EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
The European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
Test Submission Template (TST)

TABLE OF CONTENTS

Submission of confidential information

It should be noted that during the evaluation phase of test method submissions, EURL ECVAM will share information contained in the TST with its Advisory Structure, e.g. the ECVAM Scientific Advisory Committee (ESAC), and to the extent possible, with European regulatory authorities. Whoever gets access to the TST will be bound to treat all information submitted to ECVAM in a confidential manner. Nevertheless, ECVAM procedures involve that Background Review Documents on the evaluation of the test method and its outcome are compiled and made publicly available which might imply disclosing information/data submitted with the TST together with those generated during validation.

Therefore, if you consider some of the information provided in the TST as confidential (e.g. confidential business information (CBI)), please clearly identify below those paragraphs where confidential information has been entered. Briefly describe (not more than 100 words per paragraph) why this information is considered confidential. Please specify if such confidential information may be made publicly available.

No confidential information provided.

List of abbreviations

Please list the abbreviation used in the submission

BCOP:	Bovine Corneal Opacity and Permeability
BLR:	Between Laboratory Reproducibility
C:	Classified
CAS RN:	Chemical Abstracts Service Registry Number
Cat 1:	UN GHS/EU CLP classification for chemicals causing irreversible effects on the eye/serious damage to the eye
Cat 2:	UN GHS/EU CLP classification for chemicals causing reversible effects on the eye/eye irritation, sub-categorised in 2A (irritant to eyes, eye effects are not fully reversible within 7 days of observation) and 2B (mildly irritant to eyes, eye effects fully reversible within 7 days of observation)
CC:	Conjunctival chemosis
CI:	Confidence Interval
CM:	Cytosensor Microphysiometer
CO:	Corneal Opacity
Conj:	CR and/or CC
CR:	Conjunctival redness
CRL:	Charles River Laboratories
DRD:	Draize eye test Reference Database
EIT:	Eye Irritation Test
EU CLP:	European Union Regulation on Classification, Labelling and Packaging of chemicals implementing UN GHS in the EU
EURL ECVAM:	European Union Reference Laboratory for Alternatives to Animal Testing

HCE:	SkinEthic™ Human Corneal Epithelium
HPLC/UPLC:	High/Ultra Performance Liquid Chromatography
ICCVAM:	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE:	Isolated Chicken Eye
IR:	Iritis
IRE:	Isolated Rabbit Eye
LO:	L'Oréal
MTT:	(3-[4,5-dimethylthiazol-2-yl]-2,5- diphenyltetrazolium bromide)
NC:	Not Classified
NgC:	Negative Control
No Cat:	Chemicals not classified for serious eye damage/eye irritation under UNGHS/EUCLP
NICEATM:	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
No Cat:	Chemicals not classified for serious eye damage/eye irritation under UN GHS/EU CLP
NSC_{killed}:	Non Specific Colour on killed tissues
$NSC_{living:}$	Non Specific Colour on living tissues
NSMTT:	Non-specific reduction of MTT
OD:	Optical Density
OECD:	Organisation for Economic Co-operation and Development
PBS:	Phosphate-Buffered Saline
PBS^{-}:	Phosphate-buffered saline Ca^{2+} and Mg^{2+} free
PC:	Positive control
Pers:	Persistence
REACH:	EU Registration, Evaluation, Authorisation and Restriction of Chemicals
RhCE:	Reconstructed human Cornea-like Epithelium
RhT:	Reconstructed human Tissue
SCNM:	Study Criteria Not Met
SkinEthic™ HCE EITL:	Eye Irritation Test Liquid protocol
SkinEthic™ HCE EITS:	Eye Irritation Test Solid protocol
S.O.P.:	Standard Operating Procedure
STE:	Short-Time Exposure
TG:	Test Guideline
UN GHS:	United Nations Globally Harmonized System of Classification and Labelling of Chemicals
VITO:	Flemish Institute for Technological Research
VMG:	Validation Management Group
VRM:	Validation Reference Method
WLR:	Within Laboratory Reproducibility

Submitter's request to ECVAM

If you seek to enter the ECVAM validation process with this test submission, please indicate at what stage in the validation process you think the test method should enter. For a description of the validation process and the different types of validation, see the Explanatory Note on the TST. Otherwise, please specify any other type of request to ECVAM by filling the appropriate text box.

	Please tick where appropriate
Prevalidation	
Prospective validation	
Retrospective validation	
Validation based on PS	
Peer-review	✓ Peer-review of SkinEthic™ HCE Eye Irritation Test (EIT), a test method with a wide applicability domain for liquids (EITL: Eye Irritation Test Liquid protocol) and solids (EITS: Irritation Test Liquid protocol), which has been assessed in a multi-laboratory trial involving 3 laboratories.
Other	Please specify your exact request to ECVAM:

1. INFORMATION ON VALIDATION MODULES

The ECVAM modular approach to validation breaks down the information required for establishing the scientific validity of a test method into independent modules allowing greater flexibility with regard to the time of procuring information during a validation study.

1.1 MODULE 1: TEST DEFINITION

1.1.1 Human health, environmental or other biological effects addressed by the test method

Please describe which human health, environmental or other biological effects the test method will address, e.g. whether it can be used to predict repeated-dose toxicity in humans, to predict fish chronic toxicity, to measure bioaccumulation, to be used in quality control of immunobiologicals etc.

The test method can be used to predict the irritant potential of a substance on human eyes by being able to correctly identify chemicals not requiring classification and labelling for eye irritation or serious eye damage according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS).

1.1.2 Scientific basis – biological and/or mechanistic relevance

Please describe the relationship between the test method and the effect of interest. This may include a reference to a) the mechanistic relevance (e.g. the mechanism of action) and/or b) the biological relevance (e.g. how well the test method models the target organ) and/or c) an empirically observed (correlative) relationship to the effect of interest.

Eye irritation is a local, inflammatory response of the eye to direct injury caused by the application and contact of an irritant chemical. The corneal epithelium, along with the conjunctiva, are the tissues in direct contact with chemical entering the eye.
Chemical-induced serious eye damage/eye irritation, manifested *in vivo* mainly by corneal opacity, iritis, conjunctival redness and/or conjunctival chemosis, is the result of a cascade of events beginning with penetration of the chemical through the cornea and production of damage to the cells. Cell damage can occur by several modes of action, including: cell membrane lysis (e.g., by surfactants, organic solvents); coagulation of macromolecules (particularly proteins) (e.g., by surfactants, organic solvents, alkalis and acids); saponification of lipids (e.g., by alkalis); and alkylation or other covalent interactions with macromolecules (e.g., by bleaches, peroxides and alkylators) (Scott et al, 2010) (Hackett and McDonald, 1991) (Fox and Boyes, 2008). Damage to the corneal epithelium resulting from exposure to chemicals may compromise tissue function, and can result in various effects ranging from mild irritation, to the loss of cornea transparency or blindness.

It has been shown that cytotoxicity plays an important, if not the primary, mechanistic role in determining the overall serious eye damage/eye irritation response of a chemical

regardless of the physiochemical processes underlying tissue damage (Jester et al., 1998a) (Maurer et al, 2002). Moreover, the serious eye damage/eye irritation potential of a chemical is principally determined by the extent of initial injury (Jester et al., 2001), which correlates with the extent of cell death (Jester et al., 1998a) and with the extent of the subsequent responses and eventual outcomes (Jester et al., 1998b). Thus, slight irritants generally only affect the superficial corneal epithelium, the mild and moderate irritants damage principally the epithelium and superficial stroma and the severe irritants damage the epithelium, deep stroma and at times the corneal endothelium (Maurer et al, 2002) (Jester, 2006).

Serious eye damage refers to the production of tissue damage in the eye, or serious physical decay of vision, following application of a test chemical to the anterior surface of the eye, which is not fully reversible within 21 days of application, as defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS, 2013). Also according to UN GHS, eye irritation refers to the production of changes in the eye following the application of a test chemical to the anterior surface of the eye, which are fully reversible within 21 days of application. Test chemicals inducing serious eye damage are classified as UN GHS Category 1, while those inducing eye irritation are classified as UN GHS Category 2. Test chemicals not classified for eye irritation or serious eye damage are defined as those that do not meet the requirements for classification as UN GHS Category 1 or 2 (2A or 2B) i.e., they are referred to as UN GHS No Category.

References:

Fox, D.A., and Boyes, W.K. (2008). Toxic responses of the ocular and visual system. In Cassaret and Doull's Toxicology: The basic science of poisons (ed. C.D. Klaassen), 7th edn, pp. 665–697. Withby, ON, Canada: McGraw-Hill Ryerson.

Hackett, R.B., and McDonald, T.O. (1991). Eye Irritation. In Advances in Modern Toxicology: Dermatoxicology (ed. F.N. Marzulli & H.I. Maibach), 4th edn, pp. 749–815. Washington, DC, USA: Hemisphere Publishing Corporation.

Jester, J.V., Li, H.F., Petroll, W.M., Parker, R.D., Cavanagh, H.D., Carr, G.J., Smith, B., and Maurer, J.K. (1998a). Area and depth of surfactant induced corneal injury correlates with cell death. Invest. Ophthalmol. Vis. Sci. 39, 922–936.

Jester, J.V., Petroll, W.M., Bean, J., Parker, R.D., Carr, G.J., Cavanagh, H.D., and Maurer, J.K. (1998b). Area and depth of surfactant-induced corneal injury predicts extent of subsequent ocular responses. Invest. Ophthalmol. Vis. Sci. 39, 2610–2625.

Jester, J.V., Li, L., Molai, A., and Maurer, J.K. (2001). Extent of corneal injury as a mechanistic basis for alternative eye irritation tests. Toxicol. in Vitro 15, 115–130.

Jester, J.V. (2006). Extent of corneal injury as a biomarker for hazard assessment and the development of alternative models to the Draize rabbit eye test. Cutan. Ocul. Toxicol. 25, 41–54

Maurer, J.K., Parker, R.D., and Jester, J.V. (2002). Extent of corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. Reg. Tox. Pharmacol. 36, 106-117.

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

UN (2013). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30, Fifth revised edition, New York and Geneva: United Nations. Available at: [http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf].

1.1.3 Intended purpose of the test method

Please describe the intended purpose (i.e. practical use) of the test method (e.g. regulatory safety testing under REACH, for cosmetics, for pharmaceuticals, or non-regulatory applications).

The eye can be subjected by accident to contact with diverse chemicals among which cosmetic products and their ingredients. The evaluation of eye irritation potential for cosmetic products and ingredients is therefore essential in order to prevent the safety in case of any accidental exposure.

Validation and regulatory acceptance of in vitro test methods for serious eye damage/eye irritation continues to be of high priority in the EU considering legislations such as the EU Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation, the Cosmetics Regulation. While the former implemented strong legal measures that discourage testing in animals, the latter banned such tests altogether for all human health effects as well as the marketing of cosmetic products containing ingredients tested on animals for several health effects, including serious eye damage/eye irritation, as of March 2009.

The SkinEthic™ Human Corneal Epithelium (HCE) model consists of a human corneal epithelium constructed with human immortalized corneal epithelial cells. The constructed tissue obtained is a multilayered epithelium resembling to the *in vivo* epithelium with similar thickness, morphology and histology. The test method consists of a topical exposure of the test chemical onto the human corneal epithelium model SkinEthic™ HCE, followed by cell viability measurement correlated to its *vivo* irritation potential.
The measurement of viability of the SkinEthic™ HCE tissue construct after topical exposure to a chemical to discriminate chemicals not requiring classification for serious eye damage/eye irritancy (UN GHS No Category) from those requiring classification and labelling (UN GHS Categories 1 and 2) is based on the assumption that all chemicals inducing serious eye damage or eye irritation will induce cytotoxicity in the corneal epithelium.

The intended purpose of the test method is the identification of the irritant potential on liquid / solid chemicals (substances and mixtures) that represents an important component of the safety assessment of any new chemical.

1.1.4 Evidence demonstrating the need for the test method

Please summarise the need for the test method in relation to existing methods in the context of regulatory testing (relevant test guidelines and legislation) and the 3Rs. If applicable, describe whether the test method represents an improvement compared to an existing method. Possible improvements include a) better information (e.g. higher accuracy or addressing a mechanism of action), b) effectiveness in terms of throughput (e.g. amenable to high-throughput testing), c) cost.

Current OECD guideline 405 for acute eye irritation is based on the method described by Draize. (Draize et al. 1944), and generally involves a standardized protocol for instilling materials onto the corneal and conjunctiva of rabbits and scoring the effects (OECD, 2012).

The 7th amendment of the European Cosmetic Directive ban animal testing for several tests including eye irritation classification of ingredients for cosmetic purpose since March 2009. It is therefore necessary to set up and validate *in vitro* approaches in order to predict eye irritation without the need of animals.

3-Dimensionnal *in vitro* tissue cultures sustained by adapted technologies allow the testing of chemicals with wide physicochemical properties in conditions similar to *in vivo* exposure. Many efforts have been made in order to find reliable and relevant predictive methods using reconstructed tissues (Alépée et al., 2013; Pfannenbecker et al., 2013). Besides, many other complementary *in vitro* or *ex vivo* methods for eye irritation are currently under optimization (Wilson et al., 2015). Despite encouraging results, no test was found to be capable of completely replacing the *in vivo* Draize eye test.

It is currently generally accepted that, in the foreseeable future, no single *in vitro* test method will be able to fully replace the *in vivo* Draize eye test (Draize et al., 1944) (OECD, 2012) to predict across the full range of serious eye damage/eye irritation responses for different chemical classes. However, strategic combinations of several alternative test methods within (tiered) testing strategies such as the Bottom-Up/Top-Down approach may be able to fully replace the Draize eye test (Scott et al, 2010). The Bottom-Up approach is designed to be used when, based on existing information, a chemical is expected not to cause sufficient eye irritation to require a classification, while the Top-Down approach is designed to be used when, based on existing information, a chemical is expected to cause serious eye damage.

The SkinEthic™ HCE is recommended to identify chemicals that do not require classification for eye irritation or serious eye damage according to UN GHS (UN GHS

No Category) without further testing, within a testing strategy such as the Bottom-Up/Top-Down approach suggested by Scott et al. e.g., as an initial step in a Bottom-Up approach or as one of the last steps in a Top-Down approach. However, the SkinEthic™ HCE EITL (Eye Irritation Test for Liquids) and EITS (Eye Irritation Test for Solids) is not intended to differentiate between UN GHS Category 1 (serious eye damage) and UN GHS Category 2 (eye irritation). This differentiation will need to be addressed by another tier of a test strategy. A chemical that is identified as requiring classification for eye irritation/serious eye damage with SkinEthic™ HCE EITL/EITS will thus require additional testing (*in vitro* and/or *in vivo*) to establish a definitive classification.
Considering the low prevalence of ocular irritants and chemicals inducing serious eye damage, the reduction in animal use achieved with the adoption of SkinEthic™ HCE EIT will be very significant.

References

Alépée N, Bessou-Touya S, Cotovio J, de Smedt A, de Wever B, Faller C, Jones P, Le Varlet B, Marrec-Fairley M, Pfannenbecker U, Tailhardat M, van Goethem F, McNamee P. (2013).Cosmetics Europe multi-laboratory pre-validation of the SkinEthic™ reconstituted human corneal epithelium test method for the prediction of eye irritation.Toxicol In Vitro. 27(5):1476-88

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

OECD (2012). Acute Eye Irritation/Corrosion. OECD Guideline for Testing of Chemicals No. 405. Organisation for Economic Co-operation and Development, Paris. Available at:
[http://www.oecd-ilibrary.org/environment/test-no-405-acute-eye-irritation-corrosion_9789264185333-en].

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., and Van Goethem, F. (2013). 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.

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

Wilson S.L, Ahearne M., Hopkinson A. (2015). An overview of current techniques for ocular toxicity testing. Toxicology 327 (2015) 32–46

1.1.5 Technical specifications

1.1.5.1 Protocol(s) of the test method

Please append the protocol(s) as used to generate the data submitted in this TST as Attachment 1a. Furthermore please attach the protocol(s) in the DB-ALM protocol[1] *format as Attachment 1b. Please specify how the protocol(s) relate to any existing DB-ALM protocol, if applicable, i.e. whether it is identical or differing from the DB-ALM protocol. In case of deviations, please outline the major components and/or steps that differ.*

L'Oreal R&I developed the SkinEthic™ HCE Eye Irritation Test (EIT), a test method with a wide applicability domain for liquids (EITL: Eye Irritation Test Liquid protocol) and solids (EITS: Irritation Test Solid protocol), which was then assessed in a multi-laboratory trial involving 3 laboratories.

Eye Irritation Test Liquid protocol (EITL)
SkinEthic™ HCE tissues (0.5 cm²) were topically exposed to 30 µL of undiluted liquid chemical for 30 ± 2 minutes at 37°C at 5% CO_2 in a humidified incubator (standard culture conditions). Two tissues were used per test chemical (NgC, PC, or chemical). After 30 minutes treatment, tissues were rinsed at least two times with 10 mL of PBS to remove the residual test chemical from the tissue surface. After rinsing, the tissues were immersed into 1.5 mL fresh maintenance medium (750 µL underneath and 750 µL topically) for a 30 ± 2 minute incubation period in standard culture conditions. After the incubation period, duplicate tissues were assessed for tissue viability.

Eye Irritation Test Solid protocol (EITS)
SkinEthic™ HCE tissues (0.5 cm²) were topically exposed to 30 mg ± 2 mg of solid test chemical for 4 hours ± 5 minutes at 37°C at 5% CO_2 in a humidified incubator (standard culture conditions). If necessary, the test chemical was first crushed to a very fine powder and before applying the chemical on the tissue, 30 µL PBS was pipetted onto the epithelium to improve optimal contact of the powder with the epithelium. Two tissues were used per test item (NgC, PC, or chemical). After 4 hours treatment, tissues were rinsed with 25 mL of PBS to remove the residual test chemical from the tissue surface. After rinsing, the tissues were immersed into 4 mL fresh maintenance medium at room temperature for 30 minutes ± 2 minutes. At the end of the post-soak immersion, tissue were transferred to a new 6-well plate containing 1 mL of maintenance medium and were incubated for 18 hours ± 30 minutes at standard culture conditions. After the incubation period, duplicate tissues were assessed for tissue viability.

The SOPs as used to generate the data submitted are appended in Attachment 1a for liquid test chemicals and 1b for solids chemicals. The attachment 1c and 1d correspond to the DB-ALM protocol format for liquids and solids, respectively. DB-ALM protocols had been written *de novo*.

[1] The so-called "DB-ALM" protocols are ECVAM's reporting format for the dissemination of a test method protocol via the ECVAM database web service on alternative methods http://ecvam-dbalm.jrc.ec.europa.eu/

The protocol(s) should contain a list and description of the materials needed, a description of what is measured and how (specification of parameters/endpoints measured), as well as a (preliminary) prediction model (PM), if applicable. In addition, a brief description of the following key elements of the protocol(s) should be provided in the following paragraphs:

a) Brief description of the test system

Briefly specify the test system used (e.g. tissue model, specific cells grown to confluence, etc.).

The SkinEthic™ HCE model uses immortalized human corneal epithelial cells cultured in a chemically defined medium. When cultured at the air–liquid interface on a permeable synthetic membrane insert, the epithelial cells stratify and differentiate into a 3-dimensional tissue which bears close resemblance to normal human corneal epithelium. The tissue construct contains at least four viable layers including columnar basal cells, transitional wing cells and superficial squamous cells. Other structural features of corneal tissue, such as the presence of mature desmosomes and intermediate filaments, as well as the expression of corneal specific cytokeratin 64 kD (K.3) similar to that of the normal human corneal epithelium, have been described (Nguyen et al., 2003).

References

Nguyen, D.H., Beuerman, R.W., de Wever, B., Rosdy, M. (2003). Three-dimensional construct of the human corneal epithelium for in vitro toxicology. Salem, H., Katz, S.A. (Eds.), Alternative Toxicological Methods. CRC Press, pp. 147–159.

b) Parameters and endpoints measured

Please specify the parameter(s) (e.g. optical density) and endpoint(s) (e.g. cell viability, EC_{50}) measured to make predictions and describe how this parameter(s) is/are measured.

Tissue viability in SkinEthic™ HCE EIT is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue MTT formazan salt that is quantitatively measured after extraction 38 from tissues (Mosmann, 1983). The viability of the tissue following exposure to a test chemical is determined in comparison to tissues treated with the negative control substance (considered as 100% of viability), and is then used to predict the eye hazard potential of the test chemical.

The MTT assay is a standardised quantitative method that should be used to measure tissue viability. The MTT assay is performed immediately following the post-exposure incubation period. In the SkinEthic™ HCE EIT, the 2 tissue constructs are placed in 0.3 mL of MTT solution at 1 mg/mL for 180±15 minutes at standard culture conditions. The vital dye MTT is reduced into a blue MTT formazan precipitate by the viable cells of the tissue constructs. The precipitated blue MTT formazan product is then extracted from the tissue using 1.5 mL of isopropanol. Tissues tested with liquid test chemicals were

extracted from both the top (750 µL) and the bottom (750 µL) of the tissues. To minimise any potential contamination of the isopropanol extraction solution with any chemical that may have remained on the tissue, tissues tested with solid test chemicals were extracted from the bottom (1.5 mL) of the tissue only. Tissues tested with liquid chemicals that are not readily washed off may also be extracted from the bottom of the tissue only. The corresponding negative and controls were treated similarl y to the tested chemicals. The extracted MTT formazan was quantified by a standard absorbance (OD) measurement at 570 nm using a filter band pass of maximum ± 30 nm in the multi-laboratory trials involving 3 laboratories).

