HPLC/UPLC measurements, when judged against formazan levels estimated by OD-photometry, may slightly under-estimate the underlying level of cytotoxicity.

The HPLC/UPLC data were also compared to the OD-photometry data obtained during the study. Regression analysis of that for the 22 chemicals for which appropriate data were available confirmed a 91% concordance between measurement of the formazan extract samples by OD-photometry and HPLC/UPLC. This is also presented as evidence that the endpoint measurement using OD-photometry or HPLC/UPLC was consistent within the participating laboratory that performed the *in vitro* test method.

For the 21 tests substances where viability measurements could be derived from both OD-photometry and HPLC analysis there was 100% concordance.

Skin Corrosion

The low SD values calculated from the HPLC/UPLC results obtained by the three laboratories demonstrate that the cell viabilities derived from the HPLC/UPLC measurements in each of the participating laboratories were comparable. The SD values were lower than the values normally associated with endpoint measurement by OD-photometry.

In the view of ESAC WG the results detailed in the Cosmetics Europe Report demonstrate a very high level of reproducibility between the participating laboratories measuring formazan solvent extracts from the SkinEthic test method using HPLC/UPLC as an endpoint detection system.

The HPLC/UPLC data were also compared to the OD-photometry data obtained during the study. Using the prediction model set out in the RhT test method SOP classification as Corrosive or Non-Corrosive was determined from the viability values obtained by both OD-photometry (by the laboratory that performed the *in vitro* test method) and HPLC/UPLC (outcome from the three laboratories). Comparison of the classification by OD and HPLC/UPLC identified 100 % concordance for all test substances for which measurements could be made in both endpoint detection systems.

Skin Irritation

The low SD values calculated from the HPLC/UPLC results obtained by the three laboratories demonstrate that the cell viabilities derived from the HPLC/UPLC measurements in each of the participating laboratories were comparable. The SD values were lower than the values normally associated with endpoint measurement by OD-photometry.

In the view of ESAC WG the results detailed in the Cosmetics Europe Report demonstrate a very high level of reproducibility between the participating laboratories of measuring formazan solvent extracts from the EpiOcular™ EIT test method using HPLC/UPLC as an endpoint detection system.

The HPLC/UPLC data was also compared to the OD-photometry data obtained during the study. Regression analysis, having removed the four chemicals referred to at 4.3 above, of the findings for the remaining 21 chemicals for which appropriate data were available confirmed a 100% concordance between measurement of the formazan extract samples by OD-photometry and HPLC/UPLC. This is also presented as evidence that the endpoint measurement using OD-photometry or HPLC/UPLC was consistent within the participating laboratory that performed the *in vitro* test method.

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

NOTE: Are the conclusions justified by the data generated?

EIVS

SkinEthic™ HCE SE and LE protocols
The BLR acceptance criterion (BLR ≥ 80%), as established by the VMG, was fulfilled.
The SkinEthic™ HCE test method was found to be highly reproducible.

EpiOcular™ EIT
The BLR acceptance criterion (BLR ≥ 80%), as established by the VMG, was fulfilled using SOP V6.0.
The EpiOcular™ EIT test method, using the original study protocol (V6.0), was found to be highly reproducible between laboratories.

No BLR value is available for SOP V8.0.

The Cosmetics Europe Task Force Eye Irritation Report

The Cosmetics Europe Report concludes that for test substances that do not exhibit colour interference nor direct MTT reduction, the viability values obtained are almost identical whether using absorbance (OD-photometry) or HPLC/UPLC as the endpoint detection system for measurement of formazan, with the OD-photometry values being consistently, marginally above the values obtained with the HPLC/UPLC test method.

10. Predictive capacity and overall relevance (Module 5)

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

NOTE: How was the predictive capacity assessed? Where the reference data used in an appropriate manner? Are the conclusions justified based on the data evaluated and in view of the test method's purpose?

EIVS

The ESAC WG believes that the VMG's determination of confidence limits for the predictive capacity based on the number of test runs, rather than the number of chemicals tested, produces confidence limits which may be inappropriately narrow.

Furthermore, the ESAC WG considers that the accuracy figures should reflect the way decisions will be made in practice. Specifically, as chemical classification will be based upon one qualifying test run, the study results should have been evaluated to produce an accuracy range determined by the best and worst cases based on data from single runs rather than aggregating data from different test runs with the same chemical.

The ESAC WG requested additional analysis with confidence limits derived from the number of chemicals tested rather than the number of test runs, and accuracy values based on the base-case/worst-case combinations of individual test runs. See Section 1.4.

SkinEthic™ HCE

The same prediction model was used to evaluate the predictive capacity of both the SE and LE protocols. A mean estimated tissue viability of 50% was used as the threshold differentiating classified (UN GHS Category 1 and Category 2) from non-classified (UN GHS No Category) chemicals.

The predictive values obtained with the SkinEthic™ HCE protocol are summarised in the VSR as:

1. SE: 89% specificity, 43% sensitivity and 66% overall accuracy.

2. LE: 66% specificity, 72% sensitivity and 69% overall accuracy.

3. Test Strategy with EPRA (Reactive → SE; Non-Reactive → LE): 77% specificity, 55% sensitivity and 66% overall accuracy.

The SkinEthic™ HCE, SE and LE protocols, failed to meet the 'definitely acceptable' criteria for sensitivity and overall accuracy defined by the VMG (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%) for any of the options above.

The SkinEthic™ HCE sensitivity, as judged by the criteria (specificity < 50%; sensitivity < 80%, overall accuracy < 65%) by the VMG, was 'definitely unacceptable' [3].

The pre-study rationale for different LE and SE exposures based on chemical reactivity was not confirmed, indeed this component of test method optimisation may have contributed to the identification of a number of Category 1 substances as non-irritants.

The use of EPRA data to assign chemicals to an LE (for EPRA non-reactive chemicals) or SE protocol (for EPRA reactive chemicals) was found to be not useful, and the ESAC WG agrees with the VMG decision not to conduct a reproducibility assessment of EPRA data.

EpiOcular™ EIT

The means by which the values reported in the VSR were calculated are described above, and the results are summarised in the table immediately below.

