Toxicology in Vitro 31 (2016) 43–53

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

Multi-laboratory validation of SkinEthic HCE test method for testing serious eye damage/eye irritation using liquid chemicals

N. Alépée [a,*], V. Leblanc [a], E. Adriaens [b], M.H. Grandidier [a], D. Lelièvre [c], M. Meloni [d], L. Nardelli [a], C.S. Roper [e], E. Santirocco [d], F. Toner [e], A. Van Rompay [f], J. Vinall [e], J. Cotovio [a]

[a] L'Oréal Research & Innovation, Aulnay Sous Bois, France
[b] Adriaens Consulting, Aalter, Belgium
[c] Episkin SA, Lyon, France
[d] VitroScreen, Milano, Italy
[e] Charles River Laboratories, Edinburgh, United Kingdom
[f] VITO NV (Flemish Institute for Technological Research), Mol, Belgium

ARTICLE INFO

Article history:
Received 13 August 2015
Received in revised form 14 November 2015
Accepted 19 November 2015
Available online 1 December 2015

Keywords:
Inter-laboratory validation
SkinEthic HCE
Reproducibility
Performance
Serious eye damage-irritation
UN GHS Classification and Labelling of Chemicals

ABSTRACT

A prospective multicentric 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 protocol for liquids and define its predictive capacity. Briefly, 60 chemicals were three times tested (double blinded) in 3 laboratories and 45 additional chemicals were tested three times in one laboratory. Good within laboratory reproducibility was achieved of at least 88.3% (53/60) and 92.4% (97/105) for the extended data set. Furthermore, the overall concordance between the laboratories was 93.3% (56/60). The accuracy of the SkinEthic™ HCE EITL for the extended dataset, based on bootstrap resampling, was 84.4% (95% CI: 81.9% to 87.6%) with a sensitivity of 99.0% (95% CI: 96.4% to 100%) and specificity of 68.5% (95% CI: 64.0% to 74.0%), thereby meeting all acceptance criteria for predictive capacity. This efficient transferable and reproducible assay is a promising tool to be integrated within a battery of assays to perform an eye irritation risk assessment.
© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Considerable progress has been made in the partial replacement of the regulatory *in vivo* Draize rabbit eye test. Currently, four test methods are accepted by the Organisation for Economic Co-operation and Development (OECD) to classify chemicals as inducing serious eye damage according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS Category 1) (UN, 2013). Two test methods are organotypic assays, the Bovine Corneal Opacity and Permeability (BCOP) test method (OECD Test Guideline (TG) 437) and the Isolated Chicken Eye (ICE) test method (OECD TG 438) (OECD, 2013a, 2013b). Two test methods are performed on confluent monolayers: the Fluorescein Leakage test method (OECD TG 460) is a cytotoxicity and cell function-based assay and the Short-Time Exposure (STE) test method (OECD TG 491) is a cytotoxicity based assay (OECD, 2012, 2015a). Furthermore, at this time, three methods are accepted by the OECD for the identification of chemicals not requiring a classification for serious eye damage/eye irritation (UN GHS No Category). The

Abbreviations: BCOP, Bovine Corneal Opacity and Permeability; BLR, between laboratory reproducibility; C, classified; CAS RN, Chemical Abstracts Service Registry Number; Cat 1, UN GHS classification for chemicals causing irreversible effects on the eye/serious damage to the eye; Cat 2, UN GHS 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); CI, confidence interval; CM, Cytosensor Microphysiometer; CO, corneal opacity; CRL, Charles River Laboratories; EIT, Eye Irritation Test; EITL, Eye Irritation Testing of Liquids; ESAC, ECVAM Scientific Advisory Committee; EURL ECVAM, European Union Reference Laboratory for Alternatives to Animal Testing; HCE, Human Corneal Epithelium; ICCVAM, Interagency Coordinating Committee on the Validation of Alternative Methods; ICE, Isolated Chicken Eye; LO, L'Oréal; NgC, negative control; NC, not classified; 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, Non Specific Colour; NSC_{living}, Non Specific Colour on living tissues; NSC_{killed}, Non Specific Colour on killed tissues; NSMTT, non-specific reduction of MTT; MTT, 3-[4,5-dimethylthiazole-2-yl] 2,5-diphenyltetrazolium bromide; OD, optical density; OECD, Organisation for Economic Co-operation and Development; PC, positive control; PBS, phosphate-buffered saline; RhCE, Reconstructed human Cornea-like Epithelium; STE, Short-Time Exposure; S.O.P., standard operating procedure; 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.

* Corresponding author at: L'Oréal Research & Innovation, 1 Avenue Eugène Schueller, 93600 Aulnay-sous-bois, France.
E-mail address: nalepee@rd.loreal.com (N. Alépée).

http://dx.doi.org/10.1016/j.tiv.2015.11.012
0887-2333/© 2015 Elsevier Ltd. All rights reserved.

organotypic BCOP (OECD TG 437) and ICE (OECD TG 438) test methods were adopted for this purpose (OECD, 2013a, 2013b). The third test method uses a three-dimensional Reconstructed Human Cornea-like Epithelium (RhCE) (OECD TG 492) and measures cytotoxicity (OECD, 2015b). OECD TG 492 covers at this time only the commercially available EpiOcular™ Eye Irritation Test (EIT) validated by the European Union Reference Laboratory for Alternative to Animal testing (EURL ECVAM) and Cosmetics Europe in a prospective validation study (Freeman et al., 2010). In addition, the Cytosensor Microphysiometer (CM) (Hartung et al., 2010), a cytotoxicity and cell function-based method, has been endorsed as scientifically valid for the identification of Cat 1 and No Cat chemicals for limited applicability domains (ESAC, 2009; Interagency Coordinating Committee on the Validation of Alternative Methods ICCVAM, 2010). This method is currently in the process of review by the OECD. A more detailed description of the principle and background of the methods listed above is presented in the review published by Wilson et al. (2015).