In parallel to the SkinEthic™ HCE EIT validation study, L'Oreal R&I conducted a complementary study on the use of HPLC-photometry to address the current limitations of Reconstructed human Tissue/MTT-based assays for the evaluation of MTT interfering and/or coloured chemicals interfering with the MTT reduction assay. As such, the extracted MTT formazan was quantified by using an HPLC-spectrophotometry procedure to demonstrate its relevance (see paragraph 3.1) (Alépée et al., 2015).

References

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

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

c) Quality criteria applied to the test system

Please specify the quality criteria applied to the test system (e.g. appropriate stratification and barrier function of a reconstructed human epidermis in each lot/batch).

Each SkinEthic™ HCE tissue batch should meet the following specifications:
- Number of cell layers ≥ 4 on histological sections
- Optical Density at 570 nm ≥ 0.7 for cell viability (MTT reduction assay)

d) Positive control, negative control, benchmarks

Please indicate all concurrent controls used and specify if they are used as acceptance criteria for the run (a run consists of one or more test items tested concurrently with a positive and a negative control). Please include also the acceptance range for the control responses and, where available, any historical data used to establish the acceptance range.

Positive Control(s)

The positive control(s) should consist of a test item(s) well known to elicit a positive response in the test method.

Methyl acetate, a well-known irritant chemical is purchased from Sigma-Aldrich (CAS 79-20-9).
For each run using SkinEthic™ HCE tissue batches, acceptance criteria are defined for the test system as the mean viability of the two replicate tissues (2 values from each of the two tissues) treated with positive control i.e. methyl acetate, expressed as % of the negative control, is ≤ 30%.
Tissues treated with the positive control substance, should show this mean tissue viability ≤ 30% with either the liquids' or the solids' protocols, thus reflecting the ability of the tissues to respond to an irritant test chemical under the conditions of the test method

Negative Control(s)

*The negative control(s) can consist of treatment with the vehicle used and/or a test item known **not** to elicit a positive response in the test method.*

Ca^{2+}- and Mg^{2+}-free Dulbecco's phosphate-buffered saline (PBS), a well-known non-irritant solution, is purchased from Sigma-Aldrich.
For each run using SkinEthic™ HCE tissue batches, tissues treated with the negative control substance (PBS) should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes. The mean Optical Density at 570 nm (± 30nm) of the two replicate tissues treated with negative control (PBS), is ≥ 1.4 with an upper acceptance limit of ≤2.5.

Benchmarks *(if applicable)*

Benchmarks consist of test item(s) that produce a midrange response in the test method.

Not applicable

e) Acceptance criteria applied to the results

Please specify the acceptance criteria for the experimental data.

Results are expressed as mean OD, mean % viability and difference of viability between the two replicate tissues. A run was considered valid if the following criteria were met: (1) mean OD of the NgC was ≥ 1.4 and ≤ 2.5; (2) mean % viability of the PC was ≤ 30; and (3) difference of viability between two replicates tissue (NgC, PC, chemical, and all adapted controls) was ≤ 20.

For both Optical Density and HPLC-spectrophotometry endpoints, the result is accepted if:
1) The mean Optical Density (OD_{NgC}) at 570 nm (± 30nm) of the two replicate tissues treated with negative control is ≥ 1.4 with an upper acceptance limit of ≤2.5.
2) The Mean Viability of the two replicate tissues (2 values from each of the two tissues) treated with positive control, expressed as % of the negative control, is ≤ 30%.

When OD is chosen as endpoint:

The difference of viability between the two replicate tissues of a single test chemical is ≤ 20 in the same run whatever the test item (for positive control, negative control, TT and all adapted controls).

If either the negative control or positive controls included in a run fall out of the accepted ranges, the run is considered as not qualified and was repeated.
If the variability between tissue replicates of a test item falls outside of the accepted range, the test chemical should be re-tested.

In the context of the validation study, all test chemicals should have been tested in 3 independent qualified runs, and tests performed with different production tissue batches. A maximum number of two independent additional tests per test chemical was admissible (retesting) to complement missing data and achieve 3 independent qualified tests. Otherwise, the test sequence was considered as incomplete.

When HPLC/UPLC-spectrophotometry is chosen as endpoint:
Key parameters and acceptance criteria for qualification of an HPLC/UPLC-spectrophotometry system for measurement of MTT formazan extracted from tissue constructs: For details, refer to Alépée et al., 2015 and Annex III in OECD TG 492 (2015) below:

KEY PARAMETERS AND ACCEPTANCE CRITERIA FOR QUALIFICATION OF AN HPLC/UPLC-SPECTROPHOTOMETRY SYSTEM FOR MEASUREMENT OF MTT FORMAZAN EXTRACTED FROM RhCE TISSUE CONSTRUCTS

Parameter	Protocol Derived from FDA Guidance (29)(31)	Acceptance Criteria
Selectivity	Analysis of isopropanol, living blank (isopropanol extract from living RhCE tissue constructs without any treatment), dead blank (isopropanol extract from killed RhCE tissue constructs without any treatment), and of a dye (e.g., methylene blue)	$Area_{interference} \leq 20\%$ of $Area_{LLOQ}$[1]
Precision	Quality Controls (i.e., MTT formazan at 1.6 µg/mL, 16 µg/mL and 160 µg/mL) in isopropanol (n=5)	CV ≤ 15% or ≤ 20% for the LLOQ
Accuracy	Quality Controls in isopropanol (n=5)	%Dev ≤ 15% or ≤ 20% for LLOQ
Matrix Effect	Quality Controls in living blank (n=5)	85% ≤ %Matrix Effect ≤ 115%
Carryover	Analysis of isopropanol after an ULOQ[2] standard	$Area_{interference} \leq 20\%$ of $Area_{LLOQ}$
Reproducibility (intra-day)	3 independent calibration curves (based on 6 consecutive 1/3 dilutions of MTT formazan in isopropanol starting at ULOQ, i.e., 200 µg/mL); Quality Controls in isopropanol (n=5)	Calibration Curves: %Dev ≤ 15% or ≤ 20% for LLOQ Quality Controls: %Dev ≤ 15% and CV ≤ 15%
Reproducibility (inter-day)	Day 1: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 2: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 3: 1 calibration curve and Quality Controls in isopropanol (n=3)	
Short Term Stability of MTT Formazan in RhCE Tissue Extract	Quality Controls in living blank (n=3) analysed the day of the preparation and after 24 hours of storage at room temperature	%Dev ≤ 15%
Long Term Stability of MTT Formazan in RhCE Tissue Extract, if required	Quality Controls in living blank (n=3) analysed the day of the preparation and after several days of storage at -20°C	%Dev ≤ 15%

[1]LLOQ: Lower Limit of Quantification, defined to cover 1-2% tissue viability, i.e., 0.8 µg/mL.
[2]ULOQ: Upper Limit of Quantification, defined to be at least two times higher than the highest expected MTT formazan concentration in isopropanol extracts from negative controls (~70 µg/mL in the VRM), i.e., 200 µg/mL.

References

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

OECD guideline for the testing of chemicals (2015). Reconstructed Human Cornea-like Epithelium (RhCE) Test Method for Identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. DOI : 10.1787/9789264242548-en Available at: [http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788].

1.1.5.2 Data analysis and Prediction Model (PM)

Please specify how the raw data are processed and analysed. Please describe the (preliminary) procedure to translate the test method results into predictions of effects on human health or the environment. This procedure ideally should be a (preliminary) PM.

Data calculation is performed for each treated epithelium by
* Subtracting from individual tissues ODs (2 values from each of two tissues) the blank mean value (the mean OD of isopropanol solvent).
* Calculating the mean OD for each individual tissue.
* Calculating the Viability % relative to the mean OD of negative control

The OD values obtained with the replicate tissue extracts for each test chemical were used to calculate the mean percent tissue viability (mean between tissue replicates) normalised to the negative control, which is set at 100%. The percentage tissue viability cut-off value distinguishing classified from non-classified test chemicals is 60% for SkinEthic™ HCE EITL, and 50% for SkinEthic™ HCE EITS, respectively.

Results should thus be interpreted as follows:
* The test chemical is identified as not requiring classification and labelling according to UN GHS (No Category) if the mean percent tissue viability after exposure and post-exposure incubation is more than (>) 60% for EITL or > 50% for EITS. In this case no further testing in other test methods is required.
* The test chemical is identified as potentially requiring classification and labelling according to UN GHS (Category 2 or Category 1) if the mean percent tissue viability after exposure and post-exposure incubation is less than or equal (≤) to 60% for EITL or ≤ to 50% for EITS. When the final mean percent tissue viability is less than or equal (≤) to 60% for EITL or ≤ to 50% for EITS, further testing with other test methods will be required because SkinEthic™ HCE EIT cannot discriminate between UN GHS Categories 1 and 2.

1.1.5.3 Expertise required for performing the test method protocol

Please describe the level of expertise and demonstrated proficiency required by the study personnel for performing the test method protocol.

The laboratories and study personnel involved should be skilled in reconstructed tissues handling, spectrophotometry and/or HPLC/UPLC-spectrophotometry, and in the use of Excel®.

1.1.6 Test items used for developing and optimising the test method (protocol and PM)

Please indicate the number and basic physical/chemical properties of the test items used to develop and optimise the test method. Please append the full list of these test items in the form of a table, including their CAS numbers and basic physical/chemical properties and acquired data, as Attachment.

SkinEthic™ HCE EIT test method was developed within L'Oreal's laboratory using 125 chemicals including 71 liquids and 54 solids (Attachment 2).
A principal criterion for selection of test chemicals was availability of supporting complete and quality assured *in vivo* Draize eye irritation data. Moreover, the selection is limited of commercially available chemicals.
The chemicals selection is a diversify set of substances for the following parameters:
- Previously selected in EIVS validation: 35% (44/125).
- Functional group: soap/surfactant, organics (neutral, acid and base) and inorganic base.
- Several colour interfering chemicals, MTT reducers and MTT reducing colour chemical
- GHS classification: 49% of not classified (NC) and 51% of classified (C) (including 53% of Cat 1 and 47% of Cat 2).
In terms of physical state and UN GHS Categories, the 125 chemicals were distributed as described in the Table below: 34 Cat 1 (19 liquids and 15 solids), 21 Cat 2A (16 liquids and 5 solids), 9 Cat 2B (4 liquids and 5 solids) and 61 No Cat (32 liquids and 29 solids) chemicals.

	No Cat.	Cat. 2B	Cat. 2A	Cat. 1
Liquids	32 (10)	4 (3)	16 (6)	19 (7)
Solids	29 (6)	5 (2)	5 (4)	15 (7)
Total	61 (16)	9 (5)	21 (10)	34 (14)

(): number of chemicals evaluated in the multicenter studies

1.1.7 Cost and time estimates per test item

Please give an estimate of the testing cost per test item, considering that the laboratory is equipped with all necessary standard equipments and not considering labour cost. Please indicate as well an estimation of the time required to complete data acquisition for a run and specify how many chemicals can be included in a typical run.

Cost:
The consumables' cost per test item excluding labor and fixed equipment has been estimated at approximately 300 Euros (mainly the tissues' cost for NgC, PC and test item).
Contract Research Organization (CRO) testing costs are available upon request at CHARLES RIVER Laboratories or VITO.

Time required:
3 days are required to perform a run.
For SkinEthic™ HCE EITL: A run begins on Day 1 with tissue conditioning, on Day 2 with cell treatment with chemicals and the MTT viability assay and on Day 3 with data acquisition.

For SkinEthic™ HCE EITS: A run begins on Day 1 with tissue conditioning, on Day 2 with cell treatment with chemicals, and on Day 3 with the MTT viability test and data acquisition.

Applying the SkinEthic™ HCE EITL protocol, a trained experimenter can run up to 13 tests items in one series and can perform up to 3 series in one week, with the same tissue batch. The tested chemicals should be different to series 1 to 3 to consider the runs as independent.
Applying the SkinEthic™ HCE EITS protocol, a trained experimenter can run up to 13 tests item in one series and can perform up to 2 series in one week, with the same tissue batch. The tested chemicals should be different to series 1 to 2 to consider the runs as independent.

1.1.8 Occurrence of non-qualified tests

On the basis of your experience/historical data please provide:
a) an estimate (e.g. in percentage) of the frequency in occurrence of non-qualified tests (i.e. tests which do not meet the acceptance criteria),
b) an average of the number of tests which usually have to be performed to acquire the requested number of qualified tests (as described in the protocol or prediction model).

On the basis of L'Oreal experience,

a) The frequency in occurrence of non-qualified run is ≤ 1%.

b) An average of a run is required to discriminate chemicals not requiring classification for serious eye damage/eye irritancy (UN GHS No Category) from those requiring classification and labelling (UN GHS Categories 1 and 2).
A single run composed of at least two tissue replicates should be sufficient for a test chemical when the result is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements, a second test should be considered, as well as a third one in case of discordant results between the first two tests (frequency in occurrence < 1%).

1.1.9 Known limitations and drawbacks of the test methods

Please specify any known limitations of the test method beyond those addressed by the Applicability Domain (Module 6). Examples of possible limitations and drawbacks:

- *The test method requires expensive equipment and/or is based on an expensive test system*
- *Equipment / test system needed is no longer commercially available*
- *Etc.*

The SkinEthic™ HCE EIT is a simple, fast and straightforward assay that measures quantitative damage to a corneal tissue equivalent upon chemical exposure. The SkinEthic™ HCE EIT is proposed for the identification of chemicals not requiring

classification and labelling for serious eye damage/eye irritation within UN GHS (No Category) but not intended to differentiate between UN GHS/EU CLP Category 1 (serious eye damage) and UN GHS/EU CLP Category 2 (eye irritation).

Gases and aerosols have not been assessed. While it is conceivable that these can be tested using the technology, the current SkinEthic™ HCE EIT does not allow testing of gases and aerosols.

1.1.10 Intellectual Property Rights (IPRs)

Please state whether any component of the test method (e.g. protocol, test system, equipment) is patented, copyright protected, trade-marked or registered. Please specify who is holding the IPRs.

The Reconstructed Human Tissue SkinEthic™ HCE technology, associated to production of model and media are proprietary to Episkin SA, France (www.episkin.com) which is covered by IPR.
The SkinEthic™ HCE EIT protocol is however publicly available and its principles could be applicable to other similar Reconstructed human Tissue models.

1.1.11 History of test method development

Please provide any information on the process of developing the test method that might be of relevance for its validation.

In a multicenter SkinEthic™ HCE prevalidation study performed by Van Goethem and co-workers (2006), the validity of a 10-min exposure period (Short-time Exposure, SE) without post-incubation, was evaluated in four laboratories resulting in a predictive capacity of greater than 80%.
Subsequent L'Oreal in-house evaluation of this protocol with a set of about 100 cosmetic ingredients showed an increase in specificity (probability of predicting NC given the true state is No Cat) whereas the sensitivity (probability of predicting C given the true state is serious eye damage/eye irritancy (Cat 1/Cat 2)) reduced substantially (unpublished data).
In order to correctly identify the irritants which were underpredicted (false negatives) with the 10-min treatment protocol, the exposure period was prolonged to 1 hour (Long-time Exposure, LE) followed by a post-incubation period of 16 hours (Cotovio et al., 2008, 2010). Applying the UN GHS rules in combination with a threshold value of 50% viability to distinguish between NC and C, an overall predictive capacity of 82% was obtained with 81% sensitivity and 82.8% specificity with 435 substances from consumer products industry.
The assay has then undergone protocol optimization and assessment in a multi-laboratory trial managed by Cosmetics Europe (formerly COLIPA) between 2007 and 2008 leading to the refinement of the test method (Alépée et al, 2013). A high level of reproducibility within laboratory for both the SE (on 30 chemicals) and LE (on 45 chemicals) treatment procedures was observed, as well as between three laboratories.

In a next step, the validity of the SkinEthic™ HCE SE and LE protocols were evaluated with a set of 104 chemicals in a European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)/Cosmetics Europe prospective Eye Irritation Validation Study (EIVS) (Freeman et al., 2010). A total of 104 coded test chemicals were tested in both SE and LE in 3 runs and 3 replicate tissues per run, in 3 laboratories, for each protocol. None of the protocols met the acceptance criteria for predictive capacity (Barroso et al., 2014).
Finally, SkinEthic™ HCE test method was improved by L'Oreal R&I and distinct optimized protocols for the Eye Irritation Testing of Liquids (EITL protocol) and Solids (EITS protocol) were defined in 2014. Whereas in the previous test methods development the aim was to obtain a balance between sensitivity and specificity, the purpose of the current approach was to obtain a high sensitivity of at least 90%, a specificity of at least 60%, and an accuracy of at least 75%. Furthermore, none of the Cat 1 chemicals should be under-predicted in the majority of the runs.

The present TST presents the outcome of the validation exercise of the SkinEthic™ HCE EIT test method. The primary aim of this multicenter study was to assess the reliability and relevance of the test method to discriminate chemicals not requiring classification for serious eye damage/eye irritancy (No Cat) from chemicals requiring classification and labelling (Cat 1 and Cat 2).

References

Alépée N., Bessou-Touya S., Cotovio J., de Smedt A., de Wever B., Faller C., Jones P., Le Varlet B., Marrec-Fairley M., Pfannenbecker U., Tailhardat M., van Goethem F., McNamee P. (2013) Cosmetics Europe multi-laboratory pre-validation of the SkinEthic™ reconstituted human corneal epithelium test method for the prediction of eye irritation. Toxicol In Vitro 27(5), 1476-88.

Barroso, J., N. Alépée, T. Cole, C. Eskes, S. J. Freeman, R. Liška, P. McNamee, U. Pfannenbecker, A. A. Reus, C. M. Rubingh, M. W. Schaeffer and V. Zuang. (2014). EURL ECVAM – Cosmetics Europe prospective validation study of Reconstructed human Tissue-baseds test methods for serious eye damage/ eye irritation testing 9th World Congress on Animal Alternatives, Prague, Altex Proceedings (ISSN 2194-0479), Vol 3, No. 1, p. 99.

Cotovio, J., Grandidier, M.H., Lelièvre, D., Bremond, C., Flamand, N., Loisel-Joubert, S., Van Der Lee, A., Capallère, C., Meunier, J.R., Leclaire, J. (2008). The use of the reconstructed human corneal model (HCE) to assess in vitro eye irritancy of chemicals. AATEX 14, 343–350 (Special issue).

Cotovio J., Grandidier M.H., Lelièvre D., Bremond C., Amsellem C., Maloug S., Ovigne J.M., Loisel-Joubert S., Van Der Lee A., Minondo A.M., Capallere C., Bertino B., Alépée N., Tinois-Tessonneaud E., De Brugerolle de Fraissinette A., Meunier J.R., Leclaire J. (2010) In vitro assessment of eye irritancy using the Reconstructed Human Corneal Epithelial SkinEthic™ HCE model: Application to 435 substances from consumer products industry. Toxicol In Vitro 24(2), 523-37.

Freeman S.J., Alépée N., Barroso J., Cole T., Compagnoni A., Rubingh C., Eskes C., Lammers J., McNamee P., Pfannenbecker U., Zuang V. (2010). Prospective validation study of reconstructed human tissue models for eye irritation testing. ALTEX 27, Special Issue 2010, 261-266.

Van Goethem F., Adriaens E., Alépée N., Straube F., De Wever B., Cappadoro C., Catoire S., Hansen E., Wolf A., Vanparys P. (2006). Prevalidation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. Toxicol. in Vitro 20, 1–17.

1.1.12 Quality system(s) of the developing laboratory

Please state the quality system(s) in place, if any, in the laboratory that developed the test method (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

L'Oreal's R&I laboratory that developed the SkinEthic™ HCE EIT test method does not have formally implemented Good Laboratory Practices (GLP). However, the following requirements (Balls, et al., 1995a) were applied:
• Qualified personnel, and appropriate facilities, equipment and materials were available
• Records of the qualifications, training and experience, and a job description for each professional and technical individual, were maintained.
• For each study, an individual with appropriate qualifications, training and experience was appointed to be responsible for its overall conduct and for any report issued.
• Instruments used for the generation of experimental data were inspected regularly, cleaned, maintained and calibrated according to manufacturers' instructions. Records of these processes were kept, and made available for inspection on request.
• Reagents were labelled, as appropriate, to indicate their source, identity, concentration and stability. The labelling included the preparation and expiry dates, and specific storage conditions.
• All data generated during a study were recorded by the individual(s) responsible. These entries were attributable and dated.

Two additional laboratories were involved in the validation study namely CHARLES RIVER LABORATORIES (Edinburgh, United Kingdom) and VITO NV (Flemish Institute for Technological Research, Mol, Belgium). Both participating laboratories were compliant with GLP and performed the studies in accordance with the GLP standards (OECD, 1999).

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. (1995a). Practical aspects of the validation of toxicity test procedures. ECVAM Workshop Report 5. ATLA 23, 129-147.

OECD (1999) OECD Series of Good Laboratory Pactice 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].