	50 % cut-off			60% cut-off		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
(A) Results obtained during main study (before optimization of solids protocol)						
(1) Liquids	82.5%	96.2%	69.8%	**81.9%**	**98.3%**	**66.7%**
(2) Solids	73.0%	66.7%	79.7%	75.9%	76.9%	74.8%
(3) Liquids and solids	77.9%	81.4%	74.5%	79.0%	87.6%	70.5%
(B) Results obtained through post-optimisation testing (solids) and during main study (liquids)						
(4) Solids	76.8%	88.2%	64.3%	**78.0%**	**93.5%**	**60.7%**
(5) Liquids and solids	79.4%	91.8%	66.5%	**79.7%**	**95.7%**	**63.0%**

The VMG 'definitely acceptable' criteria for specificity, sensitivity and overall accuracy were - specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%.

Looking first at the data generated by testing 107 chemicals with the original ring trial using SOP V6.0:

1. Using the 50% cell viability cut-off: 74.5% specificity, 81.4% sensitivity and 77.9% overall accuracy values were reported when considering all chemicals, irrespective of physical state (solid/liquid).

[3] The VMG set used an unusual convention to assign acceptable values: they devised a three category system – "definitely acceptable", and "definitely unacceptable", with the values between these figures being borderline acceptable.

a. For liquid chemicals only the values were: 69.8% specificity, 96.2% sensitivity, and 82.5% overall accuracy.

b. For solid chemicals only the values were: 79.7% specificity, 66.7% sensitivity, and 73% overall accuracy.

The sensitivity and overall accuracy values for solid chemicals, and sensitivity value for solids and liquids taken together, failed to meet the VMG "definitely acceptable" values.

2. Using the 60% cell viability cut-off: 70.5% specificity, 87.6% sensitivity and 79% overall accuracy values were reported when considering all chemicals irrespective of physical state (solid/liquid).

a. For liquid chemicals only the values were: 66.7% specificity, 98.3% sensitivity, and 81.9% overall accuracy.

b. For solid chemicals only the values were: 74.8% specificity, 76.9% sensitivity, and 75.9% overall accuracy.

The sensitivity and overall values for solid chemicals, and for solids and liquids taken together, were improved but still failed to meet the VMG "definitely acceptable" values.

3. The values obtained when the solid chemicals were tested in one laboratory by the protocol optimised for solid chemicals (SOP V 8.0) were:

a. Using the 50% cell viability cut-off: 64.3% specificity, 88.2% sensitivity and 76.8% overall accuracy values were reported when considering all chemicals, irrespective of physical state (solid/liquid).

b. Using the 60% cell viability cut-off: 60.7% specificity, 93.5% sensitivity and 78% overall accuracy values were reported when considering all chemicals, irrespective of physical state (solid/liquid).

With the 50% cut-off the sensitivity value falls marginally short of the minimum acceptable value set by the VMG.

With the 60% cut-off all of the VMT minimum acceptable values were achieved.

4. By combining the data for liquid chemicals obtained during the ring trial (SOP V 6.0) with the data obtained by the one laboratory which retested the solid chemicals using the protocol optimised for solids (SOP V 8.0) the following values are obtained:

a. Using the 50% cell viability cut-off: 66.5% specificity, 91.8% sensitivity and 79.4% overall accuracy values were reported when considering all chemicals, irrespective of physical state (solid/liquid).

b. Using the 60% cell viability cut-off: 63.0% specificity, 95.7% sensitivity and 79.7% overall accuracy values were reported when considering all chemicals, irrespective of physical state (solid/liquid).

All of the values thus derived meet the minimal acceptance values set by the VMG, with the 60% cut-off giving the better values for overall accuracy and sensitivity.

The ESAC WG notes that the EpiOcular™ EIT optimised solid chemicals protocol (SOP V 8.0) was found to be at least as reproducible as the original solid chemicals protocol (SOP V 6.0), with 93.2% and 96.6% concordance of classifications (based on 59 chemicals) being obtained by Beiersdorf with the optimised protocol for the 50% and 60% cut-offs analysed in this study, respectively, as compared to 92.0% and 94.0% obtained by the same laboratory with the original protocol (based on 50 chemicals).

The ESAC WG can see no obvious physico-chemical feature shared by the chemicals that were misclassified (but we understand a more detailed analysis may be undertaken by others), and mindful of the limitations of the Draize eye test (from which the classification reference data was obtained) the ESAC WG notes that the false positive results (as judged on classification by the Draize eye test reference data) tended to have DET scores close to the values used to distinguish no-category chemicals from eye irritants by that method.

As stated above the ESAC WG considers that the accuracy figures should reflect the way decisions will be made in practice. Specifically, as chemical classification will be based upon one qualifying test run, the study results should have been evaluated to produce an accuracy range determined by the best and worst cases based on data from single runs rather than aggregating data from different test runs with the same chemical or by using a bootstrap approach to produce a more representative value. The ESAC WG commissioned additional analyses based on this approach.

The table below shows both the best- and worse-case predictive capacities derived when the data generated for the individual runs are used to produce an accuracy range based on the best- and worst-case values using the 60% cut-off value. Each cell gives the proportion, the absolute frequency and the exact 95%-Confidence Interval for the proportion. **The VMG 'definitely acceptable' criteria for specificity, sensitivity and overall accuracy (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%) are almost satisfied even using the worst-case outcomes.**