Despite all these efforts, full replacement of the *in vivo* Draize rabbit eye test is however yet to be achieved. During a workshop held in 2005 by the European Centre for the Validation of Alternative Methods (ECVAM), a testing scheme was proposed using a Bottom-Up or Top-Down progression of *in vitro* tests (Scott et al., 2010). The outcome of this expert meeting identified that test methods that are using RhCE could be considered for incorporation into a testing strategy as an initial step in a Bottom-Up approach or the second step in a Top-Down approach. Currently, there are two such test methods available namely, the EpiOcular™ Eye Irritation Test (EIT) (Kaluzhny et al., 2011; Pfannenbecker et al., 2013; OECD, 2015b) and the SkinEthic™ Human Corneal Epithelium (HCE) test (Van Goethem et al., 2006; Cotovio et al., 2007, 2010; Alépée et al., 2013). In the past, various test protocols using the SkinEthic™ HCE model have been evaluated in order to improve the predictive capacity of this test method. Briefly, the method consists of topical exposure of the chemical onto the SkinEthic™ HCE test system for a defined time period after which the extent of cell injury is assessed by measurement of cytotoxicity. In a multicenter prevalidation study performed by Van Goethem et al. (2006), the validity of a 10-min exposure period (Short-Time Exposure) without post-incubation, was evaluated in four laboratories resulting in a predictive capacity of greater than 80%. Subsequent in-house evaluation of this protocol with a set of about 100 cosmetic ingredients showed an increase in specificity (probability of predicting no irritant given the true state is No Category) whereas the sensitivity (probability of predicting irritant given the true state is serious eye damage causing irreversible effects (Cat 1)/reversible effects on the eye/eye irritation UN GHS Category 2 (Cat 2))reduced substantially (unpublished data). In order to correctly identify the irritants which were under predicted with the 10-min treatment protocol, the exposure period was prolonged to 1 h (Long-time Exposure) followed by a post-incubation period of 16 h (Cotovio et al., 2007, 2010). Applying the UN GHS rules, in terms of differentiating between No Cat versus Classified (Cat 1/Cat 2), in combination with a threshold value of 50% viability to distinguish between irritants and non-irritants, an overall predictive capacity of 82% was obtained with 81% sensitivity and 82.8% specificity. In a next step, the validity of the SkinEthic™ HCE Short-time and Long-time Exposure protocols was evaluated with a set of 104 chemicals in a European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)/Cosmetics Europe prospective validation study (Freeman et al., 2010). None of the protocols met the acceptance criteria for predictive capacity (Barroso et al., 2014).

The present paper presents a further optimization of the SkinEthic™ HCE test method for the Eye Irritation Testing of Liquids (EITL protocol). 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). Whereas in the previous validation studies the aim was to obtain a balance between sensitivity and specificity, the purpose of the current study 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 (No Cat) in the majority of the runs. These values were based on the acceptance criteria set by the Validation Management Group (VMG) for the prospective validation study of RhCE-based test methods conducted by EURL ECVAM and Cosmetics Europe (EC EURL ECVAM, 2014; Kaluzhny et al., 2015; Barroso et al., 2015a).

2. Materials and methods

2.1. Tissues, media and reagents

The SkinEthic™ HCE tissues and tissue maintenance medium were purchased from Episkin SA (Lyon, France). Tissues were shipped in agarose semi-solid culture medium. Upon receipt, the tissue cultures were placed into 1 mL fresh maintenance medium (6-well plate) and incubated overnight in standard culture conditions (37 °C, 5% CO_2, ≥95% humidity). Following this equilibration period, the cultures were then transferred into a 24-well plate containing 300 μL fresh maintenance medium per well.

3-[4,5-Dimethylthiazole-2-yl] 2,5-diphenyltetrazolium bromide reagent (MTT), Ca^{2+}- and Mg^{2+}-free Dulbecco's phosphate-buffered saline (PBS), and methyl acetate were purchased from Sigma (CAS RN 79-20-9, Sigma-Aldrich, France).

2.2. Chemicals

2.2.1. Validation set

A total of 60 liquid chemicals representing different chemical classes were selected and are listed in Table 1. 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). Fig. 1 shows a scheme of the management structure of the study. 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 runs, performed with different tissue batches. During each run, a maximum of 13 liquid test chemicals, NgC and PC were all tested concurrently on two tissue replicates. A test chemical concurrently tested on two tissue replicates is called a test herein after.

2.2.2. Additional chemicals

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 liquid chemicals were evaluated unblinded by L'Oréal in three independent runs. The chemicals represented 22 non-classified and 23 classified, consisting of 11 Cat 1 and 12 Cat 2 chemicals (Table 2). In total, 105 liquid chemicals (50 non-classified and 55 classified chemicals, consisting of 27 Cat 1 and 28 Cat 2 chemicals) were evaluated on SkinEthic™ HCE test method for the Eye Irritation Testing of Liquids.

2.3. Participating laboratories

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

Table 1
Overview of the chemicals tested in the multicentre study.