1.2 MODULE 2: WITHIN-LABORATORY REPRODUCIBILITY (WLR)

1.2.1 Rationale for the selection of the test items used for assessing WLR

Please describe the criteria applied for selecting the test items used for assessing WLR. Please specify the extent to which these test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible, please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. Please append the full list of these items in the form of a table, including their CAS number and relevant properties, as Attachment 3.

The principal criterion for selection of test chemicals was the availability of supporting complete and quality assured *in vivo* Draize eye irritation data, for comparative evaluation of *in vitro* method predictive capacity.

The selection includes, chemicals that:
1) cover different physical states2) cover the full range of *in vivo* serious eye damage/eye irritation responses based on high quality results obtained in the reference *in vivo* rabbit eye test (Draize 1944) and the UN GHS classification system (i.e., Categories 1, 2A, 2B, or No Category) (UN GHS 2013);
3) cover the various *in vivo* drivers of classification as reported by Adriaens et al. (2014) and Barroso et al. (2015a) ;
4) cover a good and wide representation of organic functional groups;
5) have chemical structures that are well-defined;
6) are coloured and/or direct MTT reducers;
7) are commercially available; and
8) are not associated with prohibitive acquisition and/or disposal costs.

In total, 120 double blinded chemicals (60 liquids and 60 solids) were evaluated in three laboratories. The chemicals selection is a diversify set of chemicals for the following parameters:
- UN GHS classification: 48% (58/120) non-classified and 52% (62/120) of classified chemicals (including 52% of Cat 1 and 48% of Cat 2). The 120 chemicals were distributed as follows: 32 Cat 1 (16 liquids and 16 solids), 17 Cat 2A (8 liquids and 9 solids), 13 Cat 2B (8 liquids and 5 solids) and 58 No Cat (28 liquids and 30 solids) chemicals;
- 16 different functional group: organic bases, organic acids, neutral organic (the most part), inorganic bases, soap/surfactant;
- Direct MTT reduction: 11 chemicals identified as MTT reducer and 43 as non MTT reducer during Eye Irritation Validation Study (2010) considered
- Coloration: colorant (with 2 identified as color chemicals during EIVS) and non-colorant chemicals (the most part);
- Previously used in validation: 55% (66/120) tested in EIVS.

Test items used for Liquid assessment:

A total of 60 liquid chemicals representing different chemical classes were selected and are listed in Attachment 2. The chemicals were chosen to provide a balanced representation of chemicals not requiring classification (n=28) and chemicals inducing serious eye damage/eye irritation (Cat 1, n=16; Cat 2, n=16). MTT and/or colour interfering chemicals were also selected. All chemicals were sourced and blind coded independently for each laboratory and distributed to the testing laboratories by VitroScreen (Milano, Italy).

In order to enlarge the chemical diversity and to increase the dataset for evaluating the predictive capacity of the SkinEthic™ HCE EITL protocol, 45 additional chemicals were evaluated by L'Oréal in three independent runs. The chemicals represented 22 non-classified and 23 classified chemicals (Attachment 2).

In total, 105 chemicals represented 50 non-classified and 55 classified chemicals were evaluated on SkinEthic™ HCE test method for the Eye Irritation Testing of Liquids.

Test items used for Solid assessment:
A total of 60 solid chemicals representing different chemical classes were selected and are listed in Attachment 2. The chemicals were chosen to provide a balanced representation of chemicals not requiring classification (n=30) and chemicals inducing serious eye damage/eye irritation (Cat 1, n=16; Cat 2, n=14). MTT and/or colour interfering chemicals were also selected. All chemicals were sourced and blind coded independently for each laboratory and distributed to the testing laboratories by VitroScreen (Milano, Italy).
In order to enlarge the chemical diversity and to increase the dataset for evaluating the predictive capacity of the SkinEthic™ HCE EITS protocol, 35 additional chemicals were evaluated by L'Oréal in three independent runs. The chemicals represented 23 non-classified and 12 classified chemicals (Attachment 2).
In total, 95 chemicals (53 non-classified and 42 classified chemicals consisting of 24 Cat 1 and 18 Cat 2 chemicals) were evaluated on SkinEthic™ HCE test method for the Eye Irritation Testing of Solids.

References

Adriaens E, Barroso J, Eskes C, Hoffmann S, McNamee P, Alépée N, Bessou-Touya S, de Smedt A, De Wever B, Pfannenbecker U, Tailhardat M, Zuang V (2014). Retrospective analysis of the Draize test for serious eye damage/eye irritation: importance of understanding the in vivo endpoints under UN GHS/EU CLP for the development and evaluation of in vitro test methods. Arch Toxicol 88:701-723. doi: 10.1007/s00204-013-1156-8.

Barroso, J., Alépée, N., Adriaens, E., Cluzel, M., De Smedt, A., Hibatallah, J., Klaric, M., Mewes, K.R., Millet, M., Pfannenbecker, U., Templier, M., McNamee, P. (2015a). Cosmetics Europe compilation of historical serious eye damage/eye irritation in vivo data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative in vitro methods/strategies: the Draize eye test Reference Database (DRD). Submitted to Arch Toxicol.

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

UN (2013). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30, Fifth revised edition, New York and Geneva: United Nations. Available at: [http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf].

1.2.2 Assessment of within-laboratory reproducibility of experimental data

Please provide an assessment of the within-laboratory reproducibility of experimental data, i.e. the agreement among results obtained from testing the same test items over time using the same protocol in one laboratory. Please specify possible sources of variability. Please append the data in the form of tables and/or figures, as Attachment 4.

The within laboratory reproducibility (WLR) of the SkinEthic™ HCE EIT method was assessed in three laboratories. L'Oréal (L'Oréal Research & Innovation, Aulnay sous Bois, France) acted as lead laboratory, Charles River Laboratories (CRL, Edinburgh, United Kingdom) and VITO (Flemish Institute for Technological Research, Mol, Belgium) acted as naive laboratories.

PBS and Methyl acetate were used as negative control (NgC) and positive control (PC), respectively. Each laboratory tested each chemical in at least three independent experiments, with a maximum of five independent experiments, in order to obtain three valid results for each chemical.

For each laboratory the mean viability of each run for each chemical was calculated. The WLR of the independent runs was evaluated based on the concordance of predictions (C or NC). WLR was reported with the Wilson's 95% confidence intervals (CI). The Wilson CI's based on the score test provides more reliable values for small samples and estimates close to 1.0 (Agresti and Coull, 1998).

Data of the SkinEthic™ HCE EIT method (obtained with either EITL or EITS protocols) are appended as Attachment 3a.

Assessment of WLR for Liquids:

A total of 60 chemicals were tested in three laboratories. Overall, L'Oréal produced one unqualified result (Triphenyl Phosphite, CAS RN 101-02-0) over the 18 experiments that were performed. Charles River Laboratories (CRL) performed 21 experiments; one experiment was unqualified due to high deviation between the viability of the replicate NgC tissues (difference 35%). Five additional results were unqualified, three based on the difference of viability of the replicate tissues which was > 20% (p-Methyl thiobenzaldehyde, CAS RN 3446-89-7 (two runs) and [3-(2-Aminoethylamino)propyl]trimethoxysilane, CAS RN 1760-24-3). Two other results were

unqualified because of technical issues (1-ethyl-3-methylimidazolium ethylsulphate, CAS RN 342573-75-5 and 3-methyl-pentynol, CAS RN 77-75-8). VITO performed 16 experiments and obtained two unqualified results, the difference of viability of the replicate tissues was > 20% for dicaprylyl ether (CAS RN 629-82-3) and Tween 20 (CAS RN 9005-64-5). In total 54 valid independent experiments were performed by three laboratories, the mean viability of the PC (Methyl acetate) was clearly below the acceptance threshold of 30% (range: 1.4% to 12.1%) and mean OD of the NgC was within the acceptance limit (between 1.4 and 2.5).

Among the chemicals, two colourants that were also MTT reducers, were identified (acid red 92 (22-D; S2-3) 10%, CAS RN 18472-87-2, and diethylaminopropionitrile, CAS RN 5351-04-2) (Attachment 3a), requiring the use of adapted controls for the determination of non-specific colouration and MTT reduction. Seven chemicals (CAS RN 13826-35-2, 623-51-8, 51-03-6, 3446-89-7, 1760-24-3, 2365-48-2, and 17831-71-9) were identified as MTT reducers by the three laboratories. Five chemicals (CAS RN 106-91-2, 101-02-0, 609-14-3, 542-08-5, and 1310-73-2) were identified as MTT reducers by L'Oréal and CRL. Four chemicals were identified as MTT reducer by one laboratory only (CRL CAS RN 629-82-3, 78-84-2, and 4659-45-4, and L'Oréal CAS RN 625-69-4). Both uncorrected and corrected (final) viabilities were reported in Attachment 3a.

The reliability of the SkinEthic™ HCE EITL protocol was assessed in terms of concordance in predictions for three independent experiments. The results for each laboratory are presented in Attachments 3a-3b, its statistical analysis in Attachment 8a. The WLR was 95% (95% CI: 86.3%; 98.3%) for L'Oréal, 93.3% (95% CI: 84.1%; 97.4%) for VITO, and 88.3% (95% CI: 77.8%; 94.2%) for CRL.

1-Bromo-4-chlorobutane (CAS RN 6940-78-9), ethoxydiglycol (CAS RN 111-90-0) and dipropyl disulphide (CAS RN 629-19-6) resulted in discordant results in two laboratories.

1, 9-decadiene (CAS RN 1647-16-1), iso-Propyl myristate (22-A; S2-10) (CAS RN 110-27-0), 2-methyl-1-pentanol (CAS RN 105-30-6), ethyl-2-methylacetoacetate (CAS RN 609-14-3), [3-(2-Aminoethylamino)propyl]trimethoxysilane (CAS RN 1760-24-3), and tetraethylene glycol diacrylate (CAS RN 17831-71-9) and p-methyl thiobenzaldehyde (CAS RN 3446-89-7) resulted in a discordant prediction in one laboratory. The discordant predictions obtained for 1, 9-decadiene can be attributed to the viability which fluctuated around the cut-off value of 60% (between 53% and 72%) in the three laboratories. For iso-Propyl myristate (22-A; S2-10), one result (viability: 19%), deviated clearly from all other runs (viability > 93%). Charles River reported that this chemical was hydrophobic or oily resulting in spreading and rinsing difficulties. This laboratory reported the same problem for 2-methyl-1-pentanol, low viabilities were reported for all runs (< 2%), except one run (viability: 83.6%). Ethyl-2-methylacetoacetate, [3-(2-Aminoethylamino)propyl]trimethoxysilane, and tetraethylene glycol diacrylate resulted generally in viabilities < 60%, for the discordant results, the viability varied between 62% and 67.4%.

In conclusion, low variation (WLR ≥ 88%) between the independent runs was observed within the laboratories, indicating that the SkinEthic™ HCE EITL protocol is robust.

Furthermore, the lead laboratory (L'Oréal) tested 45 additional chemicals (Attachment 3a) in three independent experiments. Twenty two chemicals did not require classification *in vivo* and 23 chemicals were classified. Concordant prediction was obtained for 41 of the 45 chemicals, resulting in a WLR of 91.1% (Attachment 3c).

Statistical report is appended as Attachment 8a.

Assessment of WLR for Solids:

The reliability of the SkinEthic™ HCE EITS protocol was assessed in terms of concordance in predictions for the independent valid runs. The results for each laboratory are presented in Attachment 3d. The WLR was 96.7% (95% CI: 88.6% - 99.1%) for L'Oréal and 95.0% (95% CI: 86.3% - 98.3%) for VITO and CRL.
4-Pyrimidinol, 2,5,6-triamino-, 4-(hydrogen sulfate) (CAS RN 1603-02-7) and 1,5-naphthalenediol (CAS RN 83-56-7) resulted in discordant results in the three laboratories. The discordant predictions obtained for these two chemicals can be attributed to the viability which was in the middle range.
Methanimidamide, N'-(2,4-dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methyl (CAS RN 33089-61-1) and benzene, 1,3-dinitro (CAS RN 99-65-0) resulted in a discordant prediction in one laboratory. At VITO, one result (viability: 81.5%) obtained for methanimidamide, N'-(2,4-dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methyl (CAS RN 33089-61-1), deviated clearly from the other two runs (36.6% and 37.4%). CRL obtained a disagreement in prediction for benzene, 1,3-dinitro (CAS RN 99-65-0) , with a lower viability (33.8%) in the first run in comparison with the other two runs (79.7% and 87.5%).
In conclusion, low variation (WLR ≥ 95%) between the independent runs was observed within the laboratories, indicating that the SkinEthic™ HCE EITS protocol is robust. This means that the WLR is higher than 95%, which is the minimum value set by the VMG (Barroso et al., 2015).

The lead laboratory (L'Oréal) tested 35 additional chemicals (Attachment 3e) in three independent runs. Twenty three chemicals did not require classification *in vivo* and 12 chemicals were classified. A concordant prediction was obtained for 34 of the 35 chemicals, resulting in a WLR of 97.1% (Attachment 3e).

Statistical report is appended as Attachment 8b.

Overall assessment of WLR:

Overall, the WLR, based on the set of 120 chemicals, was 91.7% (EITL 88.3% and EITS 95.0%) for CRL, 94.2% for VITO (EITL 93.3% and EITS 95.0%) and 95.8% for L'Oreal (EITL 95.0% and EITS 96.7%). The WLR for the extended set of 200 chemicals that were tested by L'Oreal only was 95.0% (EITL 93.3% and EITS 96.8%). The SkinEthic™ HCE EIT method met the minimum values for WLR of 85% set by the VMG (EC EURL ECVAM, 2014).

For information, a WLR of 96.3%, 98.1% and 98.1% was obtained in three laboratories for the EpiOcular™ EIT Liquids protocol. For the EpiOcular™ EIT Solids protocol, a WLR of 96.6% was obtained in one laboratory, this is comparable with the WLR obtained in the current study (EC EURL ECVAM, 2014; Barroso et al., 2015).
It is important to note that a strict comparison of the WLR should not be made since the chemical sets were different; the non-commercially available proprietary chemicals evaluated using EpiOcular™ EIT test method during the EURL ECVAM/Cosmetics Europe study (EC EURL ECVAM, 2014) were not evaluated in the current SkinEthic™ HCE EIT method.

References

Agresti, A., Coull, B.A. (1998). Approximate is better than exact for interval estimation of binomial proportions. Am. Stat. 52, 119126.

Barroso, J., Alépée, N., Cole, T., Eskes, C., Freeman, S.J., Liska, R., McNamee, P., Pfannenbecker, U., Reus, A.A., Rubingh, C.M., Schaeffer, M.W., Zuang, V. (2015). EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Tissue-based test methods for serious eye damage/eye irritation testing. Poster presented at the Congress of the European Societies of Toxicology (EUROTOX), Porto, Portugal, September, 13-16, 2015. Available from: http://dx.doi.org/10.1016/j.toxlet.2015.08.507. Accessed on 05/10/2015.

EC EURL ECVAM (2014). Validation Study Report on the EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Corneal Epithelium-based test methods for identifying chemicals not requiring classification and labelling for serious eye damage/eye irritation testing. Available at: in publication.

1.2.3 Identification and discussion of outlying values

Please identify and discuss any outlying values.

See 1.2.2

1.2.4 Quality system(s) of the testing laboratory

Please state the quality system(s) in place, if any, in the testing labo ratory (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

See 1.1.12

1.3 MODULE 3: TRANSFERABILITY (TF)

1.3.1 Rationale for the selection of the test items used for assessing the TF

Please describe the criteria applied for selecting the test items used for assessing transferability. Please specify the extent to which the test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. Please append the full list of these test items in the form of a table, including their CAS number and relevant properties, as Attachment 5.

In order to assume a good transferability of the method, commercially available test chemicals were tested. It is important to note that particular chemicals were intentionally selected for the transfer such as MTT and/or color interference test chemicals.
The chemicals selection is a diversify set of substances for the following parameters:
- UN GHS classification: 6 non-classified and 12% of classified chemicals (including 50% of Cat 1 and 50% of Cat 2) representative for the range cell viability
- Functional group: organic bases, organic acids, organic salt, neutral organic (the most part), inorganic acids, soap/surfactant
- Coloration: colorant and non-colorant chemicals (the most part)
- Direct MTT reduction: MTT reducer and standard chemicals (the most part)

Using the EITL protocol, both laboratory assistants tested 9 liquid chemicals. This set of chemicals contained a strong colourant (phloxine B-acid red 92, 10 % CAS RN 18472-87-2) and an MTT reducer (butyraldehyde CAS RN 123-72-8). The strong colourant was selected in order to evaluate the crucial rinsing step procedure and the additional controls which are needed for tissue colouring chemicals. The MTT interacting chemical was chosen with the intention to perform the specific controls for direct MTT reduction of chemicals.

Using the EITS protocol, both laboratory assistants tested 9 solid chemicals. The quinacrine dihydrochloride (CAS RN 69-05-6), a colored test chemical, was chosen targeting the crucial rinsing procedure as well as requiring adapted controls. Dihydroxy 2,6 toluene (CAS RN 608-25-3), a MTT interacting test chemical, was chosen with the intention to perform the specific controls for direct MTT reduction. Additionally, tetrabromophenol blue (CAS RN 4430-25-5) was also chosen with the intent of performing all specific controls.

The list of test items is appended as Attachment 2.

1.3.2 Training required for transferring the test method

Please provide an estimation of the amount of training that is necessary to establish the test method in a naïve laboratory (i.e. a laboratory which is familiar with the techniques involved but not with the test method). If available, please append the training protocol as Attachment 6

Two training days within the test method developer's facilities is the time recommended to establish the test method in a naïve laboratory. It includes a practical training in which (i) the trainer shows the main steps of the protocols for the Eye Irritation Testing of Liquids (EITL protocol) and Solids (EITS protocol) (ii) then the study personnel perform the same steps. It also includes in depth discussions about the detailed protocols and a practical example-based workshop on data interpretation and conclusion.

The training protocols for Liquids and Solids are appended as Attachments 4a and 4b, respectively.
The training reports of the SkinEthic™ HCE EITL protocol for VITO and CRL are appended as Attachments 5a and 5b, respectively.
The training reports of the SkinEthic™ HCE EITS protocol for VITO and CRL are appended as Attachments 5c and 5d, respectively.

1.3.3 Obstacles pertaining to transferability that are specific to the test method

Please provide a summary of expected obstacles or difficulties that may impact on the transferability of the test method, e.g. level of complexity of some procedures in the protocol(s), etc.

No particular obstacles pertaining to transferability of the test method was identified.
Nevertheless, it is important to note that particular chemicals were intentionally selected for the transfer such as:
1) MTT and/or color interference test chemicals. Pre-check of each chemical has been conducted to identify direct MTT reduction as possible interference and identify chemicals able to develop a color. The additional specific adapted controls for those chemicals allowed defining $\%NSC_{living}$, %NSMTT and $\%NSC_{killed}$ values.
2) The rinsing step of strongly colorants and sticky chemicals. Attention should be paid during the rinsing procedure since it could be difficult to obtain a homogenous rinse.

1.3.4 Organisation of the transfer phase

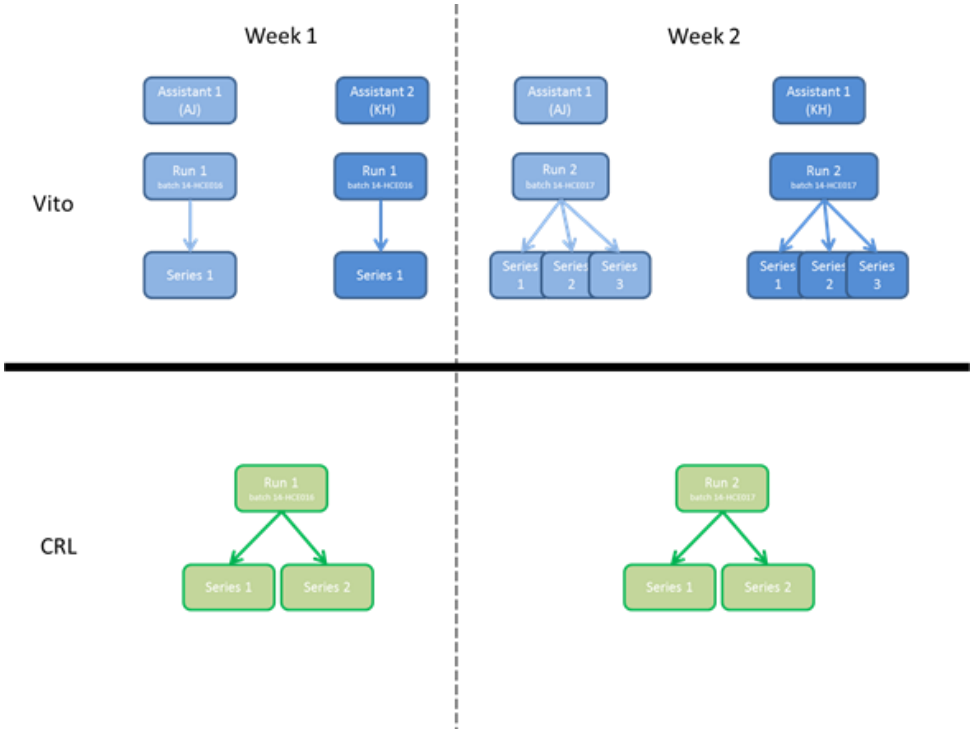
Please explain how the transfer phase was organised, including the criteria applied to assess success of transfer (see 2.3.5). If available, please append the transfer protocol as Attachment 7..