Laboratory	Protocol	Worst-case scenario (60% cut-off)			Best-case scenario (60% cut-off)		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Results obtained through post-optimisation testing (solids) and during main study (liquids)							
TOTAL (Beiersdorf + IIVS + Harlan)	**Liquids (plus #27; minus #37)**	75.5% (40/53) [0.62,0.86]	92.3% (24/26) [0.75,0.99]	59.3% (16/27) [0.39,0.78]	84.9% (45/53) [0.72,0.93]	100% (26/26) [0.87,1.00]	70.4% (19/27) [0.50,0.86]
	Liquids (plus #27; plus #37)	75.9% (41/54) [0.62,0.87]	92.3% (24/26) [0.75,0.99]	60.7% (17/28) [0.41,0.78]	85.2% (46/54) [0.73,0.93]	100% (26/26) [0.87,1.00]	71.4% (20/28) [0.51,0.87]
	Solids (plus #37)	76.3% (45/59) [0.63,0.86]	93.5% (29/31) [0.79,0.99]	57.1% (16/28) [0.37,0.76]	79.7% (47/59) [0.67,0.89]	93.5% (29/31) [0.79,0.99]	64.3% (18/28) [0.44,0.81]
	Liquids (plus #27; minus #37) And Solids (plus #37)	75.9% (85/112) [0.67,0.83]	93.0% (53/57) [0.83,0.98]	58.2% (32/55) [0.44,0.71]	82.1% (92/112) [0.74,0.89]	96.5% (55/57) [0.88,1.00]	67.3% (37/55) [0.53,0.79]
	Liquids (plus #27; plus #37) And Solids (plus #37)	76.1% (86/113) [0.67,0.84]	93.0% (53/57) [0.83,0.98]	58.9% (33/56) [0.45,0.72]	82.3% (93/113) [0.74,0.89]	96.5% (55/57) [0.88,1.00]	67.9% (38/56) [0.54,0.80]

The ESAC WG considers that the relevant predictive capacity will fall somewhere between the best- and worst-case summarised in the above table. The ESAC WG notes that even the worst-case values satisfy VMG criteria for sensitivity and accuracy, and that the worst-case specificity is close to the VMG criterion. The ESAC WG therefore considers that the VMG criteria for predictive capacity can be deemed to have been met with the 60% cut-off value.

The Cosmetics Europe Task Force Eye Irritation Report

Not applicable.

However the study co-ordinator supplied some additional information and analysis relevant to this issue at the ESAC WG meeting in March 2014.

Where comparisons can be made between HPLC/UPLC data and *in vivo* classification data, the HPLC/UPLC data did not under-estimate the *in vivo* classification. That is not surprising bearing in mind that the OD-photometry formazan values were systematically, marginally higher than those obtained by the qualified HPLC/UPLC endpoint measurement system.

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

NOTE: Are the conclusions reg. biological/mechanistic relevance and relevance in terms of making accurate predictions/measurements for the specific toxicity effect justified by the evaluated data?

EIVS

SkinEthic™ HCE

Although the test method results were highly reproducible within and between laboratories, the estimated predictive capacity with the SE and LE protocols and the Prediction Model used necessitate that predictive capacity must be improved before the method could be considered for use in any *in vitro* eye irritation testing strategy.

In consultation with the VMG the test developer has agreed to optimise the test method outside the EIVS study.

The ESAC WG believes that the VMG advice was sound, and that this was the correct decision.

EpiOcular™ EIT

All of the VMG 'definitely acceptable' criteria for specificity, sensitivity and overall accuracy defined by the VMG (specificity ≥ 60%; sensitivity ≥ 90%, overall accuracy ≥ 75%) were met by the liquids protocol and the optimized solids protocol using a 60% cut-off in the prediction model.

The Cosmetics Europe Task Force Eye Irritation Report

This study results allow comparison of the classification of the test chemicals using three RhT test methods, each of which is relevant to a different toxicological endpoint. The evidence presented suggests that for chemicals compatible with these RhT methods comparable classifications based on estimations of formazan values are obtained using both OD-photometry and HPLC/UPLC endpoint detection systems.

Furthermore, taking account of the *in vivo* classification data, the Cosmetics Europe Report has provided good evidence that with these same RhT test methods HPLC/UPLC endpoint detection measurement can be used in the case of strongly coloured chemicals which are incompatible with formazan measurement by OD-photometry.

11. Applicability domain (Module 6)

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

NOTE: When considering the objective of the study, was the study designed in a way to enable conclusions on the applicability domain, the limitations and possible exclusions (e.g. technical incompatibility of the test method with specific chemicals)?

EIVS

The EIVS study took account of the known limitation of RhT test methods that rely upon an estimation of cytotoxicity by OD-photometry measurement of the MTT reduction product formazan.

The Cosmetics Europe Task Force Eye Irritation Report

This study compared the classification of test chemicals using three RhT test methods, each of which is relevant to a different toxicological endpoint. The evidence presented suggests that for chemicals compatible with these RhT methods comparable classifications are obtained using both OD-photometry and HPLC/UPLC endpoint detection systems.

Furthermore, taking account of the *in vivo* classification data, the Cosmetics Europe Report has provided good evidence that with these same RhT test methods HPLC/UPLC endpoint detection measurement can be used in the case of strongly coloured chemicals which are incompatible with formazan measurement by OD-photometry.

The Cosmetics Europe Report also sets out plausible and reasoned arguments for why this might be applicable to other RhT test methods relying on measurement of the same endpoint.

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

NOTE: When considering the objective of the study and the data generated/analysed, have the applicability domain, the limitations and the exclusions of the method been sufficiently described?

EIVS

The EpiOcular™ EIT and SkinEthic™ HCE test methods were not developed, designed, or evaluated to differentiate between UN GHS / EU CLP Category 1 (serious eye damage) and Category 2 (eye irritation) classifications.

Gases and aerosols cannot be tested using either of the EIVS RhT test methods.

Chemicals that are MTT-reducers and/or highly coloured resulting in interference with the endpoint measurement of formazan by OD-photometry may not be appropriate for RhT test methods using this endpoint measure. See for example the findings with EIVS chemicals #33 and #96.

There is limited information available on the performance of the test methods with chemical mixtures.

EpiOcular™ EIT

The VMG and ESAC WG reviews of the chemicals that were misclassified when the EpiOcular™ EIT findings were judged against the Draize eye test reference data and identified no specific additional limitations or exclusions.

Other than the known limitations of RhT test methods that rely on the OD-photometric measurement of the MTT reduction product formazan the ESAC WG identified no specific limitation or inclusions likely to improve the test method reliability or accuracy.

The Cosmetics Europe Task Force Eye Irritation Report

The use of HPLC/UPLC would overcome the current limitation of RhT test methods with respect to problems encountered with chemicals which interfere with the measurement of formazan levels by OD-photometry.