No.	Chemical	CAS RN	Generic chemical class	Functional group class	UN GHS
1	1,9-Decadiene	1647-16-1	Neutral organic	Hydrocarbon-diene	No Cat
2	1,6-Dibromohexane	629-03-8	Neutral organic	Halogenated (electrophile)	No Cat
3	1-Bromo-4-chlorobutane	6940-78-9	Neutral organic	Halogenated (electrophile)	No Cat
4	1-Ethyl-3-methylimidazolium ethylsulphate	342573-75-5	Cation organic	Pyridinium, imidazolium	No Cat
5	Ethoxydiglycol	111-90-0	Neutral organic	Ether, alcohol	No Cat
6	2,2-Dimethyl-3-pentanol	3970-62-5	Neutral organic	Alcohol	No Cat
7	2-4-Pentanediol	625-69-4	Neutral organic	Alcohol	No Cat
8	2-Ethoxyethyl methacrylate	2370-63-0	Neutral organic	Ether, acrylates (electrophile)	No Cat
9	3-Phenoxybenzyl alcohol	13826-35-2	Neutral organic	Aromatic alcohol	No Cat
10	Dicaprylyl ether	629-82-3	Neutral organic	Ether	No Cat
11	Dipropyl disulphide	629-19-6	Neutral organic	Disulphide	No Cat
12	Ethyl thioglycolate	623-51-8	Neutral organic	Thiol ester	No Cat
13	Glycerol	56-81-5	Neutral organic	Alcohol, polyols	No Cat
14	Glycidyl methacrylate	106-91-2	Neutral organic	Acrylate, epoxide (electrophile)	No Cat
15	Iso-octyl acrylate	29590-42-9	Neutral organic	Acrylate (electrophile)	No Cat
16	Iso-propyl myristate (22-A; S2-10)	110-27-0	Neutral soap/surfactant	Ester	No Cat
17	n-Hexyl bromide	111-25-1	Neutral organic	Halogenated (electrophile)	No Cat
18	n-Octyl bromide	111-83-1	Neutral organic	Halogenated (electrophile)	No Cat
19	Octyltrimethoxysilane (SILAN 108)	3069-40-7	Neutral organic	Silicium, silane	No Cat
20	Piperonyl butoxide	51-03-6	Neutral organic	Aromatic ether	No Cat
21	p-Methyl thiobenzaldehyde	3446-89-7	Neutral organic	Aldehyde, thio-ether (electrophile)	No Cat
22	Polyethylene glycol 400	25322-68-3	Soap/surfactant	Alcohol, polyether	No Cat
23	Polyoxyethylene hydrogenated castor oil	61788-85-0	Neutral soap/surfactant	Ester, polyether	No Cat
24	Propylene glycol	57-55-6	Neutral organic	Alcohol, polyols	No Cat
25	Propylidynetrimethanol, propoxylated	25723-16-4	Neutral organic	Polyether, acrylate (electrophile)	No Cat
26	Hexamethyldisiloxane	107-46-0	Neutral inorganic	Silicium, siloxane	No Cat
27	Triphenyl phosphite	101-02-0	Neutral organic	Organophosphoric, aromatic	No Cat
28	Tween 20	9005-64-5	Neutral soap/surfactant	Ester, polyether	No Cat
29	2-Methyl-1-pentanol	105-30-6	Neutral organic	Alcohol	Cat 2B
30	3-Chloropropionitrile	542-76-7	Neutral organic	Nitrile, halogenated (electrophile)	Cat 2B
31	Di(propylene glycol) propyl ether	29911-27-1	Neutral organic	Alcohol, ether	Cat 2B
32	Diethyl toluamide	134-62-3	Neutral organic	Aromatic amide	Cat 2B
33	Ethyl-2-methylacetoacetate	609-14-3	Neutral organic	Ester, ketone	Cat 2B
34	Glycolic acid 10%	79-14-1	Organic acid	Alcohol, acid	Cat 2B
35	Iso-butanal	78-84-2	Neutral organic	Aldehyde (electrophile)	Cat 2B
36	Isopropyl acetoacetate	542-08-5	Neutral organic	Ester, cetone	Cat 2B
37	2,6-Dichlorobenzoyl chloride	4659-45-4	Neutral organic	Acid chloride, Halogenated (electrophile)	Cat 2A
38	Acetone	67-64-1	Neutral organic	Ketone	Cat 2A
39	Allyl alcohol	107-18-6	Neutral organic	Allyl alcohol (electrophile)	Cat 2A
40	Chlorhexidine gluconate 20%	18472-51-0	Organic base	Guanidine halogenated	Cat 2A
41	Cyclopentanol	96-41-3	Neutral organic	Cyclic alcohol	Cat 2A
42	Gammabutyrolactone	96-48-0	Neutral organic	Cyclic ester	Cat 2A
43	Propasol solvent P	1569-01-3	Neutral organic	Ether, alcohol	Cat 2A
44	Sodium hydroxide 1%	1310-73-2	Inorganic base	Alkali	Cat 2A
45	[3-(2-Aminoethylamino)propyl]trimethoxysilane	1760-24-3	Organic base	Amine, silicium, silane	Cat 1
46	1-Chloroctan-8-ol	23144-52-7	Neutral organic	Alcohol, halogenated (electrophile)	Cat 1
47	3-Methyl-pentynol	77-75-8	Neutral organic	Alcohol, alkyne	Cat 1
48	Acid Red 92 (22-D; S2-3) 10%[a]	18472-87-2	Organic acid	Phenol, halogenated aromatic, salt	Cat 1
49	Benzalkonium chloride 1%	63449-41-2	Cationic soap/surfactant	Ammonium salt	Cat 1
50	Benzethonium chloride 10%	121-54-0	Neutral organic	Ether, ammonium salt	Cat 1
51	Di(2-ethylhexyl)sodium sulphosuccinate 10%	577-11-7	Organic acid	Ester, sulphonic acid	Cat 1
52	Diethylaminopropionitrile	5351-04-2	Organic base	Amine, nitrile	Cat 1
53	Domiphen bromide 10%	538-71-6	Cationic soap/surfactant	Ammonium salt	Cat 1
54	Ethyl 2-hydroxyisobutyrate	80-55-7	Neutral organic	Alcohol, ester	Cat 1
55	Hexadecyltrimethylammonium bromide 10%	57-09-0	Cationic soap/surfactant	Alkylammonium salt	Cat 1
56	Hydroxyethyl acrylate	818-61-1	Neutral organic	Acrylate, alcohol (electrophile)	Cat 1
57	Lactic acid	50-21-5	Organic acid	Carboxylic acid, alcohol	Cat 1
58	Methyl thioglycolate	2365-48-2	Neutral organic	Carboxylic acid, ester, thioalcohol	Cat 1
59	Tetraethylene glycol diacrylate	17831-71-9	Neutral organic	Polyether, acrylate (electrophile)	Cat 1
60	Triton X-100 10%	9002-93-1	Soap/surfactant	Aromatic polyether	Cat 1

[a] 3,4,5,6-Tetrachloro-2-(1,4,5,8-tetrabromo-6-hydroxy-3-oxoxanthen-9-yl)-benzoic acid.