In the initial phase namely designated as training, the test method was demonstrated by L'Oreal scientists (the test method developer) who were familiar with technical details of the method. After demonstration of the protocols (EITL and EITS), participants performed the test method according to the protocols under the supervision of the L'Oreal trainer. The test method documentation was discussed and comments and suggestions for improvement of the method and documentation were collected by the trainer for later implementation into the SOP.

Eye Irritation Testing of Liquids (EITL protocol):

Two training sessions were hold from April 9th 2014 (CRL) to 16th 2014 (VITO), in which all experimental procedures, the use of Excel spreadsheets and method documentation sheets protocol were trained. The training exercises for both VITO and CHARLES RIVER Laboratories are attached to this TST in reports as Attachments 5a-5b. In the transfer phase, laboratory assistants assigned to the project performed the test method with 9 chemicals being tested 4 times (4 series). Negative and positive controls were performed in parallel in each series. Each chemical was tested in duplicates tissues. The same experiments were performed by each team in their own laboratory facilities after delivery of SkinEthic™ HCE kits.

The practical transfer phase was organised as described in the Figure below.
Briefly, during the transfer "week 1", VITO 'assistants performed a first run with one series, and 3 series were performed in order to mimic multicentre study conditions and in "week 2". In both transfer "week 1 & 2", CRL' assistants had performed two series.



Eye Irritation Testing of Solids (EITS protocol):

Before entering the transfer study, the methodology was transferred from L'Oréal to CHARLES RIVER Laboratories and to VITO within 2 days. The training session was hold from 11th and 12th December 2014 at L'Oreal, in which the SOP, all experimental

procedures, the use of Excel spreadsheets and method documentation sheets protocol were trained. The training exercises for both VITO and CHARLES RIVER Laboratories are attached to this TST in reports as Attachments 5c-5d.
In the transfer phase, laboratory assistants assigned to the project performed the test method with 9 chemicals being tested in independent series. Negative and positive controls were performed in parallel in each series. Each chemical was tested in duplicates tissues. The same experiments were performed by each team in their own laboratory facilities after delivery of SkinEthic™ HCE kits.

The practical transfer phase was organised as described below.
For VITO and CRL, the testing was performed on at least two occasions (2 series) for a total runs number of four for VITO and two for CRL, with different batches of tissues.

The criteria applied to assess success of both transfer exercises were 1) the acceptance criteria for test results on each series as described in SOP ($1.4 \leq OD_{NgC} \leq 2.5$; $Viab_{PC} \leq 30\%$; $Diff \leq 20\%$); 2) the evaluation of the within laboratory assistant reproducibility; and 3) the prediction on the classification.

The transfer reports for both VITO and CHARLES RIVER Laboratories are appended as Attachment 6a (for Liquids) and 6b (for Solids).

1.3.5 Assessment of the transferability to another laboratory

Please provide an assessment of the transferability of the submitted test method to another laboratory. If available, please append the transfer report as Attachment 8...

Identical protocols and Excel templates for data collection were transferred to each laboratory. Both naïve laboratories (CRL and VITO) received formal hands-on training in assay methodology and analysis from L'Oréal Research & Innovation.

Eye Irritation Testing of Liquids (EITL protocol):
All runs were qualified since both negative and positive controls data fulfilled the acceptance criteria requirements ($1.4 \leq OD_{NgC} \leq 2.5$; $Viab_{PC} \leq 30\%$; $Diff \leq 20\%$). These results demonstrate that the model shows a good stability and a satisfying state after shipment and adaptation to the laboratories conditions.

For VITO, the 4 series were qualified; each consisted of 9 qualified tests. The results show a good reproducibility inter-series and inter-run a very satisfactory concordance between results of the two laboratory assistants. The *in vitro* classification was for 8 test chemicals identical between series for each laboratory assistants, with only the methyl cyanoacetate showing a non-concordant classification between the series for one of the laboratory assistant.

For CRL, inter-tissue variability was observed with Triethylene Glycol Momomethyl Ether (CAS RN 112-27-6) and 1 Decanol (CAS RN 112-30-1) in a series. For all other test chemicals, and the remaining series tested, inter-tissue variability of the viability was always ≤ 20% and the test chemicals were qualified, with the acceptance criteria met, demonstrating good reproducibility. Satisfactory reproducibility was demonstrated between different series, of the same day with the same tissue batch, and between tissue batches. Therefore *in vitro* / *in vivo* concordance in classification was observed for all test chemicals.

Altogether, the results were acceptable, considering the high inter-series, inter-run and *in vivo* correlation reproducibility.

Eye Irritation Testing of Solids (EITS protocol):
For all participants from all laboratories and for all runs, negative and positive controls met always acceptance criteria qualifying all runs.

For VITO, 8 out of 9 test chemicals were consistently classified by the 2 laboratory assistants resulting in a good reproducibility and concordance. The overall *in vitro* classification was for 8 test chemicals identical between runs.
For CRL, inter-tissue variability was > 20% for a chemical in one series only. In the remaining series/runs, and for all other test chemicals, inter tissue variability was ≤ 20%.
In addition, *in vitro* classification was for all test chemicals identical between series and runs, demonstrating good reproducibility.
The between-laboratory consistency of all laboratories for final viability was good as only 1 test chemical was classified differently by the participants indicating that overall concordance between all participating laboratories was achieved during the transfer phase.

Overall, the transfer was considered successfully completed in terms of technical exchange (SOP, Excel spreadsheet discussions). As confirmed by the data, both laboratories demonstrated their proficiency in performing the SkinEthic™ HCE EIT method for the Eye Irritation Testing of Liquids (EITL protocol) and Solids (EITS protocol) sufficiently to enter the formal multicentre validation study (evaluation of 60 solid and 60 liquid coded test chemicals).

The corresponding transfer reports are appended as Attachments 6a and 6b.

1.3.6 Quality system(s) of the other laboratory

Please state the quality system(s) in place, if any, in the other laboratory (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

The study was performed complying with GLP-like principles. No internal audit was conducted due to the non-regulatory nature of the study. Documentations and raw data were kept in the test facility laboratories. All archives may be subjected to audits by the sponsor.

1.4 MODULE 4: BETWEEN-LABORATORY REPRODUCIBILITY (BLR)

1.4.1 Rationale for the selection of test items used for assessing BLR

Please describe the criteria applied for selecting the test items used for assessing BLR. Please specify the extent to which these test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible, please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. Please append the full list of these test items in the form of a table, including their CAS number and relevant properties, as Attachment 9.

The a principal criterion for selection of test chemicals is availability of supporting complete and quality assured *in vivo* Draize eye irritation data, for comparative evaluation of *in vitro* method predictive capacity. Moreover, the selection is limited of commercially available chemicals.
In total, 120 double blinded chemicals (60 liquids and 60 solids) were evaluated in three laboratories.
The chemicals selection is a diversify set of substances for the following parameters:
- GHS classification: 48% (58/120) non-classified and 52% (62/120) of classified chemicals (including 52% of Cat 1 and 48% of Cat 2). The 120 chemicals were distributed as follows: 32 Cat 1 (16 liquids and 16 solids), 17 Cat 2A (8 liquids and 9 solids), 13 Cat 2B (8 liquids and 5 solids) and 58 No Cat (28 liquids and 30 solids) chemicals;
- 16 different functional class covering organic bases, organic acids, neutral organic (the most part), inorganic bases, soap/surfactant group;
- Direct MTT reduction: 11 chemicals identified as MTT reducer and 43 as non MTT reducer during Eye Irritation Validation Study (2010) considered
- Coloration: colorant (with 2 identified as color chemicals during EIVS) and non-colorant chemicals (the most part);
- Previously used in validation: 55% (66/120) tested in EIVS and 45% of new chemicals.

1.4.2 Assessment of reproducibility

Please provide an assessment of the between-laboratory reproducibility of experimental data, i.e. the agreement among results obtained from testing the same test items using the same protocol in different laboratories. Usually at least three laboratories are requested to properly evaluate between-laboratory

reproducibility. Please specify possible sources of variability. Please append the data in the form of tables and/or figures as Attachment 10.

For each laboratory, the mean viability and standard deviation over the three independent valid runs was calculated to obtain a final classification for each chemical. The evaluation of the between-laboratory reproducibility was on the concordance of the final predictions Classified (C) or Not Classified (NC). Between laboratory reproducibility (BLR) was reported with the Wilson 95% CI.

Assessment of BLR for Liquids:

In order to assess the between-laboratory reproducibility of the method, mean viability of the three independent runs within each laboratory was calculated to determine the final classification for each laboratory. The results are presented in Attachments 7a and 8a. Fifty six of the 60 chemicals were consistently classified (NC/C) by the three laboratories resulting in a BLR of 93.3% (95% CI: 84.1% - 97.4%). The BLR for the pair-wise comparisons was 93.3% (56/60 chemicals) for L'Oréal and CRL, 95.0% (57/60 chemicals) for L'Oréal and VITO, and 98.3% (59/60 chemicals) for CRL and VITO.
1, 9-decadiene (CAS RN 1647-16-1), ethoxydiglycol (CAS RN 111-90-0), dipropyl disulphide (CAS RN 629-19-6), and p-Methyl thiobenzaldehyde (CAS RN 3446-89-7) resulted in discordant predictions.

Assessment of BLR for Solids:

Fifty eight of the 60 chemicals were consistently classified (NC/C) by the three laboratories resulting in a BLR of 96.7% (95% CI: 88.6% - 99.1%). The BLR for the pair-wise comparisons was 96.7% (58/60 chemicals) for L'Oréal and CRL and for L'Oréal and VITO, a 100% concordance was obtained between CRL and VITO. Methanimidamide, N'-(2,4-dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methyl (CAS RN 33089-61-1) and 1,5-naphthalenediol (CAS RN 83-56-7) resulted in discordant predictions.

The results are presented in Attachments 7b and 8b.

Overall assessment of BLR:

The overall BLR for the HCE EIT method, based on the set of 120 chemicals, was 95.0% (EITL 93.3% and EITS 96.7%). The BLR of the SkinEthic™ HCE EIT test method was higher than the defined minimum value of 80% set by the VMG (Barroso et al., 2015).
The EpiOcular™ EIT test method resulted in a BLR of 94.4% for the Liquids protocol and 92.0% for the Solids original protocol (EC EURL ECVAM, 2014; Barroso et al.,

2015). It is important to note that a strict comparison of the BLR should not be made since the chemical sets were different; the non-commercially available proprietary chemicals evaluated using EpiOcular™ EIT test method during the EURL ECVAM/Cosmetics Europe study (EC EURL ECVAM, 2014) were not evaluated in the current SkinEthic™ HCE EIT method.

References

Barroso, J., Alépée, N., Cole, T., Eskes, C., Freeman, S.J., Liska, R., McNamee, P., Pfannenbecker, U., Reus, A.A., Rubingh, C.M., Schaeffer, M.W., Zuang, V. (2015). EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Tissue-based test methods for serious eye damage/eye irritation testing. Poster presented at the Congress of the European Societies of Toxicology (EUROTOX), Porto, Portugal, September, 13-16, 2015. Available from: http://dx.doi.org/10.1016/j.toxlet.2015.08.507. Accessed on 05/10/2015.

EC EURL ECVAM (2014). Validation Study Report on the EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Corneal Epithelium-based test methods for identifying chemicals not requiring classification and labelling for serious eye damage/eye irritation testing. Available at: in publication.

1.4.3 Identification and discussion of outlying values

Please identify and discuss any outlying values.

Assessment for Liquids:

Four false positive results were obtained. Dipropyl disulphide (CAS RN 629-19-6), p-Methylthiobenzaldehyde (CAS RN 3446-89-7), 1,9-decadiene (CAS RN 1647-16-1) and 2-(2-ethoxyethoxy) ethanol (CAS RN 111-90-0) resulted in discordant predictions being identified as classified Cat.1/Cat.2 in at least one of the laboratory. Those four chemicals correspond with an *in vivo* UN GHS/EU CLP NC classification (with corneal opacity scores equal to 0 in all animals and all observed time points in the Draize eye test, CO = 0). We can observe that the viability obtained in was generally comprised between 48.8 to 77.7% (dots distributed around both sides of the cut off line of 60%).
Among them were organic cationic (imidazolium), neutral organic (ether and halogenated) and neutral soap / surfactant represented four different functional groups. The number of chemicals within these functional groups is too small to draw conclusions with regard to over-predictions knowing that some chemicals (e.g. CAS RN 629-03-8, CAS RN 342573-75-5, CAS RN 629-82-3, CAS RN 111-83-1, CAS RN 61788-85-0, CAS RN 9005-64-5) were predicted as non-classified in all laboratories.

Assessment for Solids:

Methanimidamide, N'-(2,4-dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methyl (CAS RN 33089-61-1) and 1,5-Naphthalenediol (CAS RN 83-56-7) resulting in discordant predictions were investigated in more detail.
1,5-Naphthalenediol (CAS RN 83-56-7) (*in vivo* Cat 2A) predicted NC in 4 out of 9 runs is a neutral organic (phenol). 1,4-Naphthalenedione (CAS RN 83-72-7), 2-hydroxy (CAS RN 83-72-7), benzoic acid, 2,4-dihydroxy (CAS RN 89-86-1), Phenol, 4-chloro-2-(phenylmethyl) (CAS RN 120-32-1), and Phenol, 4-(1,1,3,3-tetramethylbutyl) (CAS RN 140-66-9) presenting a phenol function showed concordance between the laboratories; therefore it is unlikely that the under-predictions are related with the functional group. The discordant predictions obtained for 1,5-Naphthalenediol can be attributed to the viability which was in the middle range.
Methanimidamide, N'-(2,4-dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methyl (CAS RN 33089-61-1) resulted in a discordant prediction in one laboratory. At VITO, one result (viability: 81.5%) obtained this chemical, deviated clearly from the other two runs (36.6% and 37.4%). The viability obtained also deviated clearly from the two others laboratories with a viability of 20.2% and 90.5% for L'Oreal and CRL, respectively. L'Oreal also produced one unqualified test since the difference of viability of the replicate tissues was > 20%. Variable tissue contact and/or penetration of methanimidamide, N'-(2,4-dimethylphenyl)-N-[[(2,4-dimethylphenyl)imino]methyl]-N-methyl into the tissue might potentially explain such variability. No relation was observed between the false positive and the functional group of the chemical.

1.4.4 Quality system(s) of the testing laboratories

Please state the quality system(s) in place, if any, in the testing laboratories (e.g. GLP, ISO, GCCP). For GLP-like conditions, please specify the extent and area of compliance.

See 1.1.12

1.5 MODULE 5: PREDICTIVE CAPACITY (PC)

1.5.1 Rationale for the selection of test items used for assessing PC

Please describe the criteria applied for selecting the test items used for assessing PC.. Please specify the extent to which these test items represent the range of observed effects (e.g. non-toxic to highly toxic effect). If possible, please specify the chemical classes and basic physical/chemical properties covered by your selection of test items. If some of these test items were also used to develop the prediction model, please indicate which. Please append the full list of these test items in the form of a table, including their CAS number and relevant properties, as Attachment 11.

The relevance of SkinEthic™ HCE EIT test method was determined using 200 commercially available chemicals with different physical states (105 liquids and 95 solids) representing different organic functional groups. The overall set contained several colour interfering chemicals (1 liquid and 7 solids), MTT reducers (7 liquids and 12

solids) and MTT reducing colourants (1 liquid and 10 solids). In total, 120 chemicals (60 liquids and 60 solids) covering 16 different functional groups were evaluated in the SkinEthic™ HCE EITL and SkinEthic™ HCE EITS validation study. In terms of physical state and UN GHS Categories, the 120 chemicals, were distributed as follows: 32 Cat 1 (16 liquids and 16 solids), 17 Cat 2A (8 liquids and 9 solids), 13 Cat 2B (8 liquids and 5 solids) and 58 No Cat (28 liquids and 30 solids) chemicals. Furthermore, the lead laboratory tested 80 additional chemicals (45 liquids and 35 solids) in three independent runs enlarging the number of functional groups with one additional group. The chemicals covered 19 Cat 1 (11 liquids and 8 solids), 12 Cat 2A (11 liquids and 1 solid), 4 Cat 2B (1 liquid and 3 solids) and 45 No Cat (22 liquids and 23 solids) chemicals.

See Attachment 2.

1.5.2 Assessment of the predictive capacity of the test method

Please provide all available information on the predictive capacity of the test method. Please describe the accuracy of the proposed test method with respect to its ability to measure or predict the effect of interest. The accuracy values (i.e., overall accuracy, sensitivity, specificity, positive and negative predictive values, false positive and negative proportion) of the proposed test should be compared to that obtained for the appropriate reference test method (if available) and/or other data or information, especially from the species of interest (if available). Please append the data in the form of tables and/or figures, as Attachment 12.

The predictive capacity of the assay was evaluated by comparing the prediction results, on the basis of the individual laboratory results, with the existing proposed classification. Therefore 2x2 contingency tables (C versus NC) were constructed and sensitivity (probability of predicting C given the true state is serious eye damage/eye irritancy (Category 1 and Category 2)), specificity (probability of predicting NC given the true state is No Category), and accuracy were calculated.
Bootstrap resampling (10.000 times with sample size = 1) was used to obtain 95% CI's for accuracy, sensitivity and specificity. The rationale for performing bootstrap resampling with size n=1 is that in reality a chemical will be tested only once. Therefore it was opted to calculate sensitivity, specificity, and accuracy on 10.000 simulated sets of 120 chemicals (60 liquids and 60 solids), based on observed predictions (9 predictions per chemical). Briefly, random sampling with sample size n=1 was performed per chemical (pool of 9 predictions, being 3 runs for each of the 3 laboratories) for the set of 60 liquids or 60 solidchemicals. Next, the accuracy, sensitivity and specificity were calculated for each of the 10.000 resampling sets. The mean of the bootstrap sample and 95% CI applying the percentile method was calculated for the three performance parameters.
All analyses were performed with R version 3.1.1. (R Core Team, 2015).

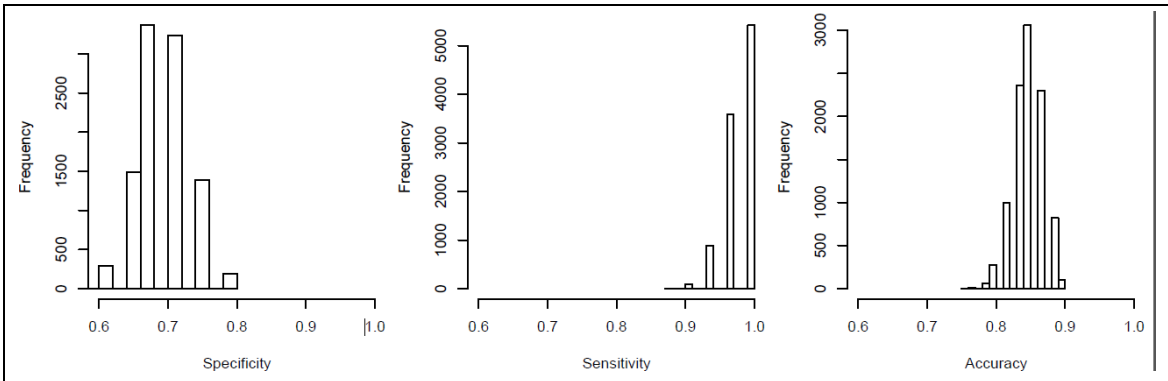
Assessment of predictive capacity for Liquids:

Predictive capacity was calculated for each laboratory and for the cumulative results of the three laboratories using the cut-off of 60% viability to distinguish between chemicals not requiring classification for serious eye damage/eye irritancy (No Cat) from chemicals requiring classification and labelling (Cat 1 and Cat 2) according to UN GHS (Table 1 below). The calculations were based on the individual predictions for each chemical in each laboratory. The sensitivity varied between 96.9% (VITO), 97.9% (CRL) and 100% (lead laboratory L'Oréal). The specificity varied between 65.5% (L'Oréal), 70.2% (VITO), and 72.6% (CRL). An accuracy of 83.9%, 86.1%, and 84.4% was obtained by L'Oréal, CRL, and VITO, respectively. In order to estimate the uncertainty of the sensitivity, specificity and accuracy estimates, the bootstrap resampling method was used. The bootstrap sample consisted of 10000 resamplings of size 1 per chemical for the set of 60 chemicals. The distribution of the bootstrap samples is presented in Figure 1 below. This resulted in an overall sensitivity of 98.2% (95% CI: 93.8% to 100%), a specificity of 69.4% (95% CI: 60.7% to 75.0%), and an accuracy of 84.8% (95% CI: 80.0% to 88.3%).