12. Performance standards (Module 7)

NOTE: This section is only relevant in case Performance Standards have been suggested upon completion of a validation study.

12.1 Adequacy of the proposed Essential Test Method Components

NOTE: Are the proposed Essential Test Method Components adequate with respect to the key elements of the validated method as evidenced by existing information and testing data generated during the study?

EIVS

Currently there is no Performance Standard for RhT eye irritation test methods, no proposal to develop such a standard, and the VSR does not propose or provide essential test method components.

Once a suitable RhT test method for eye irritation is developed and validated, its performance characteristics (WLR, BLR, specificity, sensitivity and over all accuracy) should serve as the performance benchmarks when performance standards are developed and other RhT test methods enter the validation process.

The Cosmetics Europe Task Force Eye Irritation Report

Not relevant to this study.

However, the Cosmetics Europe Report should inform on the development of Performance Standards for RhT test methods that rely on the measurement of the MTT-reduction product formazan as the endpoint detection measurement.

12.2 Adequacy of the Reference Chemicals

NOTE: Are the Reference Chemicals adequately mapping the accuracy values of the validated method? Do they provide a representative range of the applicability domain of the test substances used during validation? Do they map an appropriate range of toxicity effects of the particular health endpoint in question? Are they commercially available?

EIVS

In the view of the ESAC WG the EIVS chemicals set could form the basis of, or starting point for, a reference chemicals set when a Performance Standard for RhT test methods for evaluating chemically induced serious eye damage/eye irritation is developed.

As noted elsewhere in this report, only a limited range of chemical mixtures have to date been identified as potential reference chemicals. However, chemical mixtures, already classified, e.g. by using additivity approach according to GHS or CLP could be integrated in a reference chemicals set.

The Cosmetics Europe Task Force Eye Irritation Report

Not relevant to this study.

13. Readiness for standardised use

13.1 Assessment of the readiness for regulatory purposes

NOTE: Is the test method ready for regulatory purposes? If yes, why? If no – what impediments currently exclude application for regulatory purposes?

EIVS

SkinEthic™ HCE

The performance data for this test method in this study confirms it is not ready for consideration for use for regulatory purposes.

EpiOcular™ EIT

The EIVS study findings justify the EpiOcular™ EIT test method (SOP V 8.0) being considered within a test strategy to determine the eye irritation potential of chemicals, specifically to detect non-irritants as part of a Top-Down or Bottom-Up approach (Scott et al, 2010).

The Cosmetics Europe Task Force Eye Irritation Report

The evidence presented tends to confirm that for RhT test methods suitably qualified HPLC/UPLC endpoint measurement is as reliable as the established OD-photometry (albeit the values obtained are systematically marginally lower than those obtained by OD-photometry), and overcomes one of the current technical limitations of these RhT test methods.

13.2. Assessment of the readiness for other uses

NOTE: Is the test method ready for other uses (e.g. screening purposes, testing to gain mechanistic insight, to generate supportive information for hazard/risk assessment).

SkinEthic™ HCE

The performance data for this test method in this study confirms it is not ready for consideration for use for these purposes. More reliable test methods are already available.

EpiOcular™ EIT

The EIVS study findings justify the EpiOcular™ EIT test method being considered as a means of identifying chemicals which are not eye-irritants in non-regulatory contexts.

The Cosmetics Europe Task Force Eye Irritation Report

Not applicable.

13.3 Critical aspects impacting on standardised use

Note: What are the factors that may impact on standardised use (in regulatory or non-regulatory settings)?

EIVS

ESAC WG notes that RhT test methods of the types evaluated in the EIVS study may be sufficiently robust to withstand minor variations in the manufacturing process (e.g. changes to the inserts) and transport.

We do however consider that there are two critical aspects to standardized use:

1. It is necessary to make proper use of control materials to properly evaluate test performance to ensure that any tissue damage in transit or storage does not compromise the data obtained.
2. The post-exposure rinsing must be properly performed to prevent over-exposure of the tissue to the test material or mechanical damage to the tissue itself.

The Cosmetics Europe Task Force Eye Irritation Report

All HPLC/UPLC systems have to be suitably qualified (FDA, 2001).

13.4 Gap analysis

NOTE: Identify, if appropriate, gaps in the study design and/or conduct that may have impacted on the stated study objective or the conclusions drawn.

EIVS

1. RhT test methods could form components of future integrated testing strategies for determining the eye irritancy potential of chemicals (Scott et al, 2010). The other components of such a testing strategy and the precise role of these RhT test methods have yet to be formally defined.
2. As mentioned above the ESAC WG notes that there is an absence of chemical mixtures (other than polymers, co-polymers, and quasi-polymers) that can be used as test chemicals within validation studies of test methods to identity chemically induced eye irritation potential. In the case of the EpiOcular™ EIT test method confidence would be increased if EIVS chemical #102 was correctly classified by additional laboratories using the test protocol optimised for solids , and EIVS chemical #102 was used as a part of training and proficiency testing chemical set when the test method is introduced to naïve laboratories, and supplementary studies confirm that the test method correctly classifies a sample of labelled products.
3. The RhT test methods evaluated in the EIVS study essentially evaluate the corneal component of chemically-induced eye irritancy: additional test methods may be required to evaluate chemically-induced eye irritation produced by other mechanisms.
4. For RhT test methods that use cytotoxicity as a surrogate measure of eye irritancy potential it is plausible that related test methods could be developed to allow histological and other evaluation of the tissue damage to provide additional insights into the mechanism of injury.
5. The ESAC WG understands, although it is not immediately relevant to the determination of chemical induced eye irritation, that 3D RhT models and test methods may be developed and evaluated that include immune competent cells (e.g. Langerhans cells). Subject to defining suitable immune cascade activation markers and prediction models, these may find applications relevant to evaluating sensitisation potential and mechanisms.
6. The 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 chemicals 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 as well as the CLP, and/or the use of mixtures already assessed and classified as eye irritants or non-irritants.