2.4. Technology transfer

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, using the EITL protocol. Both laboratory assistants tested 9 chemicals in two independent runs. This set of chemicals contained a strong colourant (Phloxine B-Acid Red 92, 10%) and an MTT reducer (Butyraldehyde). 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.

2.5. SkinEthic™ HCE tissue

2.5.1. Principle of the test system

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,

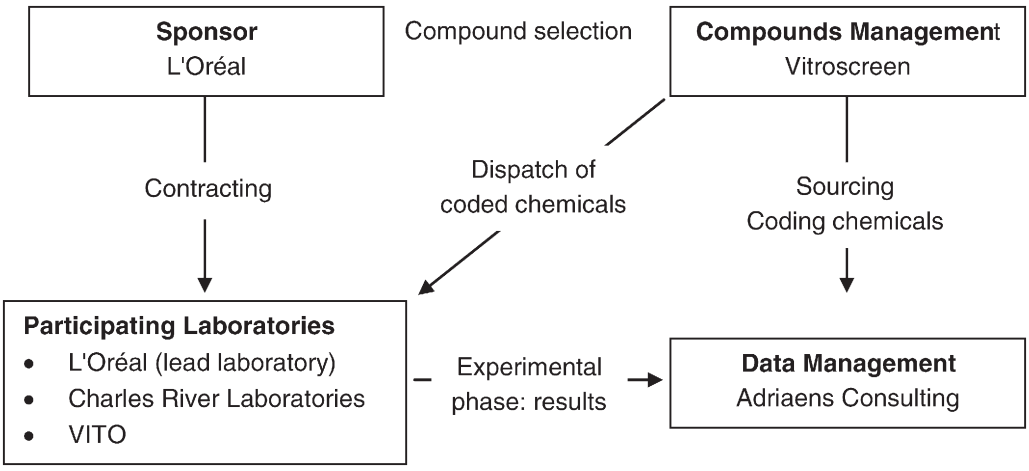

Fig. 1. Management structure of the SkinEthic™ HCE EITL validation study.

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).

2.5.2. Eye Irritation Test Liquid protocol

SkinEthic™ HCE tissues (0.5 cm^2) were topically exposed to 30 μL of undiluted liquid chemical for 30 ± 2 min at 37 °C at 5% CO_2 in a humidified incubator (standard culture conditions). Two tissues were used per test substance (NgC, PC, or chemical). After a 30 minute treatment, tissues were rinsed at least two times with 10 mL of PBS to remove the residual test substance 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.

2.5.3. Tissue viability assessment

Following the 30-minute incubation, tissues were carefully rinsed with PBS. Each tissue was transferred to a new well containing 300 μL of freshly prepared MTT (1 mg/mL) solution for a 3 h ± 15 minute incubation period under standard culture conditions. Then the tissue inserts were rinsed with 300 μL PBS and transferred into new plates containing 1.5 mL of isopropanol per well (750 μL underneath and 750 μL topically) for either 4 h at room temperature or overnight at 4 °C to extract the reduced MTT (formazan crystals) out of the tissues. 200 μL aliquots of formazan solution extracts were transferred to 96-well flat bottom microtitre plates for optical density (OD) measurement using a spectrophotometer equipped with a 570 nm filter (filter band pass ±30 nm). Isopropanol was used as a blank. The percentage viability of each of the treated cultures was calculated from the percentage MTT conversion in the treated cultures relative to the corresponding negative controls (100% viability). Results were expressed as mean OD and mean % viability and the difference of viability between the two replicate tissues.

2.5.4. Assessment for direct MTT reduction by the chemical

Possible interference of each chemical, e.g. ability to reduce MTT in absence of tissue, was verified before the start of the experiment. Therefore, 30 μL of the chemical was added to 300 μL of MTT solution (1 mg/mL), the mixture was incubated at 37 °C protected from light for 3 h. If the MTT solution colour turns blue or purple, the chemical interacts with the MTT. In case of MTT interaction, non-specific reduction of the MTT by the chemical needs to be determined in a separate experiment by using killed epithelial tissues. The killed tissues are treated with the chemical, rinsed and exposed to MTT according to the standard protocol. In addition, two killed tissues were treated with PBS as control. For the determination of the final viability, this non-specific reduction of MTT (%NSMTT) was taken into account and the viability was calculated as: the percent tissue viability obtained with living tissues exposed to the MTT reducer minus the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT).

2.5.5. Adapted controls for colouring chemicals

Coloured chemicals or chemicals able to develop a colour after contact with the tissue can generate a remaining Non Specific Colour (NSC). Therefore, each chemical was checked on a single occasion, for its colourant properties. In order to determine non-specific colouring, all steps of the EITL protocol were followed except the MTT incubation since 300 μL of maintenance medium was dispensed instead of MTT medium. The $\%NSC_{living}$ was determined after isopropanol extraction and OD reading in similar conditions. For the determination of the final viability, the $\%NSC_{living}$ was taken into account and the viability was calculated as: the percent tissue viability obtained with living tissues exposed to the colour interfering chemical and incubated with MTT solution minus the percent non-specific colour obtained with living tissues exposed to the colour interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected ($\%NSC_{living}$).