Table 1: Predictive capacity for the set of 60 liquid chemicals based on individual laboratory predictions: overall and for each laboratory

In vivo UN GHS	Cumulative		L'Oréal		Charles River laboratories		VITO	
	I	NI	I	NI	I	NI	I	NI
Classified (n)	**283**	5	**96**	0	**94**	2	**93**	3
No Category (n)	77	**175**	29	**55**	23	**61**	25	**59**
Total (n)	540		180		180		180	
Sensitivity (%)	98.3		100		97.9		96.9	
Specificity (%)	69.4		65.5		72.6		70.2	
Accuracy (%)	84.8		83.9		86.1		84.4	

Figure 1: Distribution of the bootstrap sample representing 10000 resamplings of size 1 per chemical for the set of 60 liquid chemicals (multicentre study)



Furthermore, the predictive capacity to distinguish chemicals not requiring classification from classified chemicals was determined for the extended dataset (60 chemicals of the multicenter study and 45 additional chemicals tested by the lead laboratory in three independent experiments). This resulted in an accuracy of 83.5% with a 100% sensitivity and 65.3% specificity for L'Oréal only (Table 2). Overall, the sensitivity, specificity, and accuracy based on the individual predictions of the three laboratories were 98.6%, 68.6%, and 84.4% respectively. The bootstrap estimates for this extended dataset of 105 chemicals, correspond with an overall sensitivity of 99.0% (95% CI: 96.4% to 100%), a specificity of 68.5% (95% CI: 64.0% to 74.0%), and an accuracy of 84.4% (95% CI: 81.9% to 87.6%). The distribution of the bootstrap samples is presented in Figure 2 below.

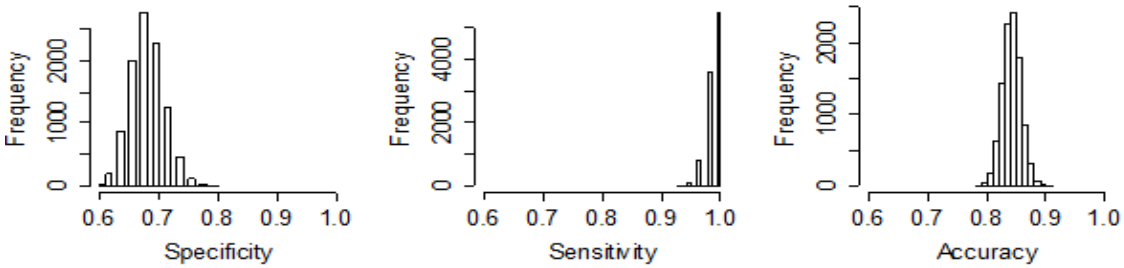
Table 2: Predictive capacity for the set of 105 chemicals based on individual laboratory predictions: overall and for each laboratory

In vivo UN GHS	Cumulative		L'Oréal [a]		Charles River laboratories [b]		VITO [b]	
	I	NI	I	NI	I	NI	I	NI
Classified (n)	**352**	5	**165**	0	**94**	2	**93**	3
No Category (n)	100	**218**	52	**98**	23	**61**	25	**59**
Total (n)	675		315		180		180	
Sensitivity (%)	98.6		100		97.9		96.9	
Specificity (%)	68.6		65.3		72.6		70.2	
Accuracy (%)	84.4		83.5		86.1		84.4	

[a] Predictions based on all chemicals (60 from the multicentre study and 45 additional chemicals)

[b] Predictions based on the 60 chemicals from the multicentre study

Figure 2: Distribution of the bootstrap sample representing 10000 resamplings of size 1 per chemical for the extended data set of 105 liquid chemicals.



Data are appended as Attachments 8a and 9.

Assessment of predictive capacity for Solids:

The predictive capacity was calculated for each laboratory and for the cumulative results of the three laboratories using the cut-off of 50% viability to distinguish NC (No Cat) from chemicals requiring classification and labelling (Cat 1 and Cat 2) according to UN GHS (Table 3). The calculations were based on the individual predictions derived from the qualified tests for each chemical in each laboratory. The three laboratories obtained a sensitivity of 92.2%. The specificity varied between 75.3% (VITO), 75.6% (L'Oréal), and 78.9% (CRL). An accuracy of 83.9%, 85.6%, and 83.3% was obtained by L'Oréal, CRL, and VITO, respectively.

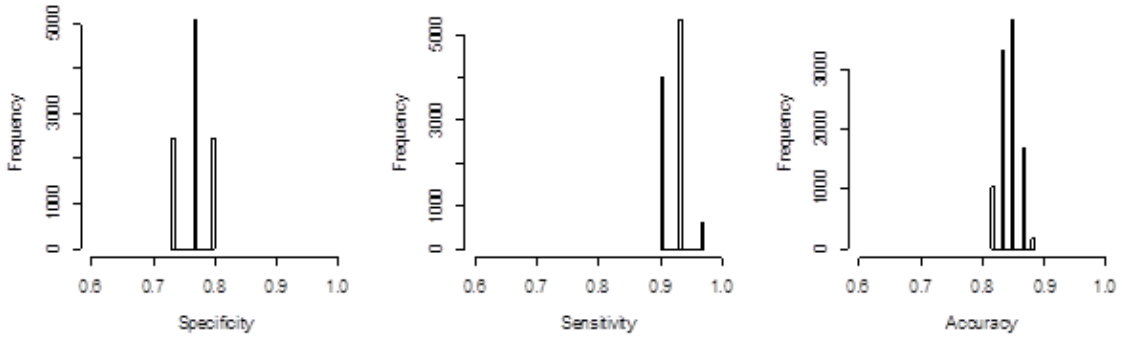
Table 3: Predictive capacity for the set of 60 solid chemicals based on individual laboratory predictions: overall and for each laboratory

In vivo UN GHS	Cumulative		L'Oréal		CRL		VITO	
	C	NC	C	NC	C	NC	C	NC
Classified (n)	**249**	21	**83**	7	**83**	7	**83**	7
No Category (n)	63	**206**	22	**68**	19	**71**	22	**67**
Total (n)	539		180		180		179 [a]	
Sensitivity (%)	92.2		92.2		92.2		92.2	
Specificity (%)	76.6		75.6		78.9		75.3	
Accuracy (%)	84.4		83.9		85.6		83.3	

a For chemical No. 2 only two valid runs were obtained over the five runs

In order to estimate the uncertainty of the sensitivity, specificity and accuracy estimates, the bootstrap resampling method was used. The distribution of the bootstrap samples is presented in Fig. 3. This resulted in an overall sensitivity of 91.9% (95% CI: 90.0% to 93.3%), a specificity of 76.6% (95% CI: 73.3% to 80.0%), and an accuracy of 84.3% (95% CI: 81.7% to 86.7%).

Fig. 3. Distribution of the bootstrap sample representing 10000 resamplings of size 1 per chemical for the set of 60 solid chemical (multicentre study)



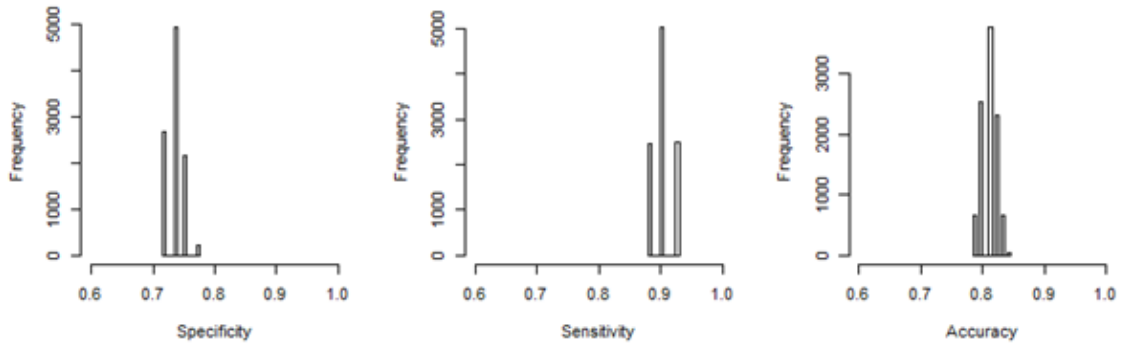
Furthermore, the predictive capacity was also determined for the extended dataset (60 chemicals of the multicenter study and 35 additional solid chemicals). This resulted in an accuracy of 80.7% with a 89.7% sensitivity and a 73.6% specificity for L'Oréal only (Table 4). Overall, the sensitivity, specificity, and accuracy based on the individual predictions of the three laboratories were 91.2%, 75.4%, and 82.9% respectively. The bootstrap estimates for this extended dataset of 95 chemicals, correspond with an overall sensitivity of 90.5% (95% CI: 88.1% to 92.9%), a specificity of 73.6% (95% CI: 71.7% to 75.5%), and an accuracy of 81.0% (95% CI: 78.9% to 83.2%). The distribution of the bootstrap samples is presented in Fig. 4.

Table 4: Predictive capacity for the set of 95 solid chemicals based on individual laboratory predictions: overall and for each laboratory

In vivo UN GHS	Cumulative		L'Oréal [a]		Charles River laboratories [b]		VITO [b]	
	C	NC	C	NC	C	NC	C	NC
Classified (n)	**279**	27	**113**	13	**83**	7	**83**	7
No Category (n)	83	**255**	42	**117**	19	**71**	22	**67**
Total (n)	644		285		180		179 [c]	
Sensitivity (%)	91.2		89.7		92.2		92.2	
Specificity (%)	75.4		73.6		78.9		75.3	
Accuracy (%)	82.9		80.7		85.6		83.3	

a Predictions based on all chemicals (60 from the multicentre study and 35 additional chemicals)
b Predictions based on the 60 chemicals from the multicentre study
c For chemical No. 2 only two valid runs were obtained over the five runs

Fig. 4. Distribution of the bootstrap sample representing 10000 resamplings of size 1 per chemical for the extended data set of 95 solid chemicals.



Data are appended as Attachments 8b and 9.

In conclusion, the accuracy based on the individual predictions obtained in the three laboratories for the set of 120 chemicals (60 liquids and 60 solids) was 84.6% with a specificity of 73.1% and sensitivity of 95.3%. For the liquids (EITL) and solids (EITS) the accuracy was 84.8% and 84.4%, with a specificity of 69.4% and 76.6%, and sensitivity of 98.3% and 92.2%, respectively.

Overall assessment of predictive capacity:

The predictive capacity of SkinEthic™ HCE EIT test method was determined using 200 commercially available chemicals (105 liquids and 95 solids). In terms of physical state and UN GHS Categories, the 200 chemicals, were distributed as follows: 51 Cat 1 (27 liquids and 24 solids), 29 Cat 2A (19 liquids and 10 solids), 17 Cat 2B (9 liquids and 8 solids) and 103 No Cat (50 liquids and 53 solids) chemicals. An accuracy of 83.7% was obtained with a specificity of 72.1% (based on 103 chemicals) and sensitivity of 95.2% (based on 97 chemicals).
The EpiOcular™ EIT has an overall accuracy of 80% (based on 112 chemicals), sensitivity of 96% (based on 57 chemicals), specificity of 63% (based on 55 chemicals) when compared to reference *in vivo* rabbit eye test data (EC EURL ECVAM, 2014; OECD, 2015). However, it is important to note that a strict comparison of the predictive capacity values should not be made since the number of chemicals tested and sets are different.

References

EC EURL ECVAM (2014). Validation Study Report on the EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Corneal Epithelium-based test methods for identifying chemicals not requiring classification and labelling for serious eye damage/eye irritation testing. Available at: in publication.

OECD (2015a). Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. OECD Guideline for the Testing of Chemicals No. 492, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines].

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

1.5.3 Identification and discussion of false predictions

Please identify and discuss any false predictions (e.g. false positives and false negatives).

Identification and discussion of false predictions for Liquids:

The misclassified chemicals were investigated in more detail by taking into account the functional group. Besides this, the Draize eye test irritation data of the misclassified chemicals were also evaluated for the following reason. In total, out of 55 classified chemicals that were tested, four false negative results were obtained. 2-methyl-1-pentanol (CAS RN 105-30-6), ethyl-2-methylacetoacetate (CAS RN 609-14-3) and tetraethylene glycol diacrylate (CAS RN 17831-71-9) were predicted NC in 1 out of 9 runs. [3-(2-Aminoethylamino)propyl]trimethoxysilane (CAS RN 1760-24-3) was predicted NC in 2 out of 9 runs (Attachment 3b). Furthermore, the four false negatives represented four different functional groups (alcohol; ester, ketone; amine-silane; and polyether-acrylate). Two chemicals (2-methyl-1-pentanol CAS RN 105-30-6 and ethyl-2-methylacetoacetate CAS RN 609-14-3) correspond with an *in vivo* UN GHS/EU CLP Cat 2B classification. 2-methyl-1-pentanol resulted in an abnormal high viability (83.6%) in one experiment performed by Charles River in comparison with all other experiments (viability $\leq$ 2%). A single false negative result was obtained for ethyl-2-methylacetoacetate with a viability of 67.4% (VITO). Two other false negatives ([3-(2-Aminoethylamino)propyl]-trimethoxysilane (CAS RN 1760-24-3) and tetraethylene glycol diacrylate (CAS RN 17831-71-9)) correspond with an *in vivo* Cat 1 classification. [3-(2-Aminoethylamino)propyl]-trimethoxysilane resulted two times in a NC prediction (viability: 65.1% and 65.7% in runs 2 and 3, respectively) by VITO. For this chemical crystal formation in the sample was reported upon storage. The first run (viability: 25.1%) was performed in the beginning of the experimental phase whereas the second and third runs were performed at the end of the experimental phase (more than 60 days later). The effect of storage condition on the stability of this chemical ([3-(2-Aminoethylamino)propyl]trimethoxysilane) was evaluated after the validation study. Indeed, the viability increased when the container was not closed properly. After 14 and 30 days of storage with half open or open lid, mean viability increased above 50% (51.5% to 66.3%). In the two other laboratories, the independent runs for the [3-(2-Aminoethylamino)propyl]-trimethoxysilane were performed within a period of less than 30 days. Tetraethylene glycol diacrylate resulted in a NC prediction for one experiment performed by CRL, the viability of 62% was just above the classification cut-off of 60%.

In terms of the *in vivo* driver of classification, 2-methyl-1-pentanol and ethyl-2-methylacetoacetate were classified Cat 2B based on corneal opacity in the Draize eye test. [3-(2-Aminoethylamino)propyl]trimethoxysilane, classified Cat 1 in the Draize eye test based on persistent conjunctival and corneal effects on day 21 in the majority of the animals. Tetraethylene glycol diacrylate was classified as Cat 1 based on iritis and resulted in severe but delayed corneal opacity in the Draize eye test. It is important to note that the false negative results were only obtained for 1 or 2 out of the 9 independent runs. Therefore, we can conclude that the false negative results are not related to the driver of *in vivo* classification.

Of the 50 *in vivo* No Cat chemicals, 28 were correctly predicted. Twenty two chemicals not requiring classification resulted in a false positive prediction in at least one experiment. Twelve *in vivo* UN GHS No Cat chemicals (CAS RN 3970-62-5, 2370-63-0, 13826-35-2, 623-51-8, 106-91-2, 111-25-1, 109-64-8, 931-87-3, 141-78-6, 3938-95-2, 123-86-4, and 9002-93-1) were consistently predicted C (mean viability < 60%) in all runs. Among them were three esters, ethyl acetate (CAS RN 141-78-6), ethyl trimethyl acetate (CAS RN 3938-95-2), n-Butyl acetate (CAS RN 123-86-4) which resulted in a mean viability < 10% (Attachment 3c). These esters also resulted in a false positive prediction in the BCOP (Balls et al., 1995b) and EpiOcular™ EIT (Kaluzhny et al., 2011). The other 9 false positives represent 8 different functional groups, the number of chemicals within a functional group is too small to draw conclusions with regard to over-predictions.

Nine additional *in vivo* UN GHS No Cat chemicals were sometimes predicted C but the viability was in the majority of the cases between 50% and 60%. In particular, 1,9-decadiene (CAS RN 1647-16-1) was seven times predicted C. Dipropyl disulphide (CAS RN 629-19-6) was six times predicted C and ethoxydiglycol (CAS RN 111-90-0) was four times predicted C, p-Methyl thiobenzaldehyde (CAS RN 3446-89-7) was three times predicted C, and 1-Bromo-4-chlorobutane (CAS RN 6940-78-9) was predicted C twice in the multicentre study. The false positive results for dimethyl sulfoxide (CAS RN 67-68-5), triethanolamine orthovanadate 30 (CAS RN 13476-99-8), triethylene glycol monomethyl ether (CAS RN 112-35-6), and triethylene glycol (CAS RN 112-27-6) all resulted in a mean viability of ≥ 56.9%. The single false positive result obtained for iso-propyl myristate (CAS RN 110-27-0) (mean viability of 19%) was an exception, for all other runs the mean viability was > 90%. With respect to the *in vivo* No Cat chemicals, an interesting relation was found between the SkinEthic™ HCE EITL data and the Draize eye test data. Of the 50 *in vivo* No Cat chemicals that were tested, 40 chemicals showed corneal opacity (CO) scores equal to 0 in all animals and all observed time points in the Draize eye test (CO = 0) (Barroso et al., 2015). Twenty six out of those 40 chemicals were consistently predicted NC with the SkinEthic™ HCE EITL method by all laboratories. For 9 out of 40 chemicals, the false positive result corresponded often with a mean viability between 50% and 60%. Another five chemicals resulted in a false positive result in all runs (mean viability < 50%). Ten of the 50 *in vivo* No Cat chemicals showed

CO scores equal greater than 0 in at least one animal for at least one observed time point in the Draize eye test (CO > 0) (Barroso et al., 2015). Seven out of those 10 chemicals were consistently predicted C with the SkinEthic™ HCE EITL assay by all laboratories (mean viability < 35%).
Since several false positives resulted in a mean viability between 50% and 60%, the effect of decreasing the cut-off value to 50% for distinguishing chemicals that have C potential (Cat 1/Cat 2) from No Cat chemicals was evaluated. A cut-off of 50% would result in an increase of the specificity from 68.6% (60% cut-off) to 76.1% with a slight decrease in sensitivity from 98.6% to 97.2%. However, one Cat 1 chemical (tetraethylene glycol diacrylate) would result in an overall false negative prediction by VITO. As a consequence, the performance criteria were not met with the 50% cut-off value since none of the Cat 1 chemicals should be under-predicted in the majority of the runs (OECD, 2015a). However, with the 60% cut-off value, the SkinEthic™ HCE EITL method resulted in a similar sensitivity (98.6%), specificity (68.6%) and accuracy (84.4%) as obtained with the RhCE EpiOcular™ Eye Irritation Test (EIT) test method validated in the EURL ECVAM/Cosmetics Europe study and being accepted for identifying No Cat chemicals (OECD, 2015b).

Identification and discussion of false predictions for Solids:

The misclassified chemicals were investigated in more detail by taking into account the functional group and the UN GHS category. In total, out of 42 classified chemicals that were tested, five false negative results were obtained. 1,5-Naphthalenediol CAS RN 83-56-7 *(in vivo* Cat 2A) was predicted NC in 4 out of 9 runs and benzene, 1,3-dinitro CAS RN 99-65-0 (*in vivo* Cat 2B) was predicted NC in 8 out of 9 runs. 2-Azetidinone, 4-(acetyloxy)-3-[(1R)-1-[[(1,1-dimethylethyl)dimethylsilyl]oxy]ethyl]-, (3R,4R) (CAS RN 76855-69-1), benzene, 1,4-dibutoxy (CAS RN 104-36-9) and cyclopropanecarboxylic acid, 3-[(1Z)-2-chloro-3,3,3-trifluoro-1-propen-1-yl]-2,2-dimethyl-, (2-methyl[1,1'-biphenyl]-3-yl)methyl ester, (1R,3R)-rel- (CAS 82657-04-3) (*in vivo* Cat 2A, 2B and 2B, respectively) were always predicted NC (Attachments 8b and 9). The five false negatives correspond with four different functional groups (nitro-compound, two esters, phenol and ether), therefore it is unlikely that the under-predictions are related with the functional group.
Of the 53 *in vivo* No Cat chemicals, 38 were correctly identified as NC. Twelve *in vivo* UN GHS No Cat chemicals (CAS RN 84540-47-6, 598-65-2, 94-13-3, 66170-10-3, 7631-90-5, 120-14-9, 100-97-0, 92-43-3, 350-30-1, 19285-83-7, 25102-12-9, and 156028-26-1) were consistently predicted C (mean viability < 50%) in all runs. Three additional *in vivo* No Cat chemicals (CAS RN 1603-02-7, 33089-61-1, and 59997-51-2) resulted in a false positive prediction in at least two runs. No relation was observed

between the false positives and the functional group of the chemical (12 different functional groups).