EpiOcular™ EIT

1. The ESAC WG believes confidence in the test method would have been enhanced had the performance of the SOP V 8.0 protocol optimised for solids been evaluated in more than one laboratory.
2. The VMG acceptance criteria require that no Category 1 chemicals be misclassified. Solid chemical #102, a Category 1 eye irritant, was misclassified in the ring-trial (SOP 6.0) by all 3 laboratories when the 50% cut-off was applied, and by 2 of the laboratories when the 60% cut-off was applied. Using the protocol optimised for solid chemicals and the 60% cut-off at one of the 2 laboratories that misclassified #102 using the original protocol gave the correct classification. The other laboratory that misclassified this chemical did not retest it using the optimised for solids protocol. However, chemical #102 was one of the test materials blind-tested at the MatTek laboratory to optimise the protocol for testing solid chemicals, and using SOP V 8.0, it was reliably and consistently identified as a chemical requiring classification with tissue viabilities very similar to those obtained by the one laboratory which performed the post-protocol optimisation testing of solid chemicals.

The Cosmetics Europe Task Force Eye Irritation Report

The Report does not address the potential problem of co-elution that can occur with HPLC/UPLC systems and how this is recognised and remedied.

14. Other considerations

NOTE: Please address any other consideration you might have in relation to the proposed approach under this section.

EIVS

1. The ESAC WG notes that within the SOPs used within the EIVS and Cosmetics Europe study there are references to procedural activities that should or should not take place at "room temperature". As some elements of the testing procedures are temperature dependent the ESAC WG believes that whenever possible this term should be replaced by a specified temperature range.

2. The ESAC WG endorses the view set out by the VMG that due to the variability of individual animal responses within the *in vivo* Draize eye test that there is >= 12% probability, if chemicals are retested, of chemicals currently classified as UN GHS Category 2 by the *in vivo* test being reclassified 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 irritation.

3. The SkinEthic™ HCE and the EpiOcular™ EIT test methods were developed for maximum sensitivity (ability to detect positives, with low rate of false negatives) rather than for optimal overall accuracy with balanced sensitivity and specificity (ability to detect negatives, with an acceptably low rate of false positives), in view of their likely role of identifying non-irritant chemicals.

SkinEthic™ HCE

1. The study findings failed to show that the use of the EPRA to determine chemical reactivity for protocol optimisation was useful, however there are early indications that there may be a trend for LogP values tending to equate to liquid chemical eye irritancy potential.

2. The performance values obtained with the EpiOcular test method should serve as the required performance benchmarks when the optimised SkinEthic™ HCE test method re-enters the validation process.

EpiOcular™ EIT

1. The VMG observed that a change in the insert used in the EpiOcular™ EIT test kit resulted in minor changes in reported test method performance with respect to reproducibility. The ESAC WG considers this as evidence in support of the external validity of these test methods, but believes that as and when a Performance Standard is developed consideration should be given as to whether any similar manufacturing change to a validated test method would require formal catch-up re-evaluation of the validation status of the test method with the expectation being that any such change should result in test performance equivalent or better to the original test method.

Cosmetics Europe UPLC/HPLC Study

1. The ESAC WG notes that the Cosmetics Europe HPLC/UPLC report does not describe the outcome of a study intended, designed, conducted or reported in support of the formal validation of a test method. It is an assessment of the performance of a suitably qualified bio-analytical system for the measurement of the MTT reduction product formazan.
2. This study addresses Scientific Committee on Consumer Safety (SCCS) of the European Commission Directorate General for Health and Consumers advice that *"...for coloured substances, a different endpoint, not involving Optical Density (OD) quantification, should be envisaged. Analytical methods such as HPLC/UPLC might be more appropriate to detect formazan in the in vitro assay...*" (McNamee et al. 2009).

3. HPLC/UPLC should not be required or used as a routine replacement for OD-photometric determination of formazan levels – rather, it should be used as an alternative endpoint measurement system when OD-photometry is not a suitable or reliable measurement system.

4. The development and acceptance of HPLC/UPLC should not preclude the development and consideration of other appropriate analytical methods for measuring formazan levels.

15. Conclusions on the study

NOTE: This section should present a brief summary of the study results and conclusions as described in the VSR (subsection 15.1), discuss to which extent the conclusions drawn in the study reports are justified by the study results on their own (subsection 15.2) and evaluate to which extent the conclusions are plausible with respect to other information (subsection 15.3).

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

EIVS

The RhT test methods evaluated with EIVS were developed to identify eye-irritancy potential.

EpiOcular™ EIT

With respect to the VMG conclusions and recommendations:

1. The VMG considered that the EpiOcular™ EIT original liquids protocol and the optimised solids protocol are scientifically valid (reproducible and accurate) to identify chemicals not requiring classification for serious eye damage/eye irritation under UN GHS.

2. The VMG recommended that the 60% cut-off (as submitted by the method developer) is used rather than the 50% cut-off due to higher sensitivity and accuracy, and lower number of total misclassifications.

   The ESAC WG believes the study data fully support this conclusion and recommendation.

3. The VMG considered that although a wide range of chemical types, chemical classes, molecular weights, LogP, chemical structures, etc were tested, no clear additional limitations regarding applicability could be identified. The VMG therefore recommended that EpiOcular™ EIT is considered applicable to the testing of all types of chemicals, until proven contrary.

   The ESAC WG has noted that currently there is only a limited range of chemical mixtures available for use as test substances within eye 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 irritancy potential.

4. The VMG recommended the use of positive control(s) and associated acceptance criteria that are strict enough to allow easy detection of inappropriate conduct of the assay (e.g., in the optimised SkinEthic™ HCE).

   The ESAC WG agrees.

5. The VMG concluded that the use of 2 tissue replicates in similar or modified RhT/MTT-based test method aiming at identifying chemicals not requiring classification for serious eye damage/eye irritation is statistically and scientifically justified.

   The ESAC WG believes the data (submissions by the test developer, the study findings, and the conclusions of the VMG) tends to support this, as does the principle of lean study design.

However, we do not accept that this can be generalised to all RhT test methods on the basis of the available evidence.