Therefore a coloured chemical can, in some cases, interfere with the MTT pre-check. In that case, each colouring chemical was applied onto two killed tissues and incubated in maintenance medium instead of MTT solution to determine the Non Specific Colour on killed tissues (NSC_{killed}). The final viability was calculated as: the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT minus $\%NSC_{living}$ plus the percent non-specific colour obtained with killed tissues exposed to the colour interfering test chemical and incubated with medium without MTT, calculated relative to the negative control ran concurrently to the test being corrected ($\%NSC_{killed}$).

2.6. Prediction model

Based on the relative viability the SkinEthic™ HCE EITL protocol can 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. No Cat (NC;

Table 2
Overview of the additional chemicals.

No.	Test substance	CAS RN	Generic class	Functional group class	UN GHS
61	1,2,6-Hexanetriol	106-69-4	Neutral organic	Alcohol, polyols	No Cat
62	1,3-Di-iso-propylbenzene	99-62-7	Neutral organic	Cyclic, aromatic phenyl	No Cat
63	1,3-Dibromopropane	109-64-8	Neutral organic	Halogenated (electrophile)	No Cat
64	1,4-Dibromobutane	110-52-1	Neutral organic	Halogenated (electrophile)	No Cat
65	1,5-Dibromopentane	111-24-0	Neutral organic	Halogenated (electrophile)	No Cat
66	3-Methoxy-1,2-propanediol	623-39-2	Neutral organic	Alcohol, ether polyols	No Cat
67	Cetylpyridinium bromide 0.1%	140-72-7	Cationic soap/surfactant	Onium compound, heterocyclic ammonium, aromatic, alkyl-pyridinium	No Cat
68	Cis-cyclooctene	931-87-3	Neutral organic	Hydrocarbon cyclic, cycloalkene	No Cat
69	DMSO	67-68-5	Neutral inorganic	Sulfoxide	No Cat
70	Ethanol 10%	64-17-5	Neutral organic	Alcohol	No Cat
71	Ethyl acetate	141-78-6	Neutral organic	ester	No Cat
72	Ethyl trimethyl acetate	3938-95-2	Neutral organic	Ester	No Cat
73	Gamma-methacryloxypropyltrimethoxy silane	2530-85-0	Neutral organic	Silicium, acrylate (electrophile)	No Cat
74	Methyltetraglycol	9004-74-4	Neutral organic	Polyether alcohol	No Cat
75	n-Butyl acetate	123-86-4	Neutral organic	ester	No Cat
76	Phloxine B/Acid Red 92 (22-D; S2-3)[a] 1%	18472-87-2	Organic acid	Phenol, halogenated	No Cat
77	Polyoxyethyelene 23 lauryl ether (Brij-35) (10%)	9002-92-0	Neutral soap/surfactant	Polyether	No Cat
78	Sodium salicylate (22-I; S2-14) 10%	54-21-7	Organic salts	Carboxylic acid, phenol	No Cat
79	Triethanolamine orthovanadate (30%)	13476-99-8	Neutral inorganic	Cetone	No Cat
80	Triethylene glycol monomethyl ether	112-35-6	Neutral organic	Ether, alcohol	No Cat
81	Triethylene glycol	112-27-6	Neutral organic	Ether, alcohol	No Cat
82	Triton X-100 1%	9002-93-1	Soap/surfactant	Aromatic polyether, alcohol	No Cat
83	n-Butanal	123-72-8	Neutral organic	Aldehyde (electrophile)	Cat 2B
84	1-Decanol	112-30-1	Neutral organic	Alcohol	Cat 2A
85	1-Octanol	111-87-5	Neutral organic	Alcohol	Cat 2A
86	2-Ethyl-1-hexanol	104-76-7	Neutral organic	Alcohol	Cat 2A
87	2-Methyl-1-propanol	78-83-1	Neutral organic	Alcohol	Cat 2A
88	Cetylpyridinium bromide 1%	140-72-7	Cationic soap/surfactant	See No. 67	Cat 2A
89	Ethanol	64-17-5	Neutral organic	Alcohol	Cat 2A
90	Isopropanol	67-63-0	Neutral organic	Alcohol	Cat 2A
91	Methyl acetate	79-20-9	Neutral organic	Ester	Cat 2A
92	Methyl cyanoacetate	105-34-0	Neutral organic	Ester, nitrile	Cat 2A
93	n-Hexanol	111-27-3	Neutral organic	Alcohol	Cat 2A
94	Poly(ethylene glycol) butyl ether	9004-77-7	Neutral organic	Polymer, polyether alcohol	Cat 2A[b]
95	Acetic acid (10%)	64-19-7	Organic acid	Carboxylic acid	Cat 1
96	Anisole	100-66-3	Neutral organic	Aromatic ether, phenyl ether	SCNM[c]
97	Cetylpyridinium bromide 6%	140-72-7	Cationic soap/surfactant	See No. 67	Cat 1
98	Cetylpyridinium bromide 10%	140-72-7	Cationic soap/surfactant	See No. 67	Cat 1
99	Gamma-aminopropyl triethoxy silane	919-30-2	Organic base	Amine, silicium, silane	Cat 1
100	n-Butanol-10%	71-36-3	Neutral organic	Alcohol	Cat 1
101	n-Butanol	71-36-3	Neutral organic	Alcohol	Cat 1
102	Sodium hydroxide (10%)	1310-73-2	Inorganic base	Alkali	Cat 1
103	Stearyltrimethylammonium chloride (10%)	112-03-8	Cationic soap/surfactant	Alkylammonium salt	Cat 1
104	Surfonic N-102	9016-45-9	Neutral soap/surfactant	Polyether	Cat 1
105	Trichloroacetic acid (30%)	76-03-9	Organic acid	Carboxylic acid, halogenated	Cat 1

[a] 3,4,5,6-Tetrachloro-2-(1,4,5,8-tetrabromo-6-hydroxy-3-oxoxanthen-9-yl)-benzoic acid.
[b] Study criteria were not met for the *in vivo* Draize rabbit eye test, the UN GHS/EU CLP classifications corresponds with at least a Cat 2A.
[c] Study criteria were not met for the *in vivo* Draize rabbit eye test, the UN GHS/EU CLP classifications correspond with classified. The summary results of this study were published in the DRD (Barroso et al., 2015b). The study was terminated on day 7 with CO = 2, CC = 2 and CC = 2 in 1/1 animal.