Depth analyses with respect to the drivers of *in vivo* classification:

Recent papers have shown the importance of understanding these effects for the evaluation of alternative methods (Barrosso et al., 2013; Adriaens et al., 2014). A full description of all the ocular effects that drive classification is available for a large set of reference chemicals, the so-called Draize eye test Reference Database (DRD) published by Barroso and co-workers (2015). In order to evaluate the predictive capacity of the SkinEthic™ HCE EIT test method and its limitations, the misclassified chemicals were correlated with the *in vivo* drivers of classification as presented in the DRD (Barroso et al., 2015). Extensive analyses of *in vivo* UN GHS Cat 1 studies, presented in the DRD, showed that the most important drivers for Cat 1 classification are CO mean ≥ 3 (mean scores calculated from grading at day 1, 2 and 3 after instillation of the chemical in the eye) and CO persistence on day 21 in the absence of severity (CO mean ≥ 3) (Adriaens et al., 2014; Barroso et al., 2015). The most important drivers for a Cat 2 classification are CO mean ≥ 1 and conjunctival redness mean ≥ 2. Barroso and co-workers (2015) also suggested a critical revision of the current UN GHS decision criteria, one of the key conclusions of this analysis was that all classifiable Cat 1 effects should be present in more than 60% of the animals. The most important drivers of Cat 1 and Cat 2 classification named above were well represented in the solids and liquids chemicals set evaluated with the SkinEthic™ HCE EITS and EITL protocol (Fig. 1 and Fig. 2). The results of 9 chemicals [(4 solids: Manganese, [N-[2-[(dithiocarboxy)amino]ethyl]carbamodithioato(2-)-κS,κS'] CAS RN 12427-38-2; Phenol, 4-chloro-2-(phenylmethyl) CAS RN 120-32-1; 4,7-Methanoisobenzofuran-1,3-dione, 4,5,6,7,8,8-hexachloro-3a,4,7,7a-tetrahydro CAS RN 115-27-5; and quinacrine dihydrochloride CAS RN 69-05-6) and (5 liquids: 1-decanol CAS RN 112-30-1; ethanol CAS RN 64-17-5; poly(ethylene glycol) butyl ether CAS RN 9004-77-7, anisole CAS RN 100-66-3; and n-butanol CAS RN 71-36-3)] were not included in the pie charts since the driver could not be identified or because multiple studies were available for the same chemical and the driver differed between the repeat studies. From the regulatory perspective, it is important to note that out of the 51 *in vivo* UN GHS Cat 1 chemicals (24 solids and 27 liquids) that were tested with the optimized SkinEthic™ HCE EITL or EITS protocols, 49 (96.1%) chemicals were always correctly identified as C. Of note, Fig. 1 and Fig. 2 includes the results of 21 Cat 1 solids and 25 Cat 1 liquids. The false negative result obtained by VITO for the liquid 3-(2-aminoethylamino)propyl]-trimethoxysilane, was probably related to instability of the chemical. The second chemical with a false negative prediction (1 out of 9 runs) was the liquid tetraethylene glycol diacrylate, this chemical was classified Cat 1 based on iritis in the Draize eye test,

an endpoint know to be of minor importance in driving Cat 1 classification (Barroso et al., 2015). With respect to the most important drivers of Cat 1 classification, an excellent predictive capacity was obtained for the solids with 100% correct predictions (Fig. 1) and a very high predictive capacity was obtained for the liquids with 100% correct predictions for the driver CO mean ≥ 3 and 97.4% correct predictions for the liquids that were classified Cat 1 based on persistence (Fig. 2). Overall 84.8% (39/46) of the *in vivo* UN GHS Cat 2 chemicals (18 solids and 28 liquids) were always predicted C. Of note, Fig. 1 and Fig. 2 show the results of 17 solids and 25 liquids, as mentioned before, the results of 1 solid and 3 liquids were omitted. For two *in vivo* UN GHS Cat 2B liquid chemicals (2-methyl-1-pentanol CAS RN 105-30-6 and ethyl-2-methylacetoacetate CAS RN 609-14-3), only 1 out of 9 runs resulted in an under-prediction. The viability for 2-methyl-1-pentanol was 83.6% in one run whereas in 8 other runs, the viability was below 2%. For ethyl-2-methylacetoacetate a borderline classification was obtained in one run (62%), whereas the viability in the other 8 runs was < 60%. The solid, 1,5-naphthalenediol (CAS RN 83-56-7) was under-predicted in 4 out of 9 runs, this chemical was predicted Cat 2A based on conjunctival effects only. Four additional solid chemicals (Benzene, 1,3-dinitro (CAS RN 99-65-0); 2-Azetidinone, 4-(acetyloxy)-3-[(1R)-1-[[(1,1-dimethylethyl)dimethylsilyl]oxy]ethyl]-(3R,4R) (CAS RN 76855-69-1); Benzene, 1,4-dibutoxy (CAS RN 104-36-9); and Cyclopropanecarboxylic acid, 3-[(1Z)-2-chloro-3,3,3-trifluoro-1-propen-1-yl]-2,2-dimethyl-, (2-methyl[1,1'-biphenyl]-3-yl)methyl ester, (1R,3R)-rel) (CAS 82657-04-3) were under-predicted in the majority of the runs. Two chemicals (Benzene, 1,3-dinitro and Benzene, 1,4-dibutoxy) were classified Cat 2 based on conjunctival effects only. The two others were both classified Cat 2 based on corneal opacity. The performance of the SkinEthic™ HCE EIT method in terms of most important drivers of Cat 2 classification was high (Fig. 1 and Fig. 2). Cat 2 chemicals (6 solids and 20 liquids) that were classified based on corneal opacity were correctly predicted in 71.4% (solids) and 96.5% (liquids) of the runs. Furthermore, although the SkinEthic™ HCE EIT method models the cornea, a substantial proportion of the Cat 2 chemicals (11 solids and 5 liquids) that were classified based on conjunctival effects only were also identified correctly (82.8% and 100% of the runs for solids and liquids, respectively). This provides evidence that the SkinEthic™ HCE EIT method can also identify chemicals which result in *in vivo* conjunctival effects only. Concerning the false positives, it is interesting to note that the liquids of the subgroups CO > 0 and CO > 0 ** were often over-predicted by the SkinEthic™ HCE EITL method (44% and 100% of the runs, respectively). This relationship was also observed for the solids, chemicals of the subgroups CO > 0 and CO > 0 ** were over-predicted in 40% and 75% of the runs, respectively (Fig. 1). In the Draize eye test, those chemicals induced CO scores greater than 0 in at least one animal for at least one observed time point. Moreover, for two solid chemicals and five liquid chemicals (Fig. 2), CO mean over the first three days was equal to or greater than 1 in one animal (subgroup CO > 0 **). This means that the SkinEthic™ HCE EIT method is very sensitive in detecting such *in vivo* effects. On the

other hand, solid and liquid chemicals from the subgroup CO = 0 (CO scores equal to 0 in all animals and all observed time points in the Draize eye test) were identified correctly in 80.1% and 79.1% of the runs, respectively.

Fig. 1. Distribution of the solid chemicals (chem.) according to the drivers of classification (UN GHS Cat 1 and Cat 2) and according to the subgroups (UN GHS No Cat) as defined by Barroso and co-workers (2015). The proportion correct predictions corresponds with the number of runs that were correctly predicted over the total number of runs that were performed.

a The data of 17/18 Cat 2 chemicals are included in the chart, chemical No. 43 was excluded since the driver could not be identified (SCNM).

b The data of 21/24 Cat 1 chemicals are included in the chart, chemicals No. 49 and No. 93 were excluded since multiple studies were available that resulted in a different driver and No. 90 was excluded since the driver could not be identified (SCNM).

"CO = 0" Corneal Opacity (CO) scores equal to 0 in all animals and all observed time points in the Draize eye test; "CO > 0" CO scores greater than 0 in at least one animal for at least one observed time point, ** correspond with No Cat studies for which at least one animal had a mean of the scores of days 1-3 above the classification cut-off for at least one endpoint but not enough animals to generate a classification; "CO mean ≥ 1" mean CO scores of days 1-3 ≥ 1 in ≥ 60% of the animals; "Conj mean ≥ 2" mean Conjunctival Redness (CR) and/or Conjunctival Chemosis (CC) during the first three observation days ≥ 2 in ≥ 60% of the animals in absence of "CO mean ≥ 1"; "CO mean ≥ 3" mean CO scores of days 1-3 ≥ 3 in ≥ 60% of the animals; "IR mean > 1.5" mean Iritis (IR) scores of days 1-3 > 1.5 in ≥ 60% of the animals in absence of "CO mean ≥ 3"; "Pers D21" persistence of any ocular effect on day 21 in the absence of severity ("CO mean ≥ 3" and "IR mean > 1.5"); "CO = 4" at any observation time during the study in the absence of both severity and persistence.

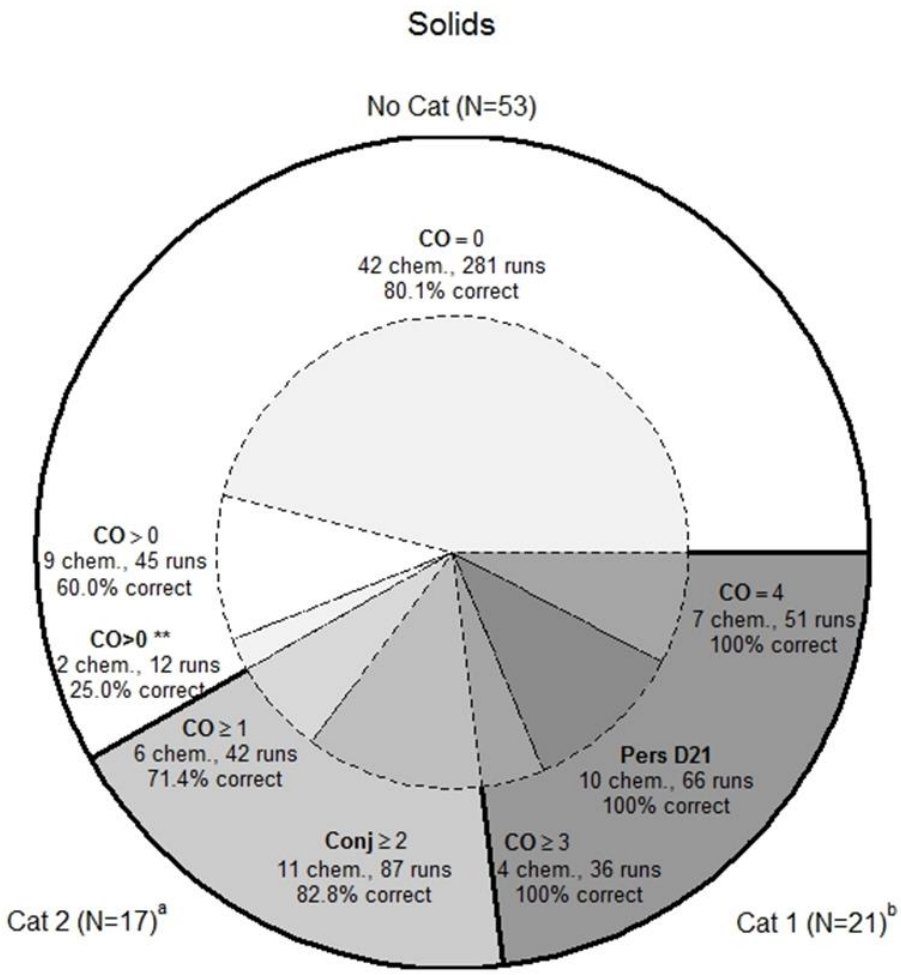
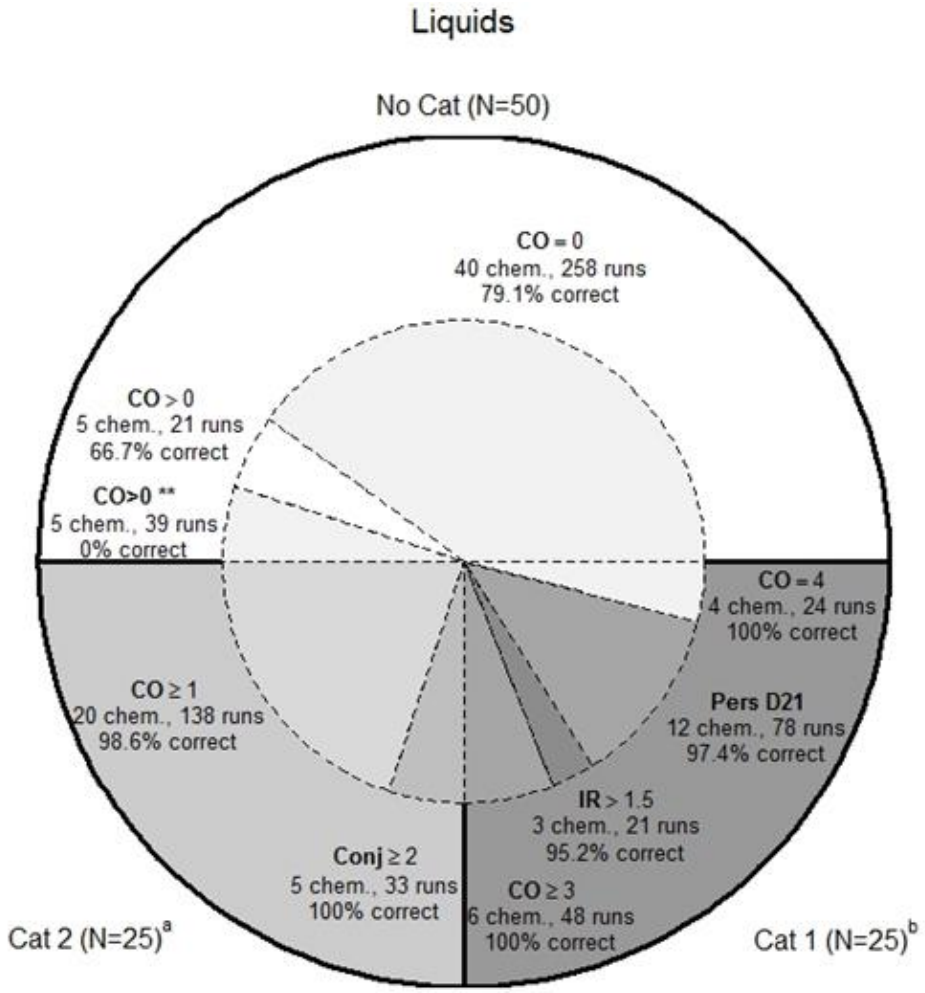


Fig. 2. Distribution of the liquid chemicals according to the drivers of classification.
a The data of 25/28 Cat 2 chemicals are included in the chart, 1-decanol and poly(ethylene glycol) butyl ether were excluded since the driver could not be identified (SCNM), ethanol was excluded since multiple studies were available resulting in different drivers/classifications.
b The data of 25/27 Cat 1 chemicals are included in the chart, anisole was excluded since the driver could not be identified (SCNM) and n-butanol was excluded since multiple studies were available that resulted in a different driver.



References

Adriaens E, Barroso J, Eskes C, Hoffmann S, McNamee P, Alépée N, Bessou-Touya S, de Smedt A, De Wever B, Pfannenbecker U, Tailhardat M, Zuang V (2014). Retrospective analysis of the Draize test for serious eye damage/eye irritation: importance of understanding the *in vivo* endpoints under UN GHS/EU CLP for the development and evaluation of *in vitro* test methods. Arch Toxicol 88:701-723. doi: 10.1007/s00204-013-1156-8.

Balls., M., Botham, P.A., Bruner, L.H., Spielmann, H. (1995b). The EC/HO international validation study on alternatives to the draize eye irritation test. Toxicol. In Vitro 9, 871-929. doi: 10.1016/0887-2333(95)00092-5.

Barroso, J., Alépée, N., De Smedt, A., De Wever, B., Hibatallah, J., McNamee, P., Mewes, K.R., Millet, M., Pfannenbecker, U., Tailhardat, M., Templier, M. (2013). The importance of understanding drivers of irritation in vivo for selection of chemicals used in the development and evaluation of in vitro eye irritation assays: Cosmetics Europe analysis. Poster presented at the In Vitro Testing Industrial Platform (IVTIP) Spring 2013 Meeting, Southampton, UK, May 14–16, 2013. Available from: https://www.cosmeticseurope.eu/downloads/6057.html. Accessed on 05/10/2015.

Barroso, J., Alépée, N., Adriaens, E., Cluzel, M., De Smedt, A., Hibatallah, J., Klaric, M., Mewes, K.R., Millet, M., Pfannenbecker, U., Templier, M., McNamee, P. (2015a). Cosmetics Europe compilation of historical serious eye damage/eye irritation *in vivo* data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative *in vitro* methods/strategies: the Draize eye test Reference Database (DRD). Submitted.

Kaluzhny, Y., Kandarova, H., Handa, P., Kunilus, J., d'Argembeau-Thornton, L., Klausner, M. (2011). DEvelopment of the EpiOcularTM Eye Irritation Test (EpiOcular-EIT) for hazard identification and labelling of eye irritating chemicals in response to the requirements of the cosmetics directive and REACH legislation. ATLA 39, 339-364.

OECD (2015a). Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. OECD Guideline for the Testing of Chemicals No. 492, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines]

OECD (2015b). Performance standards, Series on Testing & Assessment No. 216. Performance Standards For The Assessment Of Proposed Similar Or Modified *In Vitro* Reconstructed Human Cornea-Like Epithelium (Rhce) Test Methods For Eye Hazard. Available at: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)23&docLanguage=En]

1.6 MODULE 6: APPLICABILITY DOMAIN (AD)

1.6.1 Applicability of the test method identified through testing

Please identify the applicability of the test method on the basis of experimental evidence. In the case of chemicals, indicate the chemicals and/or chemical categories (e.g. based on functional groups and/or physicochemical properties) for which the test method makes reliable and relevant predictions.

The applicability of SkinEthic™ HCE EIT test method was determined using 200 commercially available chemicals with different physical states (105 liquids and 95 solids). The chemicals represent different functional groups (cationic soap / surfactant, neutral soap / surfactant, anionic soap / surfactant, neutral organic, organic acid, organic base, organic cationic, organic salt, organic acid/organic base, inorganic acid, neutral inorganic, inorganic base, and inorganic salt). Different functional class were represented such as carboxylic acid & derivatives (ester, amide) , alcohol, allyl alcohol, cyclic alcohol, alkanes, amine, ammonium salt, aromatic derivatives, phosphorus derivatives, electrophile (acrylate, aldehyde, ketone), ether & polyether, halogenates, heterocyclic, nitrile, silicium derivatives, thiol, di-sulfur, sulfur oxides, nitro derivatives, urea derivatives and metal derivatives.The majority of these chemicals represented mono-constituent substances, but several multi-constituent substances (including 10 polymers e.g. Acrylamide Copolymer, Polyquaternium-10, Nonoxynol-2) were also included in the study. The majority of the 200 chemicals were tested neat (175) or in dilution (0.1 to 30%). The overall set contained several colour interfering chemicals (1 liquid and 7 solids), MTT reducers (7 liquids and 12 solids) and MTT reducing colorants (1 liquid and 10 solids). In terms of physical state and UN GHS Categories, the 200 chemicals, were distributed as follows: 51 Cat 1 (27 liquids and 24 solids), 29 Cat 2A (19 liquids and 10 solids), 17 Cat 2B (9 liquids and 8 solids) and 103 No Cat (50 liquids and 53 solids) chemicals.

Results generated in the validation study have demonstrated that SkinEthic™ HCE EIT is applicable those categories.

1.6.2 Limitations of the test method identified through testing

Please identify on the basis of experimental evidence the limitations of the test method. In the case of chemicals, indicate the chemicals and/or chemical categories (e.g. based on functional groups and/or physicochemical properties) for which the test method has been shown not to be applicable to.

Gases and aerosols have not been assessed in a validation study. While it is conceivable that these can be tested using this technology, the current SkinEthic™ HCE EIT test method does not allow testing of gases and aerosols.
SkinEthic™ HCE EIT test method is not intended to differentiate between UN GHS Category 1 (serious eye damage) and UN GHS Category 2 (eye irritation). This differentiation will need to be addressed by another tier of a test strategy (Scott et al, 2010). A chemical that is identified as requiring classification for eye irritation/serious eye damage with SkinEthic™ HCE EIT will thus require additional testing (*in vitro* and/or *in vivo*) to establish a definitive classification, using e.g., OECD TG 437, 438, 460 or 492.

References

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

1.6.3 Suggested Applicability Domain (AD) of the test method

Please define to the extent possible the (preliminary) AD of the test method on the basis of the information given in 1.6.1 and 1.6.2. Indicate whether the AD is considered to be wider than the one experimentally demonstrated and provide justification for this.

The SkinEthic™ HCE EIT test method allows the identification of chemicals not requiring classification and labelling for eye irritation or serious eye damage in accordance with UN GHS independently of the types of ocular effects observed *in vivo* (i.e., corneal, iridal and conjunctival injuries).
The SkinEthic™ HCE EIT is applicable to substances and mixtures, and to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water.

1.7 MODULE 7: PERFORMANCE STANDARDS (PS)

Performance Standards, based on a validated test method, can be used to evaluate the reliability and accuracy of proposed similar or updated test methods. The three elements of PS are: a) essential test method components, b) a minimum list of reference test items and c) the defined accuracy and reliability values.
Please consider this section as optional, since these standards are normally defined at the end of validation studies.

Please indicate the essential structural, functional and procedural components of a validated test method that should be retained in a proposed test method to be judged as similar.

[…]

1.7.2 Suggestions of possible reference test items

Please suggest test items which should be used to assess the accuracy and reliability of a similar test method. These test items should be a representative subset of those used to demonstrate the reliability and accuracy of the validated test method.

[…]

1.7.3 Suggestions of defined accuracy and reliability values

Please suggest the target accuracy and reliability values that should be achieved by the similar test method when evaluated using the reference test items.