6. The VMG considered the current endpoint detection system using standard OD-photometry as appropriate to assess direct MTT-reducers and coloured chemicals, when tissue viability falls within the linear range of the spectrophotometer (e.g., < 140%), or when the uncorrected viabilities already identify the test chemical as requiring classification.

   The ESAC WG tends to agree.

7. For coloured chemicals interfering too strongly with the MTT-reduction assay an alternative endpoint detection system should be used (e.g., HPLC/UPLC-photometry).

   The ESAC WG agrees, and believes the suitably qualified HPLC/UPLC endpoint measurement system described in the Cosmetics Europe study report is suitable for this purpose.

The ESAC WG concludes that with respect to the results and conclusions of the EIVS study and VSR that in relation to the EpiOcular test method:

1. The EpiOcular test method does not represent a stand-alone full replacement for the currently accepted *in vivo* test method.

2. Its primary purpose will be as a means of identifying UN GHS eye irritation "no category" solid and liquid chemicals, as a tool used within an integrated testing strategy.

3. The test method has been shown to be sufficiently relevant and reliable to be considered for that purpose.

The Cosmetics Europe Task Force Eye Irritation Report

Section 5 of the Cosmetics Europe Report sets out 5 conclusions:

1. Using the approach based on the FDA guidance on validation of bio-analytical methods, HPLC/UPLC systems were established and qualified through meeting acceptance criteria of key parameters in the three participating laboratories. Any new HPLC/UPLC system could be considered for such analysis once the system has been suitably qualified. This approach is transferable and applicable to any laboratory.

2. The reproducibility of measuring formazan extract samples by HPLC/UPLC was established as being very high through evaluation of the same samples in three independent laboratories.

3. For test substances that do not exhibit colour interference nor direct MTT reduction, cell viability values are almost identical using either absorbance (OD-photometry) or HPLC/UPLC as endpoint detection systems for measurement of formazan.

4. HPLC/UPLC is capable of measuring formazan resulting in a viability measurement that can be translated into a classification for strongly coloured test substances when this is not possible by using OD-photometry absorbance for endpoint detection. Examples of this in one or more of the RhT test methods are test substance #10 that was excluded from the EURL ECVAM EIVS based on pre-checks colour interference and test substance #26 identified from the SCCS memorandum (SCCS, 2010).

5. Analysis of the use HPLC/UPLC as an endpoint detection system for coloured test substances was conducted using selected test systems for the different endpoint *in vitro* RhT test methods.

On the basis of the results obtained, it is concluded that this analytical endpoint detection system is relevant to all RhT test methods irrespective of test system and test method and can be applied to any of the other RhT test systems within OECD Test Guideline 431 (skin corrosion) and 439 (skin irritation). For eye irritation, this will be equally applicable to different test systems in an OECD Test Guideline developed based on the EIVS outcome.

The ESAC WG agrees on all five points.

The ESAC WG also concludes that with respect to the Cosmetics Europe Report:

1. The qualified HPLC/UPLC systems described in the Cosmetics Europe study report are capable of reliably and reproducibly detecting and accurately quantifying the MTT-reduction product formazan in extracts prepared using the RhT models used within the study: and that it is plausible that this will be true of if the same, or similarly qualified, HPLC/UPLC systems, are applied to these and other RhT test methods which rely on the detection and quantification of the MTT-reduction product formazan.
2. The use of HPLC and HPLC technology overcomes one of the current limitations of RhT test methods which rely on the OD-photometry detection and quantification of the MTT-reduction product formazan: that is the assessment of the relevant toxicological properties of highly-coloured chemicals which themselves produce colour interference with the OD-photometric methods of detecting and quantifying formazan.

3. When using the RhT test methods and test chemicals used for this study HPLC/UPLC was capable of consistently and reliability identifying and quantifying formazan, with the values obtained permitting an estimation of percentage cell viability that translated into a toxicity endpoint classification that did not under-classify the toxicity of the reference chemicals; and for strongly coloured test chemicals where the formazan measurement by OD-photometric methods would not have permitted an *in vitro* classification using these RhT systems.

4. The formazan levels as measured by HPLC/UPLC were systematically slightly below the values obtained when formazan levels were estimated by OD photometry.
5. With the RhT test methods used within the study the derived cell viability values were also almost identical using OD-photometric absorbance and HPLC/UPLC as the endpoint detection systems for measurement of formazan test substances not producing colour interference nor direct MTT reduction, that suitably qualified HPLC/UPLC should be appropriate as an alternative endpoint detection system with this larger class of chemicals. However, the ESAC WG notes that the values obtained using HPLC/UPLC were systematically, marginally lower than those obtained by OD photometry: there may be a case for considering whether this necessitates re-evaluation of the cut-offs developed using OD photometry.

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

All of the ESAC WG conclusions given above can be derived and argued or justified from the study results reported in the EIVS VSR and the Cosmetics Europe study report, the appendices to those reports, and the additional information supplied to or commissioned by the ESAC WG.

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

In reaching its conclusions the ESAC WG has also taken account of the larger body of information and knowledge set out in the references cited in the study documents. 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.

The Cosmetics Europe Task Force Eye Irritation Report

All of the above conclusions are plausible in the context of the study findings and supporting documents.

16. Recommendations

Note: This section should provide recommendations on the test method (e.g. further work, possible use) and their constituting elements (e.g. test system, prediction model, SOP).

16.1 General recommendations

1. 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.

2. The ESAC WG supports the VMG recommendation that *"…RhT/MTT-based test method aiming at identifying chemicals not requiring classification for serious eye damage/eye irritation (using tissues modelling the corneal epithelium), including an optimised SkinEthic™ HCE test method, use positive control(s) and associated acceptance criteria that are strict enough to allow easy detection of inappropriate conduct of the assay…".*

3. The ESAC WG, considers that when evaluating multiple tests runs for each test chemical the predictive capacity values should be calculated on the basis of the best and worse outcomes for all qualified test runs, or using a "boot-strap" approach.

4. 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 base of the properties of their components.