UN GHS No Cat) correspond with chemical that result in a mean tissue viability >60%, Classified (C; UN GHS Cat 1/Cat 2) correspond with chemical that result in a mean tissue viability ≤60%.

2.7. Acceptance criteria

A run was considered qualified if the following criteria were met, mean OD of the NgC was ≥1.4 and ≤2.5 and mean % viability of the PC was ≤30. In addition, the difference of viability between the two replicate tissues of a single test chemical was ≤20 in the same run whatever the test item (for PC, NgC, test chemical and all adapted controls) (S.O.P. HCE-EITL version 1.0, 2015). As mentioned before, each chemical was tested three times in three independent runs in each laboratory. If a test did not meet the acceptance criteria in a run, a maximum of two additional independent tests was performed for each chemical in each laboratory.

2.8. Statistical data analyses

The frequency of non-qualified runs and non-qualified tests per laboratory was reported. The within-laboratory reproducibility (WLR), between-laboratory reproducibility (BLR), accuracy, sensitivity and specificity of the SkinEthic™ HCE EITL test method were calculated according to the rules described in the OECD Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Reconstructed Human Cornea-like Epithelium (RhCE) Test Methods for Eye Hazard (ENV/JM/MONO, 2015). Furthermore, the Validation Management Group (VMG) specified the following minimum values

Table 3
WLR for SkinEthic™ HCE EITL protocol. Mean cell viability (n = 2) for 60 chemicals and concordance of prediction within a laboratory.