[…]

2. ESSENTIAL INFORMATION FOR A SPECIFIC VALIDATION PROCESS

2.1 External Validation Studies (not coordinated by ECVAM)

All validation studies, also those which are not coordinated by ECVAM, should follow some good validation principles. The following sections of the TST (from 2.1.1 to 2.1.5) apply to all validation studies not coordinated by ECVAM including external retrospective validation studies and external performance standards-based validation studies.

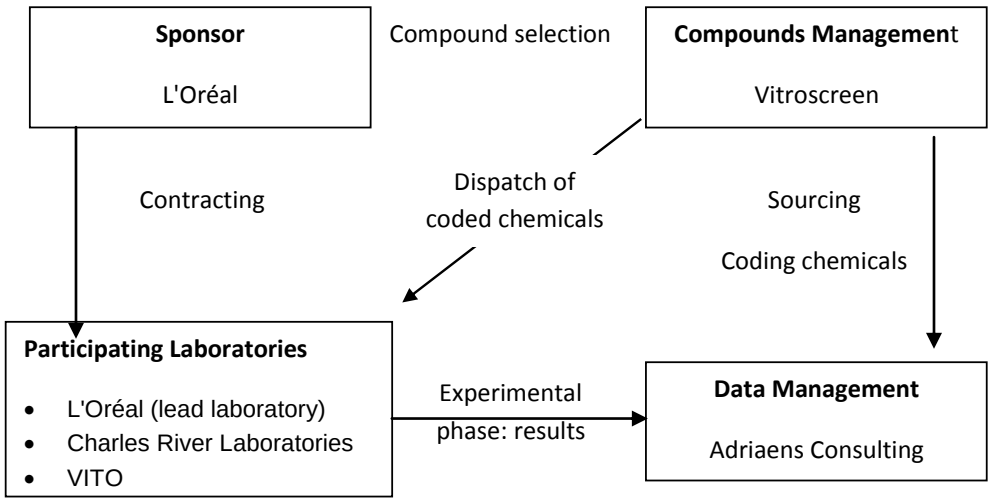
2.1.1 Study organisation and management (Project Plan)

For external validation studies (not coordinated by ECVAM) a detailed project plan, describing the study organisation and management should be produced prior to the initiation of the study and provided as Attachment 13 (see Explanatory Note to the TST on pages 2-3). Please briefly summarise here below how the study was designed (e.g. number of participating laboratories, number of replicates within a single experiment, number of independent repetitions, etc.) and managed.

The study coordination was conducted by L'Oréal R&I including the organisation of all necessary meetings or teleconferences. Independent coding and distribution of chemicals was contracted out to VitroScreen. Independent statistical analyses of the data were contracted out to Adriaens consulting BVBA. The SkinEthic™ HCE EIT method was assessed in three laboratories. L'Oréal (L'Oréal Research & Innovation, Aulnay sous Bois, France) acted as lead laboratory, Charles River Laboratories (CRL, Edinburgh, United Kingdom) and VITO (Flemish Institute for Technological Research, Mol, Belgium) acted as naive laboratories (Figure 1).
Training of the participating laboratories has been provided by the test method developer (L'Oréal). Upon completion of the training phase, participating laboratories tested 9 chemicals to demonstrate transferability of the assay. In the testing phase of multicentre study, each of the chemicals (60 liquids and 60 solids) were tested in at least three independent tests (using different tissue batches and performed in separate runs, 2 tissues replicates/run/chemical) by each of the three independent laboratories. The Study Directors forwarded the data acquired by their laboratories to the Study Data Coordinator for analysis. Once completed, these statistical analyses and their conclusions were provided to the sponsor.
Vitroscreen was the Study Products Coordinator that had responsibility for the organisation of all the different aspects of the test items including coding, shipment and safety data sheets. Random coding of the chemicals (i.e. different codes used for each laboratory in each study) was established by Vitroscreen. During the conduct of the study, the above mentioned Study Products Coordinator was the single point of contact for the Test Facility regarding the test items including coding, shipment and Safety Data Sheets.

Els Adriaens, from Adriaens consulting BVBA was the Study Data Coordinator that has organized the communication flow between all parties and has coordinated the study conduction. Els Adriaens had the responsibility of appropriate data storage, reporting and archiving, the statistical supervision of the laboratories and follow-up of issue encountered during the experimental phase, the statistical analysis of the study data and the chemicals decoding.
The participating laboratories were allowed to freely communicate and meet during the training and transfer phases of the multicentre study. However, during the testing phase, the participating laboratories and the personnel responsible for providing training on the test method were no longer contact each other regarding this multicentre study. During the conduct of the study, the Study Data Coordinator was the single point of contact for the Test Facility. No exchange was made with the Study Sponsorship representatives. All experiments were performed according to the SOPs that were sent by the Study Data Coordinator.

Figure 1: Management structure of the SkinEthic™ HCE EIT validation study



Project Plans of the SkinEthic™ HCE EIT validation exercises are appended as Attachments 10a (EITL) and 10b (EITS).

2.1.2 Study objective and goals

Please specify the objective and goals of the study (e.g. evaluation of the transferability of the test method, evaluation of WLR and/or BLR, and/or assessment of PC) here below.

The objective of this study was to formally assess the relevance (predictive capacity) and reliability (reproducibility within and between laboratories) of the SkinEthic™ HCE EIT test method, to facilitate its international acceptance in regulatory schemes for hazard assessment of chemicals. For this purpose, an inter-laboratory validation study was performed by testing a statistically significant number of coded chemicals (substances and mixtures), supported by *in vivo* Draize eye irritation data for comparative evaluation of results.
In particular, this test method shall be incorporated into a tiered test strategy (so-called Bottom-Up/Top-Down test strategy, as defined in an ECVAM workshop held in 2005, Scott L. *et al.*, 2010) as e.g. the initial step in a Bottom-Up approach or the second step in a Top-Down Approach. The ultimate purpose of a tiered test strategy will be to replace the traditional *in vivo* Draize eye irritation test [Method B.5 of EC Regulation 440/2008 (EC, 2008) or OECD TG 405 (OECD, 2012)].

References:

EC (2008). Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Off J Eu Union L353:1–1355.

OECD (2012): Acute Eye Irritation/Corrosion. OECD Guideline for Testing of Chemicals No. 405. Organisation for Economic Co-operation and Development, Paris. Available at: [http://www.oecd-ilibrary.org/environment/test-no-405-acute-eye-irritation-corrosion_9789264185333-en].

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

2.1.3 Summary of the study results

Please briefly summarise the study main results, from the information included in the other sections of the TST, in light of the objectives and goals set for the study (e.g. agreement of the results for WLR and BLR when experimental data have been generated using the same test items and adhering to the same protocol, accuracy of the test method to measure or predict the effect of interest).

A prospective multicentre study of the reconstructed human corneal epithelial tissue-based *in vitro* test method (SkinEthic™ HCE) was conducted to evaluate its usefulness to identify chemicals as either not classified for serious eye damage/eye irritation (No Cat.) or as classified (Cat. 1/Cat. 2) within UN GHS. The aim of this study was to demonstrate the transferability and reproducibility of the SkinEthic™ HCE EITL and EITS protocols for liquids and solids, respectively and define its overall predictive capacity. Briefly, 120 (60 liquids, 60 solids) chemicals were three times tested (double blinded) in 3

laboratories. Additional chemicals (45 liquids and 35 solids) were tested three times in the lead laboratory.
The WLR, based on the set of 120 chemicals, was 91.7% (EITL 88.3% and EITS 95.0%) for CRL, 94.2% for VITO (EITL 93.3% and EITS 95.0%) and 95.8% for L'Oreal (EITL 95.0% and EITS 96.7%). The WLR for the extended set of 200 chemicals that were tested by L'Oreal only was 95.0% (EITL 93.3% and EITS 96.8%). Furthermore, the overall concordance between the laboratories (BLR) for the HCE EIT method, based on the set of 120 chemicals, was 95.0% (EITL 93.3% and EITS 96.7%). The accuracy based on the individual predictions obtained in the three laboratories for the set of 120 chemicals was 84.6% with a specificity of 73.1% and sensitivity of 95.3%. For the liquids (EITL) and solids (EITS) the accuracy was 84.8% and 84.4%, with a specificity of 69.4% and 76.6%, and sensitivity of 98.3% and 92.2%, respectively.
At present, 200 chemicals were tested (105 liquids and 95 solids) resulting in a sensitivity of 95.2%, specificity of 72.1% and accuracy of 83.7%, thereby meeting all acceptance criteria for predictive capacity.

2.1.4 List of test items used in the validation study

Please provide as Attachment 14, a table listing the test items used in the validation study and specify for each test item whether it has been used for a) test method development, b) WLR and/or BLR assessment and c) predictive capacity assessment.

The list of test items is appended as Attachment 2.
Descriptions are reported in the respective sections 1.1.6 (optimisation), 1.2.1 (WLR), 1.3.1 (Transferability), 1.4.1 (BLR) and 1.5.1 (Predictive capacity).

2.1.5 Study conclusions

Please provide your conclusions regarding the outcome of the study [e.g. is the test method easily transferable to other laboratories? Is the test method generating reproducible results within a single laboratory and between laboratories? Is the test method sufficiently accurate compared to the reference data (e.g. in vivo data)/ target accuracy and reliability values (for performance standards-based validation studies)?], in the light of the study objectives and in consideration of the extent to which such conclusions are supported by the study results.

See section 2.1.3

2.1.6 Recommendations

Considering the study objectives and outcome, please provide recommendations, where applicable, regarding a) possible improvements of the test method (e.g. in relation to the SOP, prediction model etc.), b) future activities to be undertaken to better characterise the performance of the test method in view of its envisaged use (e.g. better characterisation of the PC by testing additional coded chemicals), c) the possible current use of the test method (e.g. screening method, partial replacement method as part of a testing strategy, full replacement method), d) any other recommendation on future activities.

SkinEthic™ HCE EIT test method is intended to discriminate chemicals not requiring classification for serious eye damage/eye irritancy (No Cat) from chemicals requiring classification and labelling (Cat 1 and Cat 2). The SkinEthic™ HCE EIT test method is however not intended to differentiate between UN GHS Cat 1 (serious eye damage) and UN GHS Cat 2 (eye irritation). In fact, a chemical that is identified as requiring classification for eye irritation/serious eye damage with SkinEthic™ HCE EIT will require additional testing by another tier of a test strategy (Scott et al., 2010). A definitive classification can be established using e.g., OECD TG 437, 438, 460 or 492.

2.2 Essential information for a Retrospective Validation Study

Retrospective validation studies of test methods are performed on the basis of available information/data collected from different sources (such as validation studies, in-house data, unpublished company data, publications, communications, etc).

Not applicable

2.3 Essential information for a Validation Study based on Performance Standards (for similar or updated test methods)

Not applicable

3. ADDITIONAL INFORMATION

3.1 Additional information - HPLC/UPLC-spectrophotometry

Please add any additional information that was not covered in the TST.

The endpoint to identify chemical effects in all of *in vitro* test methods based on Reconstructed human Tissues (RhT) is measurement of tissue viability in treated tissues after topical application onto the tissue surface. Tissue viability is determined by enzymatic reduction of yellow 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) tetrazolium salt to purple reduced MTT (formazan) (Mosmann, 1983). Formazan is then quantified photometrically with the results being expressed as % viability of the test chemical treated tissues relative to negative control treated tissues and converting this to a classification using a prediction model based on a percentage (%) viability cut-off value. A known limitation of the photometric MTT-reduction assay is possible interference of coloured test chemicals with the absorbance measurement of

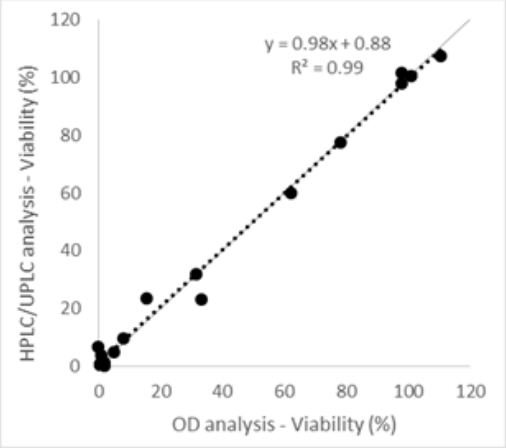
formazan. This limitation has been addressed in a project completed by Cosmetics Europe in which the use High/Ultra High Performance Liquid Chromatography Performance (HPLC-UPLC)-spectrophotometry for endpoint detection of formazan was established (Alépée et al., 2015). In this study, HPLC-UPLC-spectrophotometry endpoint detection system for measurement of formazan was shown to be highly reproducible between different laboratories. Based on these findings, it was concluded that HPLC/UPLC-spectrophotometry is relevant to all *in vitro* RhT test methods irrespective of the test system and test method and can be applied to any of the other RhT test systems within the relevant OECD Test Guidelines. OECD TGs 431 (skin corrosion) and 439 (skin irritation) have been revised and the newly adopted 492 (eye irritation / serious eye damage) written to incorporate use of HPLC/UPLC-spectrophotometry as an alternative endpoint detection system for measurement of formazan (OECD 2015a; 2015c; 2015d).

Applying the approach, the current project describes evaluation of chemicals in the SkinEthic™ HCE EIT test method in which both standard photometry (OD) and HPLC/UPLC-spectrophotometry are used as endpoint detection systems for measurement of formazan thereby enabling a direct comparison.
A total of 24 chemicals, 11 liquids and 13 solids, representing non-coloured and coloured chemicals were selected and are listed in Attachment 2. Prior to the testing, each chemical was checked for its colourant and/or MTT reducing properties as described in the SOP in order to determine the use of adapted controls for the determination of non-specific colouration and/or MTT reduction. The viability of the chemicals was assessed using photometric MTT-reduction and HPLC/UPLC-spectophotometry. The SkinEthic™ HCE EITL or EITS protocols were performed for liquids and solids, respectively. The resulting formazan tissue extracts were analysed by OD and HPLC/UPLC-spectrophotometric analysis in a laboratory. The agreement in viability between the MTT and HPLC/UPLC-spectrophotometric method was assessed with a scatter plot. The line of equality was used as a visual tool for agreement. A dot that falls on the line or that is close to the line corresponds with a chemical with equal viability values or values close to each for the different analytical methods.

The MTT reducing and colour interfering properties are presented in Attachment 11.
Of the 24 chemicals evaluated, measurement of formazan by OD resulted in an *in vitro* classification for 19 (79 %) chemicals whilst for HPLC/UPLC-spectrophotometry an *in vitro* classification could be derived for all of the test chemicals (100%).
Photometric measurements (MTT) could not be obtained for 5 test items, 3 chemicals (CAS RN 134429-57-5, 1686090-84-5, 74578-10-2) and Benzenamine, 4,4'-[(4-imino-3-methyl-2,5-cyclohexadien-1-ylidene)methylene]bis[2-methyl-, hydrochloride (1:1) (CAS RN 3248-91-7) which was evaluated neat and in dilution (1% (w/v) aqueous). For three chemicals (CAS RN, 1686090-84-5, 74578-10-2 and 3248-91-7) the % viability observed

of either chemical-treated tissue and/or of adapted controls (%NSMTT, $\%NSC_{living}$, $\%NSC_{killed}$) was identified to be outside the linearity range of the spectrophotometer.
For the remaining chemical (CAS RN 134429-57-5) high variability of the adapted $\%NSC_{living}$ controls viability values between tissues was observed for which the difference was >20. This means that it was not possible for these 5 chemicals to determine final % viability values due to the experiments being identified as unqualified.
The remaining 19 chemicals covered 9 chemicals without MTT-reducing and colouring properties and 10 chemicals with MTT-reducing and/or colouring properties. The figure 1 demonstrates that very similar or identical viability measurements were obtained for both methods since the viabilities are close to the line of unity. In addition, the regression line coincides with the line of unity, the values are the same on average. This supports the findings of the study published by Alépée and co-workers (2015) that HPLC/UPLC-spectrophotometry and OD measurements gave similar results in terms of tissue viability.

Fig. 1. Tissue viability (%) quantified from formazan tissue extracts by OD and HPLC/UPLC-spectrophotometry for 19 chemicals. The straight line corresponds with the line of equality and the dotted line corresponds with the linear regression line.



Of the 24 test items for which historic *in vivo* data are available and for which the formazan could be measured by HPLC/UPLC-spectrophotometry, the concordance between the *in vitro* and *in vivo* classifications was 83% (20/24). Four chemicals (CAS RN 2370-63-0, 66170-10-3, 23920-15-2, and 60207-90-1) were identified as Classified *in vitro* which is discordant with the *in vivo* classification. Within these test chemicals known to be Not-classified *in vivo*, 2 chemicals (66170-10-3 and 60207-90-1) were also classified by the EpiOcular™ EIT test method using OD and/or HPLC/UPLC-spectrophotometry (Alépée et al, 2015). In addition, the results identify that, for all chemicals known to be classified (Cat.1 / Cat.2) *in vivo*, the *in vitro* classification based on formazan measurement by both OD (9/9) and HPLC/UPLC-spectrophotometry (13/13) correlated correctly with the classification as classified. Overall, these results

demonstrate that HPLC/UPLC-spectrophotometry does not over- not under-predict relative to OD as the endpoint detection system.

In conclusion, HPLC/UPLC-spectrophotometry may be used with all types of test chemicals (coloured, non-coloured, MTT-reducers and non-MTT reducers) for measurement of MTT formazan. As strong colour chemicals can interfere with the optical detection method, HPLC/UPLC-spectrophotometry is recommended to overcome this problem. Besides, this alternative endpoint detection can be used for all chemicals belonging to the applicability domain of the SkinEthic™ HCE EIT test method.

References

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

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

OECD (2015a). Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. OECD Guideline for the Testing of Chemicals No. 492, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines]

OECD (2015c). *In vitro* skin corrosion: Reconstructed Human Epidermis Test Method. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines]

OECD (2015d). *In vitro* skin irritation: Reconstructed Human Epidermis Test Method. OECD Guideline for the Testing of Chemicals No. 439, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines]

3.2 Additional Information - Proficiency

Please add any additional information that was not covered in the TST.

[…]

3.3 List of references

Please provide a list of scientific or other appropriate references, briefly describing their relevance with regard to the submitted test method. Please include all publications that provide background information on the test method's biological and mechanistic relevance. If available, please include publications that provide direct information of the performance of the proposed test method. In cases of submissions of putative similar or updated test methods, please include publications relating to the validated reference test method(s) on which – if available – the performance standards (PS) are based. Please append publications as electronic files (pdf or scans of paper copies) as Attachment 17.

Adriaens E, Barroso J, Eskes C, Hoffmann S, McNamee P, Alépée N, Bessou-Touya S, de Smedt A, De Wever B, Pfannenbecker U, Tailhardat M, Zuang V (2014). Retrospective analysis of the Draize test for serious eye damage/eye irritation: importance of understanding the in vivo endpoints under UN GHS/EU CLP for the development and evaluation of in vitro test methods. Arch Toxicol 88:701-723. doi: 10.1007/s00204-013-1156-8.

Agresti, A., Coull, B.A. (1998). Approximate is better than exact for interval estimation of binomial proportions. Am. Stat. 52, 119126.

Alépée N, Bessou-Touya S, Cotovio J, de Smedt A, de Wever B, Faller C, Jones P, Le Varlet B, Marrec-Fairley M, Pfannenbecker U, Tailhardat M, van Goethem F, McNamee P. (2013).Cosmetics Europe multi-laboratory pre-validation of the SkinEthic™ reconstituted human corneal epithelium test method for the prediction of eye irritation.Toxicol In Vitro. 27(5):1476-88.

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

Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, Reinhardt CA, Repetto G, Sladowski D, Spielmann H & Zucco F. (1995a). Practical Aspects of the Validation of Toxicity Test Procedures. The Report and Recommendations of ECVAM Workshop 5. ATLA 23: 129-147.

Balls., M., Botham, P.A., Bruner, L.H., Spielmann, H., 1995b. The EC/HO international validation study on alternatives to the draize eye irritation test. Toxicol. In Vitro 9, 871-929. doi: 10.1016/0887-2333(95)00092-5.

Barroso, J., Alépée, N., De Smedt, A., De Wever, B., Hibatallah, J., McNamee, P., Mewes, K.R., Millet, M., Pfannenbecker, U., Tailhardat, M., Templier, M. (2013). The importance of understanding drivers of irritation in vivo for selection of chemicals used in the development and evaluation of in vitro eye irritation assays: Cosmetics Europe analysis. Poster presented at the In Vitro Testing Industrial Platform (IVTIP) Spring 2013 Meeting, Southampton, UK, May 14–16, 2013. Available from: https://www.cosmeticseurope.eu/downloads/6057.html. Accessed on 05/10/2015.

Barroso, J., N. Alépée, T. Cole, C. Eskes, S. J. Freeman, R. Liška, P. McNamee, U. Pfannenbecker, A. A. Reus, C. M. Rubingh, M. W. Schaeffer and V. Zuang. (2014). EURL ECVAM – Cosmetics Europe prospective validation study of Reconstructed human Tissue-baseds test methods for serious eye damage/ eye irritation testing 9th World Congress on Animal Alternatives, Prague, Altex Proceedings (ISSN 2194-0479), Vol 3, No. 1, p. 99.