5. The VMG concluded that based on the highly reproducible data acquired with both EpiOcular™ EIT and SkinEthic™ HCE in EIVS using multiple exposure times and post-treatment incubation periods, it is reasonable to conclude that the reproducibility of this type of test methods is not affected by varying the exposure or the post-treatment incubation times. The ESAC WG would recommend a more conservative approach: variations in exposure and post-treatment incubation periods do alter the test results: we would therefore recommend that any such protocol variations are evaluated on a case by case basis.

6. *"The VMG considers the current endpoint detection system using standard photometry as appropriate to assess direct MTT-reducers and coloured chemicals, when the observed interference with formazan is not extreme i.e., the viabilities of the tests before correcting for direct MTT reduction and/or colour interference (with the viabilities obtained with the killed and/or living/colour controls) are within the linear range of the spectrophotometer (e.g., below 140% of the negative control) or when the uncorrected viabilities already identify the test chemical as irritant. Moreover, the VMG recommends that, for MTT reducers and colorants that strongly interfere with the MTT measurement (i.e., killed and/or living/colour control values > 50% of the negative control), three independent tests are conducted and that within test*

variability is compared to that obtained with non-interfering chemicals in EIVS. If the variability of the interfering chemical is not significantly higher than normal, correction using adapted controls should be allowed as long as the interference is not extreme (as explained above). If variability is significantly higher than normal, it is assumed that the amount of chemical retained by the tissue after exposure and post-treatment incubation varies significantly between different tests. In this situation the VMG recommends that the following rules are applied:

a. *RULE 1 – IF the mean of % Non-Specific Color (NSC) or % Non-Specific MTT reduction (NSMTT) of three qualified tests is less than or equal to (≤) 50%, THEN the chemical is considered to be compatible with the test method.*
b. *RULE 2 – IF the mean of %NSC or %NSMTT of three qualified tests is greater than (>) 50% AND their classification (I or NI) remains the same upon correction, THEN the chemical is considered to be compatible with the test method.*
c. *RULE 3 – IF the mean of %NSC or %NSMTT of three qualified tests is greater than (>) 50% AND the classification of at least one of the qualified tests changes upon correction, THEN this chemical is considered to be incompatible with the test method. In this case, use of another method (analytical or test method not relying on MTT reduction) or of a default classification as irritant should be considered."*

The ESAC WG tends to agree.

7. The VMG commented: *"For coloured chemicals interfering too strongly with the MTT-reduction assay an alternative endpoint detection system may be required (e.g., HPLC/photometry, see complementary report on Cosmetics Europe HPLC-Project). In this case, one single test (the chosen measurement system) should be sufficient independently of how strong the colour interference is, unless the chemical is also a strong MTT reducer (i.e., killed control values > 50% of the negative control) and correction from control tissues is required. If the latter occurs, the conduct of three independent tests is still recommended."*

   The ESAC WG agrees in principle BUT both endpoints might be measured and compared to *in vivo* reference data, and the Cosmetics Europe report suggests HPLC/UPLC might be useful for this. In addition, as the study showed formazan levels were maintained in samples stored for one month, supplementary HPLC/UPLC measurements could be arranged for test samples where the colour of the test material is believed to have interfered with the formazan estimation by OD photometry.

8. EURL ECVAM should consider providing supplementary statistical guidance to those planning and managing validation studies – specifically in relation to the identification of the appropriate experimental unit, calculation of appropriate confidence limits and accuracy values based on the number of test chemicals rather than the number of test runs, and means of supplementing regression analysis as a means of addressing correlation between the performance of two test methods particularly when the true value of the variable of interest is unknown (Bland and Altman, 1986).

The Cosmetics Europe Task Force Eye Irritation Report

That ESAC WG accepts the conclusions set out in the Study Report, but in addition recommends that in future where different endpoint measurement systems are being evaluated and compared that the statistical methods recommended by Bland and Altman (1986) be used instead of or in addition to standard correlation measures and regression analyses.

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

The ESAC WG offers the following recommendations for consideration.

EIVS

Not all of the relevant *in vivo* mechanisms of ocular toxicity are addressed by the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ HCE test methods. These RhT test methods are by design most relevant for the assessment of epithelial responses. If these test methods are to be used to identify non-irritants within an integrated testing strategy the ESAC EG Eye Irritation believes other test methods will be needed to assess other adverse outcome pathways.

The ESAC WG supports the VMG recommendation: *"...that all viscous, waxy and gel-like chemicals are placed in a water bath for 15 minutes at 37°C before deciding if they should be tested with the liquids or the solids protocol. Moreover, the test chemical should not be brought to room temperature before testing and should be applied directly from the water bath."* , and would also recommend that within SOPs the term "room temperature" be replaced by an appropriate specified temperature range.

SkinEthic™ HCE

The ESAC WG concludes, as does the VMG, that the SkinEthic™ HCE TS LE and the SE protocols as used in the validation study are not considered suitable to identify chemicals not requiring classification for serious eye damage/eye irritation. We note the test method proved to be highly reproducible within and between laboratories, and that the EPRA data was not useful, support the VMG recommendations for optimisation of the SkinEthic™ HCE test method, including considering developing and evaluating different protocols for liquid chemicals and solid chemicals. We note that the test method optimisation is not being done within EIVS.

The ESAC WG agrees with the VMG conclusion that the use of EPRA to orient chemicals to the LE (non-reactive) or SE (reactive) protocols as proposed in the SkinEthic™ HCE TS is not useful in the context of eye irritation with this RhT test method.

On the basis of the independent statistical analysis of the data acquired in EIVS with SkinEthic™ HCE SE and LE protocols using three replicate tissues per test the VMG concluded and recommended the use of two tissue replicates per test in any similar or modified RhT/MTT-based test method aiming at identifying chemicals not requiring classification for serious eye damage/eye irritation is statistically and scientifically justified. The ESAC WG would recommend a more conservative approach – with the appropriate minimum number of replicates being identified for each RhT test method on a case by case basis.

The Cosmetics Europe Task Force Eye Irritation Report

Mindful that OD-photometry estimation of formazan levels is systematically slightly higher than formazan values estimated from HPLC/UPLC measurements, consideration should be given as to whether and to what extent this should impact on the prediction models used with different endpoint measurement systems.