No.	LO				CRL				VITO			
	Run 1	Run 2	Run 3	WLR concord.	Run 1	Run 2	Run 3	WLR concord.	Run 1	Run 2	Run 3	WLR concord.
1	57.9	59.2	56.2	Yes	71.2	66.5	57.5	No	57.5	53.2	55.9	Yes
2	99.4	106.7	101.9	Yes	121.1	102.1	110.0	Yes	105.4	91.3	126.9	Yes
3	62.5	80.0	58.0	No	73.7	72.0	54.2	No	98.0	74.4	87.8	Yes
4	73.3	72.5	71.6	Yes	84.3	86.6	82.0	Yes	90.4	76.1	78.2	Yes
5	34.4	59.6	70.5	No	85.2	76.9	70.9	Yes	52.0	59.1	75.3	No
6	1.0	1.1	1.1	Yes	2.7	2.0	1.2	Yes	1.3	1.5	1.3	Yes
7[a]	62.0 (62.4)	63.1 (63.7)	61.4 (61.8)	Yes	79.5	84.3	73.0	Yes	94.3	81.3	83.2	Yes
8	5.3	4.2	4.5	Yes	20.5	15.9	24.7	Yes	37.3	48.3	35.5	Yes
9[a]	1.9 (2.7)	2.1 (3.0)	2.1 (3.0)	Yes	1.4 (2.8)	1.5 (2.8)	1.9 (3.3)	Yes	2.0 (3.0)	3.7 (4.8)	3.6 (4.8)	Yes
10[a]	94.1	97.9	94.8	Yes	95.3 (95.4)	100.2 (100.3)	96.2 (96.3)	Yes	91.1	90.8	96.6	Yes
11	48.2	54.5	43.7	Yes	54.7	68.7	62.9	No	81.4	51.9	58.5	No
12[a]	2.0 (28.0)	0.0 (25.4)	0.3 (26.7)	Yes	0.0 (30.5)	0.0 (27.4)	0.0 (28.0)	Yes	4.3 (22.9)	3.6 (19.4)	9.3 (26.4)	Yes
13	88.1	89.1	99.1	Yes	86.6	93.0	94.8	Yes	91.7	95.6	64.3	Yes
14[a]	11.5 (11.5)	15.8 (15.8)	13.0 (13.0)	Yes	16.8 (16.8)	19.3 (19.3)	21.0 (21.0)	Yes	12.8	34.3	9.5	Yes
15	66.5	91.3	86.2	Yes	90.9	97.6	94.7	Yes	102.5	82.2	92.6	Yes
16	93.2	99.3	91.0	Yes	102.0	19.0	94.2	No	107.8	97.5	103.7	Yes
17	39.5	42.8	37.0	Yes	34.9	14.8	42.0	Yes	43.8	37.3	37.5	Yes
18	85.1	109.7	98.6	Yes	75.8	85.0	87.6	Yes	83.8	70.4	99.4	Yes
19	92.4	92.7	95.4	Yes	108.5	100.0	98.1	Yes	105.3	97.8	96.0	Yes
20[a]	98.7 (98.9)	98.9 (99.1)	95.7 (95.9)	Yes	97.5 (97.7)	100.4 (100.5)	99.9 (100.1)	Yes	81.9 (81.9)	105.3 (105.3)	90.4 (90.4)	Yes
21[a]	66.4 (66.8)	55.2 (55.5)	56.4 (56.8)	No	89.2 (89.5)	66.8 (67.0)	45.2 (45.5)	No	60.6 (61.1)	75.0 (95.4)	92.8 (93.2)	Yes
22	89.2	93.6	90.0	Yes	83.0	96.4	91.4	Yes	110.8	78.0	95.8	Yes
23	92.5	86.1	81.8	Yes	87.7	83.7	92.9	Yes	115.3	73.9	87.6	Yes
24	89.7	97.5	92.6	Yes	87.4	83.2	97.1	Yes	100.7	92.7	97.2	Yes
25	79.1	86.0	82.0	Yes	83.9	104.0	109.8	Yes	76.5	94.4	87.2	Yes
26	70.4	100.6	98.5	Yes	88.6	92.3	85.2	Yes	95.6	88.7	103.2	Yes
27[a]	92.3 (92.5)	93.4 (93.6)	87.6 (87.7)	Yes	88.3 (88.6)	84.7 (85.0)	77.7 (78.0)	Yes	103.0	89.3	89.9	Yes
28	85.3	89.2	90.4	Yes	81.5	90.0	90.8	Yes	75.1	80.8	80.2	Yes
29	1.0	1.2	1.3	Yes	83.6	1.3	1.1	No	1.3	2.0	1.4	Yes
30	2.1	2.9	3.0	Yes	2.8	2.1	2.0	Yes	15.1	4.2	9.2	Yes
31	1.4	1.5	1.1	Yes	3.2	1.1	1.0	Yes	2.4	9.7	1.5	Yes
32	1.7	1.6	1.9	Yes	2.1	5.3	2.1	Yes	3.4	3.5	3.6	Yes
33[a]	27.6 (27.9)	32.1 (32.3)	23.8 (24.0)	Yes	37.2 (37.2)	42.3 (42.3)	38.5 (38.5)	Yes	67.4	55.4	49.3	No
34	2.1	2.0	2.1	Yes	2.5	2.6	3.0	Yes	2.6	2.3	2.6	Yes
35[a]	0.9	1.0	0.8	Yes	0.9 (1.1)	3.5 (3.6)	4.9 (5.0)	Yes	1.0	1.3	2.7	Yes
36[a]	6.4 (6.5)	11.7 (11.8)	6.0 (6.1)	Yes	26.8 (27.1)	16.3 (16.5)	11.6 (11.9)	Yes	17.7	32.3	26.5	Yes
37[a]	2.6	4.2	5.1	Yes	3.8 (3.8)	2.8 (2.8)	4.5 (4.5)	Yes	6.6	9.4	8.2	Yes
38	3.1	3.9	3.2	Yes	3.4	4.9	4.7	Yes	9.2	8.0	12.3	Yes
39	1.5	2.1	2.2	Yes	0.8	0.9	1.0	Yes	0.8	0.9	1.0	Yes
40	0.8	1.0	0.9	Yes	0.9	1.2	1.4	Yes	3.6	1.2	1.2	Yes
41	1.6	1.6	1.6	Yes	1.6	1.5	1.4	Yes	1.4	1.6	1.4	Yes
42	3.6	1.7	3.5	Yes	5.2	2.7	6.5	Yes	7.4	10.4	8.5	Yes
43	1.2	1.4	1.5	Yes	1.9	2.6	1.4	Yes	1.9	1.9	1.5	Yes
44[a]	0.8 (0.8)	0.6 (0.6)	0.6 (0.6)	Yes	4.5 (4.6)	1.7 (1.9)	1.4 (1.6)	Yes	1.4	5.2	10.7	Yes
45[a]	14.1 (14.3)	14.7 (14.9)	24.3 (24.5)	Yes	43.9 (44.0)	27.4 (27.5)	44.4 (44.5)	Yes	25.1 (25.1)	65.1 (65.1)	65.7 (65.7)	No
46	2.2	1.6	1.7	Yes	6.7	1.9	1.6	Yes	2.9	1.8	1.6	Yes
47	1.8	2.1	1.7	Yes	1.6	1.9	1.6	Yes	1.8	1.7	1.7	Yes
48[a,b]	0.0 (3.9)	0.0 (3.6)	0.0 (6.0)	Yes	0.0 (12.2)	0.0 (25.3)	0.0 (18.6)	Yes	8.0 (9.6)	3.6 (8.7)	2.3 (5.9)	Yes
49	1.8	1.7	1.8	Yes	2.0	2.3	1.9	Yes	2.4	2.3	3.1	Yes
50	1.4	1.8	1.6	Yes	1.6	2.1	2.3	Yes	1.5	2.5	1.9	Yes
51	3.8	4.7	2.9	Yes	2.7	4.2	2.5	Yes	3.9	2.6	3.4	Yes
52[a,b]	2.1 (4.2)	1.4 (3.9)	2.1 (4.4)	Yes	1.7 (3.2)	1.3 (2.9)	0.0 (1.1)	Yes	0.4 (0.9)	0.7 (1.3)	0.6 (1.3)	Yes
53	1.9	2.4	1.7	Yes	3.7	1.8	2.1	Yes	2.3	2.6	2.8	Yes
54	2.3	2.0	2.0	Yes	3.0	2.8	2.7	Yes	2.4	2.9	4.1	Yes
55	1.7	2.4	1.6	Yes	1.7	2.2	2.1	Yes	1.8	4.2	1.8	Yes
56	0.8	0.6	0.5	Yes	1.7	4.2	1.9	Yes	1.1	2.8	0.7	Yes
57	2.2	2.7	2.7	Yes	4.0	1.8	2.5	Yes	2.4	2.7	2.4	Yes
58[a]	0.0 (28.0)	0.0 (24.4)	0.0 (24.2)	Yes	1.5 (28.4)	3.8 (32.6)	2.2 (31.9)	Yes	17.5 (32.4)	12.6 (26.9)	11.6 (26.4)	Yes
59[a]	28.0 (28.0)	15.5 (15.6)	22.1 (22.1)	Yes	51.7 (51.9)	62.0 (62.2)	27.9 (28.1)	No	53.6 (53.6)	52.4 (52.4)	59.5 (59.5)	Yes
60	0.9	0.9	0.8	Yes	0.8	1.0	0.8	Yes	0.6	0.8	0.8	Yes

for WLR (≥85%), BLR (≥80%), accuracy (≥75%), sensitivity (≥90%) and specificity (≥60%) for the prospective validation study of RhCE-based test methods conducted by EURL ECVAM and Cosmetics Europe (EC EURL ECVAM, 2014; Kaluzhny et al., 2015; Barroso et al., 2015a).