Barroso, J., Alépée, N., Cole, T., Eskes, C., Freeman, S.J., Liska, R., McNamee, P., Pfannenbecker, U., Reus, A.A., Rubingh, C.M., Schaeffer, M.W., Zuang, V. (2015). EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Tissue-based test methods for serious eye damage/eye irritation testing. Poster presented at the Congress of the European Societies of Toxicology (EUROTOX), Porto, Portugal, September, 13-16, 2015. Available from: http://dx.doi.org/10.1016/j.toxlet.2015.08.507. Accessed on 05/10/2015.

Curren, R.D., Southee, J.A., Spielmann, H., Liebsch,. M., Fentem, J.H. & Balls, M. (1995).The role of prevalidation in the development, validation and acceptance of alternative methods. Altern. Lab. Anim. 23, 211-217.

Cotovio, J., Grandidier, M.H., Lelièvre, D., Bremond, C., Flamand, N., Loisel-Joubert, S., Van Der Lee, A., Capallère, C., Meunier, J.R., Leclaire, J. (2008). The use of the reconstructed human corneal model (HCE) to assess in vitro eye irritancy of chemicals. AATEX 14, 343–350 (Special issue).

Cotovio J., Grandidier M.H., Lelièvre D., Bremond C., Amsellem C., Maloug S., Ovigne J.M., Loisel-Joubert S., Van Der Lee A., Minondo A.M., Capallere C., Bertino B., Alépée N., Tinois-Tessonneaud E., De Brugerolle de Fraissinette A., Meunier J.R., Leclaire J. (2010). In vitro assessment of eye irritancy using the Reconstructed Human Corneal Epithelial SkinEthic™ HCE model: Application to 435 substances from consumer products industry. Toxicol In Vitro 24(2), 523-37.

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

EC (2008). Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Off J Eu Union L353:1–1355.

EC EURL ECVAM (2014). Validation Study Report on the EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Corneal Epithelium-based test methods for identifying chemicals not requiring classification and labelling for serious eye damage/eye irritation testing. Available at: in publication.

Fox, D.A., and Boyes, W.K. (2008). Toxic responses of the ocular and visual system. In Cassaret and Doull's Toxicology: The basic science of poisons (ed. C.D. Klaassen), 7th edn, pp. 665–697. Withby, ON, Canada: McGraw-Hill Ryerson.

Freeman S.J., Alépée N., Barroso J., Cole T., Compagnoni A., Rubingh C., Eskes C., Lammers J., McNamee P., Pfannenbecker U., Zuang V. (2010). Prospective validation study of reconstructed human tissue models for eye irritation testing. ALTEX 27, Special Issue 2010, 261-266.

Hackett, R.B., and McDonald, T.O. (1991). Eye Irritation. In Advances in Modern Toxicology: Dermatoxicology (ed. F.N. Marzulli & H.I. Maibach), 4th edn, pp. 749–815. Washington, DC, USA: Hemisphere Publishing Corporation.

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.

Jester, J.V., Li, H.F., Petroll, W.M., Parker, R.D., Cavanagh, H.D., Carr, G.J., Smith, B., and Maurer, J.K. (1998a). Area and depth of surfactant induced corneal injury correlates with cell death. Invest. Ophthalmol. Vis. Sci. 39, 922–936.

Jester, J.V., Petroll, W.M., Bean, J., Parker, R.D., Carr, G.J., Cavanagh, H.D., and Maurer, J.K. (1998b). Area and depth of surfactant-induced corneal injury predicts extent of subsequent ocular responses. Invest. Ophthalmol. Vis. Sci. 39, 2610–2625.

Jester, J.V., Li, L., Molai, A., and Maurer, J.K. (2001). Extent of corneal injury as a mechanistic basis for alternative eye irritation tests. Toxicol. in Vitro 15, 115–130.

Jester, J.V. (2006). Extent of corneal injury as a biomarker for hazard assessment and the development of alternative models to the Draize rabbit eye test. Cutan. Ocul. Toxicol. 25, 41–54

Kaluzhny, Y., Kandarova, H., Handa, P., Kunilus, J., d'Argembeau-Thornton, L., Klausner, M., 2011. DEvelopment of the EpiOcularTM Eye Irritation Test (EpiOcular-EIT) for hazard identification and labelling of eye irritating chemicals in response to the requirements of the cosmetics directive and REACH legislation. ATLA 39, 339-364.

Maurer, J.K., Parker, R.D., and Jester, J.V. (2002). Extent of corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. Reg. Tox. Pharmacol. 36, 106-117.

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

Nguyen, D.H., Beuerman, R.W., de Wever, B., Rosdy, M. (2003). Three-dimensional construct of the human corneal epithelium for in vitro toxicology. Salem, H., Katz, S.A. (Eds.), Alternative Toxicological Methods. CRC Press, pp. 147–159.

OECD (1999) OECD Series of Good Laboratory Pactice 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). 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 (2012). Acute Eye Irritation/Corrosion. OECD Guideline for Testing of Chemicals No. 405. Organisation for Economic Co-operation and Development, Paris. Available at: [http://www.oecd-ilibrary.org/environment/test-no-405-acute-eye-irritation-corrosion_9789264185333-en].

OECD (2015a). Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying Chemicals Not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. OECD Guideline for the Testing of Chemicals No. 492, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines]

OECD (2015b). Performance standards, Series on Testing & Assessment No. 216. Performance Standards For The Assessment Of Proposed Similar Or Modified In Vitro Reconstructed Human Cornea-Like Epithelium (Rhce) Test Methods For Eye Hazard. Available at: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)23&docLanguage=En]

OECD (2015c). In vitro skin corrosion: Reconstructed Human Epidermis Test Method. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris. Available at URL: [http://www.oecd.org/env/testguidelines]

OECD (2015d). In vitro skin irritation: Reconstructed Human Epidermis Test Method. OECD Guideline for the Testing of Chemicals No. 439, OECD, Paris. Available at URL: [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., and Van Goethem, F. (2013). 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.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundati on for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

UN (2013). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30, Fifth revised edition, New York and Geneva: United Nations. Available at: [http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf].

Van Goethem F., Adriaens E., Alépée N., Straube F., De Wever B., Cappadoro C., Catoire S., Hansen E., Wolf A., Vanparys P. (2006). Prevalidation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. Toxicol. in Vitro 20, 1–17.

Wilson S.L, Ahearne M., Hopkinson A. (2015). An overview of current techniques for ocular toxicity testing. Toxicology 327 (2015) 32–46

4. ATTACHMENTS

Attachment Number	Description	Tick if attached	File name
Attachment 1a	Protocol of the SkinEthic™ HCE EITL (Liquid)	X	
Attachment 1b	Protocol of the SkinEthic™ HCE EITS (Solid)	X	
Attachment 1c	SkinEthic™ HCE EITL (Liquid) DB-ALM protocol	X	
Attachment 1d	SkinEthic™ HCE EITS (Solid) DB-ALM protocol	X	
Attachment 2	SkinEthic™ HCE EIT - List of test items including their CAS number and basic physical/chemical properties for optimisation/transfer/WLR/BLR/Predictive capacity	X	
Attachment 3a	SkinEthic™ HCE EIT - Data used for relevance and reliability assessment (EITL and EITS)	X	
Attachment 3b	SkinEthic™ HCE EITL - WLR assessment (60 chemicals – 3 labs)	X	
Attachment 3c	SkinEthic™ HCE EITL - WLR assessment (45 additional chemicals – 1 lab)	X	
Attachment 3d	SkinEthic™ HCE EITS - WLR assessment (60 chemicals – 3 labs)	X	
Attachment 3e	SkinEthic™ HCE EITS - WLR assessment (35 additional chemicals – 1 lab)	X	
Attachment 4a	Training protocol of the SkinEthic™ HCE EITL (Liquids)	X	
Attachment 4b	Training protocol of the SkinEthic™ HCE EITS (Solids)	X	

Attachment 5a	Training report of the SkinEthic™ HCE EITL (Liquids) - VITO	X	
Attachment 5b	Training report of the SkinEthic™ HCE EITL (Liquids) - CRL	X	
Attachment 5c	Training report of the SkinEthic™ HCE EITS (Solids) - VITO	X	
Attachment 5d	Training report of the SkinEthic™ HCE EITS (Solids) – CRL	X	
Attachment 6a	Transfer report of the SkinEthic™ HCE EITL (Liquids) – VITO & CRL	X	
Attachment 6b	Transfer report of the SkinEthic™ HCE EITS (Solids) – VITO & CRL	X	
Attachment 7a	SkinEthic™ HCE EITL - BLR assessment (60 chemicals – 3 labs)	X	
Attachment 7b	SkinEthic™ HCE EITS - BLR assessment (60 chemicals – 3 labs)	X	
Attachment 8a	Statistical analysis and reporting of the SkinEthic™ HCE EITL (Liquids)	X	
Attachment 8b	Statistical analysis and reporting of the SkinEthic™ HCE EITS (Solids)	X	
Attachment 9	SkinEthic™ HCE EIT – Predictive capacity assessment	X	
Attachment 10a	Project plan of the SkinEthic™ HCE EITL (Liquids)	X	
Attachment 10b	Project plan of the SkinEthic™ HCE EITS (Solids)	X	
Attachment 11	SkinEthic™ HCE EIT - HPLC/UPLC-spectrophotometry (24 chemicals – 1 lab)	X	

NOTE: *Please label appended files by indicating the relevant Attachment Number at the beginning of the file name. If more than one file needs to be attached for each description, please use ZIP compression*

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
The European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Explanatory Note on the ECVAM Test Submission Template (TST)

1. Purpose of the TST

This Test Submission Template (TST) is the ECVAM standard reporting format for Step 2 of the ECVAM process for submission of test methods for pre-validation, validation, evaluation and/or peer-review.

The TST and all accompanying material (Annexes) will be used by ECVAM to evaluate test methods proposed for (pre-)validation. It will enable ECVAM to decide on the type of validation process most suitable for assessing the scientific validity of the proposed test method. In addition, the TST can be used to submit finalised external validation studies for ECVAM evaluation of the test method's readiness to enter peer review. In the specific case of performance standards-based external validation of similar ("me-too") or updated test methods, the test submitter may be asked to submit a test pre-submission followed by a full submission using the test submission template, before initiation of the validation study to allow for an assessment of similarity (compliance with Essential Test Method Components) and a preliminary assessment of compliance with the target Reliability and Accuracy values. However, ECVAM will not comment at this stage on study design (project plan), which is out of the scope for evaluating compliance with the performance standards as described above. Such information should only be submitted after completion of the external performance standards-based validation study to allow for an evaluation of the test method's readiness to enter peer review.

It should be noted though that all validation studies, either coordinated by ECVAM or externally should follow some good validation practices which should be laid down in a well-defined project plan. Being aware of differences which may occur between the test methods which need to be validated, a good project plan should include (but does not need to be limited to) the following points:

1) Study objective and goals

2) Test method description

3) Management structure (composition and role of the different actors involved in the study): Validation Management Group (VMG); study coordinator; study sponsor(s); chemicals selection group; entity/person responsible for chemicals acquisition, coding and distribution; participating laboratories (experienced and naïve); entity/person responsible for biostatistics; etc.

4) Chemicals selection criteria

5) Procedure for chemical acquisition, coding, distribution, receipt and handling

6) Identification of study directors, safety officers, quality assurance directors and experimental team in each participating laboratory and description of their roles and responsibilities

7) Study design (including instructions on the number of replicates measurements per test, the number of tests per test item, the way in which unqualified data are dealt with, the number of re-testing allowed, etc.)

8) Data collection, handling and analysis

9) Quality assurance/good laboratory practice in the participating laboratories and at the test system producers (quality assurance audits at the production sites may be considered if production is not done under GMP)

10) Timelines of the study.

The types of validation processes include (1) Prospective Pre-validation, (2) Prospective Validation, (3) Retrospective Validation, (4) Validation based on Performance Standards, i.e. validation of new or updated test methods found sufficiently similar to the validated reference method (colloquially referred to as "catch-up" validation).

As a general rule, the information submitted in the TST should be:

- as complete as possible, depending also on the type of validation sought (for details see Table 1)
- as detailed as necessary, but
- as concise as possible.

Please note that the TST can be used for submission of single test methods as well as of testing strategies composed of more than one test method.

2. Validation of alternative methods

Alternative test methods are intended to reduce, refine or replace the use of experimental animals, and are commonly referred to as the Three Rs methods, as first described by Russell and Burch in "*The Principles of Humane Experimental Technique*" (Methuen, London, 1959).

Based on a requirement outlined in EU Directive 86/609/EEC on the protection of animals used for experimental and other scientific purposes, which states that the Commission and the Member States should actively support the development, validation and acceptance of alternative test methods, in 1991, the European Commission established the European Centre for the Validation of Alternative Methods (ECVAM) *[Communication SEC(91)1794 of the European Commission to the Council and the European Parliament]*.

Validation is defined as the process by which the reliability and relevance of a particular approach, method, process or assessment, is established for a defined purpose (OECD, 2005). Since the validation process generates and/or assesses empirical information on reliability and relevance of a test method under standardised and controlled conditions, it is generally accepted to facilitate and/or accelerate the international (regulatory) acceptance of alternative test methods.

In 1995, based upon experience from large-scale validation studies, and in consultation with various international experts, ECVAM published recommendations concerning the practical and logistical aspects of validating alternative test methods in prospective studies (Balls et al, 1995). These criteria were then taken up internationally and are described in the OECD Guidance Document 34 (OECD, 2004).

Different **types of validation** processes exist and are described below:

Pre-validation: The initial phase(s) of a validation study. A small-scale study intended to obtain preliminary information on the relevance and reliability of a test method. Based on the outcome of a pre-validation study, the test method protocol may be modified or optimised to increase intra- and/or inter-laboratory reproducibility and accuracy in subsequent validation studies (OECD, 2005).

Prospective validation: An approach to validation when some or all information necessary to assess the validity of a test are not available, and therefore new experimental work is required (OECD, 2005).

Retrospective validation: An assessment of the validation status of a test method carried out by considering all available information and data generated from one or more validation studies (OECD, 2005).

Validation Based on Performance Standards: A validation study for a test method that is structurally and functionally similar to a previously validated and accepted reference test method. The candidate test method should incorporate the essential test method components included in performance standards developed for the reference test method, and should have comparable performance when evaluated using the reference chemicals provided in the performance standards (OECD, 2005).

3. Information requirements

ECVAM developed a **modular approach** to the validation of alternative methods (Hartung et al., 2004), where the various information requirements for peer-review and as generated during the validation process are broken down into 7 independent modules.

Note that according to this modular approach, the information requirements can be fulfilled by using data obtained from a prospective study, by a retrospective evaluation of already existing data/information, or by a combination of both.

The TST is structured according to these modules and the information necessary to complete each module is summarised below:

Module 1 – Test Definition
This module defines the scientific purpose of the test and describes the mechanistic basis of the test in view of broader current scientific knowledge of the test endpoint and the definitive protocol which should include specification of the endpoints, endpoint measurement, derivation and expression of results, interpretation of results via a prediction model and the inclusion of adequate controls.

Module 2 – Within-Laboratory Reproducibility
This module addresses the reproducibility of results within a single laboratory over time, using a defined protocol and the same laboratory set-up.

Module 3 – Transferability
The transferability measures the ability of a test method protocol to be accurately and reliably performed in independent competent laboratories. It therefore provides an estimate of how much training is needed to be able to perform the test in a naïve laboratory and produce reproducible results. A naïve laboratory refers to a laboratory that is inexperienced in performing the test method. The transferability gives an indication on the robustness of a test method (e.g. its reliable performance under different conditions).

Module 4 – Between-Laboratory Reproducibility
This module addresses the reproducibility of results in different qualified laboratories, using the same protocol and testing the same test items. The between-laboratory reproducibility is usually assessed in three well-trained laboratories. However, the number of laboratories as well as the number and type of test items should be decided according to the purpose of the validation study.

Module 5 – Predictive Capacity
The predictive capacity determines the ability of a test method to predict the *in vivo* result and/or a human health effect of concern. This is usually done by relating the predictions to the result obtained with a reference method. The predictive capacity calculated is influenced by the number of test items (sample size) and the quality of the reference method.

Module 6 – Applicability Domain

The Applicability Domain of an *in vitro* test method is defined by its limitations and by considerations of the physico-chemical or other properties of the substances for which a method is applicable for use as determined from empirical testing.

Module 7 – Performance standards

Upon completion of a validation study, performance standards of the validated test method are defined. These are (i) essential test method components, (ii) defined reference (test) chemicals, (iii) defined accuracy values. Performance standards allow the validation of tests found to be sufficiently similar with respect to the validated test method. Performance standards can also be applied to assess modifications of already validated tests.

The following table provides an indication of the information required for entering a specific validation process. Depending on a number of factors such as, for example, the purpose of the test method (screening, partial replacement, full replacement), as well as practical and economical considerations, the extent of the information requirements may vary.

Table 1. Information required for entering a specific validation process.

Nr.	Module	Pre-validation Study	Prospective Validation Study	Retrospective Validation Study	Performance Standard-Based Validation Study	Peer Review
1	**TD**	Required (however only preliminary PM, if applicable)	Required	Required	Required	Required
2	**WLR**	Preliminary information required	Preliminary information required	Required	Required	Required
3	**TR**	If available	Preliminary information required	Preliminary information, if available	If available	Required
4	**BLR**	If available	Preliminary information required	Preliminary information required	If available	Required
5	**PC**	Preliminary information, if available	Preliminary information required	Preliminary information required	Required (at least for reference chemicals)	Required
6	**AD**	If available	Preliminary information on the test method's limitations required	Preliminary information on the test method's limitations required	Not applicable (assumed to be the same as for the reference method)	Required
7	**PS**	Not required	Not required	Not required	Not applicable	If available

4. Abbreviations used in the TST

- AD Applicability Domain
- BLR Between Laboratory Reproducibility
- ECVAM European Centre for the Validation of Alternative Methods
- ESAC ECVAM's Scientific Advisory Committee
- GCCP Good Cell Culture Practice
- GLP Good Laboratory Practice
- IPRs Intellectual Property Rights
- PC Predictive Capacity
- PM Prediction Model
- PS Performance Standards
- TF Transferability
- TST Test Submission Template
- WLR Within Laboratory Reproducibility

5. Note regarding terms

Below is a list of terms related to test method validation, as defined in OECD Guidance Document No 34 (OECD, 2005) or defined for the purpose of the TST (identified with an *).

***Benchmark:** A test item that produces a midrange response in the test method, i.e. to assess variability of the test system over time. Please note that if positive controls elicit a too strong response, they cannot be used as benchmark.

Between-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results.

***Biological relevance:** Relates to the extent to which the test methods models or reproduces the biological properties of target organ/system or species of interest (e.g. mechanism of action, cell types, cytoarchitecture).

Endpoint: The biological or chemical process, response, or effect, assessed by a test.

False negative: A substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

False negative rate: The proportion of all positive substances falsely identified by a test method as negative. It is one indicator of test method performance.

False positive: A substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

False positive rate: The proportion of all negative (non-active) substances that are falsely identified as positive. It is one indicator of test performance.

***Negative Control:** The vehicle used and/or a test item known not to elicit a positive response in the test method.

***Non-qualified test:** A test that does not meet the acceptance criteria defined in the protocol.

Performance standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (1) essential test method components; (2) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of reference chemicals.

***Positive Control:** A test item well known to elicit a positive response in the test method.

***Predictive capacity:** The ability of a test method to make relevant predictions on defined biological effects (e.g. human health effects).

Prediction Model: a formula or algorithm (e.g., formula, rule or set of rules) used to convert the results generated by a test method into a prediction of the (toxic) effect of interest. Also referred to as decision criteria. A prediction model contains four elements: (1) a definition of the specific purpose(s) for which the test method is to be used; (2) specifications of all possible results that may be obtained, (3) an algorithm that converts each study result into a prediction of the (toxic) effect of interest, and (4) specifications as to the accuracy of the prediction model (e.g., sensitivity, specificity, and false positive and false negative rates). Prediction models are generally not used in in vivo ecotoxicological tests.

***Qualified test:** A test that meets the acceptance criteria defined in the protocol.

Relevance: 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 the accuracy (concordance) of a test method.

Reliability: Measures of the extent that a test method can be performed reproducibly within- and between laboratories over time, when performed by using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

Reproducibility: The agreement among results obtained from testing the same substance using the same test protocol (see Reliability).

***Run:** A run consists of one or more test items tested concurrently, and if applicable, also with a positive and a negative control.

Sensitivity: The proportion of all positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method.

Specificity: The proportion of all negative/inactive substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method.

***Test:** The use of a test method for testing a single test item within a single experiment (can be composed of one single measurement or several measurements conducted in parallel, i.e. "replicates").

***Test Item:** Any entity to be tested with the test method. These may be single substances, mixtures, biologicals, etc.

Test method: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay".

***Test System:** A test method is usually composed of three elements: (i) test system, (ii) protocol, (iii) prediction model. The test system is the biological/chemical system that is exposed to the test items to obtain experimental data.

Transferability: The ability of a test procedure to be accurately and reliably performed in independent, competent laboratories.

***Variability:** Within- and between laboratory variability are also referred to as within- and between laboratory reproducibility. Although reproducibility is generally the preferred term, it refers to the same concept as variability, often used in the literature. The latin prefixed intra- and inter- are often replaced with the English translation within- and between.

Within-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times.

6. References