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. 2014 Mar;88(3):701-23.

Balls, M., Blaauboer, B.J., Fentem, J.H., *et al.* (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of the ECVAM Workshop 5. *Altern. Lab. Anim.*23, 129-147.

Bland JM, Altman DG. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **i**, 307-310.

Cormier EM, Parker RD, Henson C, Cruse LW, Merritt AK, Bruce RD, Osborne R. Determination of the intra- and interlaboratory reproducibility of the low volume eye test and its statistical relationship to the Draize eye test. Regul Toxicol Pharmacol. 1996;23:156–161.

Eskes, C., Bessou, S., Bruner, L., et al. (2005). Eye irritation. *Altern. Lab. Anim.* 33, Suppl. 1, 47-81.

Food and Drug Administration (2001) Guidance for Industry: Bioanalytical Method Validation. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research.

Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A., Zuang V. (2004). A modular approach to the ECVAM principles on test validity. *Altern. Lab. Anim.* 32, 467-472.

Marzulli F, Ruggles D. 1973. Rabbit eye irritation test: collaborative study. Journal of the Association of Analytical Chemists 56:905-914.

Mc Namee P, Faller IC, Goebel C, Pfuhler S, Tieber T, Egham I. Use of HPLC/UPLC instead of photometry for evaluation of MTT in in vitro RhT assays for irritation; assessment of colored materials, in Altex, volume 26, special issue, Abstracts 7th World Congress, Rome, 2009, p281.

Mosmann, T. Rapid colorimetric assay for cellular growth and survival:application to proliferation and cytotoxicity assays. J. Immunol. Methods, 65:55-63, 1983.

Nguyen DH, Beuerman RW, DeWever B, Rosdy M: Three-dimensional construct of the human corneal epithelium for in vitro toxicology. In: Alternative toxicological methods, 1st ed., chapter 14, pp.147-159, Ed. Salem H and Katz SA, CRC, New York, 2003

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.

OECD (2013), In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris

OECD (2013). Test Guideline No. 439: In Vitro Skin Irritation – Reconstructed Human Epidermis Test Method.

SCCS/1392/10: Memorandum (addendum) on the in vitro test EPISKIN™ for skin irritation testing, adopted by the SCCS during the 9th plenary meeting of 14 December 2010.

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 (2009), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Third revised edition, UN New York and Geneva.

Weil CS, Scala RA (June 1971). "Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests". *Toxicol. Appl. Pharmacol.* **19** (2): 276–360.

Appendix 1 – Description of contentious issue (Majority/Minority Position)

The ESAC Working Group has reached consensus on all but one point as outlined below.

A) BACKGROUND

- One of the acceptance criteria defined at the beginning of the validation study was that no UN GHS Category 1 chemical should be under-classified (i.e. no Category 1 false negatives).However, during EIVS, one Cat. 1 chemical was, by two of the three laboratories (Beiersdorf and IIVS), predicted as a non-irritant (i.e. false negative prediction). The chemical was disodium 2,2'-([1,1'-biphenyl]-4,4'-diyldivinylene) bis(benzene sulphonate; INCI name: DISODIUM DISTYRYLBIPHENYL DISULFONATE) (EIVS Chemical No. 102).
- These data were generated using the EpiOcular™ protocol V6.0, i.e. the version before optimisation of the protocol's specifications for testing of solids that took place in the context of EIVS.
- Following optimisation, the final protocol version (V 8.0) was then only used at Beiersdorf which produced a correct Category 1 classification with chemical No. 102. In contrast, IIVS, which had also under-classified the chemical using protocol V 6.0, did not retest the chemical using the optimised protocol.
- However, during this optimisation phase, the test developer (MatTek) tested No. 102 in a coded fashion with an exposure time of 6 hours, i.e. the exposure time that is now set in the final optimised protocol. The Mattek testing data obtained during optimisation classified EIVS chemical No. 102 as a Category 1 eye irritant (true positive). The testing data by Mattek thus were obtained (1) with the same parameters as those of the final optimised protocol and (2) under blind conditions.
- This information is not, however, to be found within the EIVS VSR, but was communicated to the ESAC WG during the course of the scientific review by EURL ECVAM. Thus, when judging only on the basis of the validation study report, apparently only one laboratory (Beiersdorf) had tested No. 102 with the optimised protocol, while in reality additional testing data by Mattek with the same protocol, but from the optimisation phase, were available but not reported in the VSR.
- When reflecting on this it should be noted that the *in vivo* classification of EIVS chemical No. 102 as a UN GHS Category 1 chemical is on the basis of persistence of conjunctival effects only: this *in vivo* classification is based on a conjunctival score of 1 in two out of three animals at day 21, by which time all the other acute tissue effects had resolved. For the purposes of the US EPA classification, which does not take account of conjunctival effects, the chemical is classified as EPA Category II.

B) WG MAJORITY POSITION: On the basis that two laboratories used the EpiOcular™ optimised for solids protocol for testing coded chemicals and produced the correct classification for chemical No. 102, a majority of the Working Group is satisfied that there is sufficient information to conclude that protocol V 8.0 is capable of correctly classifying this chemical and that false predictions of Category 1 substances are unlikely to occur (c.f. EIVS acceptance criterion).

C) WG MINORITY POSITION: One WG member considers that, as IIVS did not subsequently re-test using the amended protocol, the relevant EIVS acceptance criterion cannot be deemed to have been met, and support for the use of the test method for regulatory purposes must be conditional (i.e. depending on additional testing of No. 102 in another laboratory). Specifically, support for the use of the EpiOcular™ protocol V8.0 for regulatory use (for testing solids) must be conditional upon one or more laboratories correctly classifying EIVS chemical No. 102 as UN GHS Category 1. In addition confidence in the reliability of the test method would be increased if EIVS chemical No. 102 is used as part of a required training and proficiency protocol for transferring the test method to other laboratories, and supplementary studies confirm that the protocol can correctly classify a sample of labelled products.

LA-NA-28173-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:10.2787/043697
ISBN 978-92-79-63195-5