Calculation of basic statistical parameters such as difference and standard deviation was performed. The difference was used to analyse how consistent two values are such as for the % tissue viability between two tissue replicates and the standard deviation (a measure of the amount of dispersion (or variability) around the mean in a dataset) was used to evaluate the variation in % tissue viability values obtained by the three participating laboratories.

2.8.1. Within laboratory reproducibility (WLR)

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) of the qualified tests. WLR was reported with the Wilson's 95% confidence intervals (CI) for proportions. 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) and does not violate the constraints of probability estimates (i.e., estimates must be between 0 and 1).

2.8.2. Between laboratory reproducibility (BLR)

For each laboratory, the mean viability and standard deviation of the independent qualified tests 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 No Cat (NC). Between- laboratory reproducibility was reported with the Wilson 95% CI.

2.8.3. Predictive capacity

The predictive capacity of the assay was evaluated by comparing the prediction results, on the basis of the individual laboratory results using all qualified tests obtained for each chemical (as recommended by the OECD guidance document ENV/JM/MONO, 2015), with the existing proposed classification. Therefore 2 × 2 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.

In addition to the calculation of predictive capacity as recommended by the OECD guidance document, the uncertainty of the point estimates for accuracy, sensitivity and specificity was assessed with the bootstrap method. 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 60 chemicals, 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 chemicals. 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).

Notes to Table
LO: L'Oréal; CRL: Charles River Laboratories; concord.: concordance. Cells with a grey background correspond to irritant prediction (mean cell viability ≤60%).
[a]Chemical was identified as an MTT reducer by at least one laboratory. Values in brackets correspond to uncorrected viabilities, i.e. before subtraction of viability measured in killed epithelial tissue.
[b]Chemical corresponds with a colourant. Values in brackets correspond to uncorrected viabilities, i.e. before subtraction of non-specific colouring.

3. Results and discussions

3.1. Multicenter study: SkinEthic™ HCE EITL test method

Before starting the multicenter study, the effectiveness of the training phase was assessed in a transferability study. During this study, the laboratory technicians from CRL and VITO tested each 9 liquids under blinded conditions in at least three independent runs. These data were submitted to the test method developer (L'Oréal) for quality check. The results (data not shown) demonstrated effectiveness of the training.

Next, a total of 60 chemicals were tested in three laboratories. Overall, L'Oréal produced one unqualified result (chemical No. 27) over the 18 runs that were performed. Charles River Laboratories (CRL) performed 21 runs; one run was unqualified due to high deviation between the viability of the replicate NC tissues (difference 35%). Five

Table 4
BLR for SkinEthic™ HCE EITL protocol. Mean cell viability of 3 independent runs for 60 chemicals and agreement of prediction between the laboratories.

No.	*In vivo* UN GHS/EU CLP	LO		CRL		VITO		BLR concord.
		Cell viability (%)	*In vitro* class	Cell viability (%)	*In vitro* class	Cell viability (%)	*In vitro* class	
1	No Cat	57.8 ± 1.5		65.1 ± 7.0	NC	55.5 ± 2.2		No
2	No Cat	102.6 ± 3.7	NC	111.1 ± 9.6	NC	107.8 ± 17.9	NC	Yes
3	No Cat	66.8 ± 11.6	NC	66.6 ± 10.8	NC	86.7 ± 11.8	NC	Yes
4	No Cat	72.4 ± 0.8	NC	84.3 ± 2.3	NC	81.6 ± 7.7	NC	Yes
5	No Cat	54.8 ± 18.5		77.7 ± 7.2	NC	62.1 ± 12.0	NC	No
6	No Cat	1.1 ± 0.1	C	2.0 ± 0.7	C	1.4 ± 0.2	C	Yes
7	No Cat	62.2 ± 0.9	NC	78.9 ± 5.7	NC	86.2 ± 7.0	NC	Yes
8	No Cat	4.7 ± 0.6	C	20.4 ± 4.4	C	40.4 ± 6.9	C	Yes
9	No Cat	2.0 ± 0.2	C	1.6 ± 0.3	C	3.1 ± 0.9	C	Yes
10	No Cat	95.6 ± 2.0	NC	97.2 ± 2.6	NC	92.8 ± 3.3	NC	Yes
11	No Cat	48.8 ± 5.4	C	62.1 ± 7.1	NC	63.9 ± 15.5	NC	No
12	No Cat	0.8 ± 1.1	C	0.0 ± 0.0	C	5.7 ± 3.1	C	Yes
13	No Cat	92.1 ± 6.1	NC	91.4 ± 4.3	NC	83.9 ± 17.0	NC	Yes
14	No Cat	13.4 ± 2.2	C	19.0 ± 2.1	C	18.9 ± 13.5	C	Yes
15	No Cat	81.3 ± 13.1	NC	94.4 ± 3.4	NC	92.4 ± 10.2	NC	Yes
16	No Cat	94.5 ± 4.3	NC	71.7 ± 45.9	NC	103.0 ± 5.2	NC	Yes
17	No Cat	39.8 ± 2.9	C	30.6 ± 14.1	C	39.5 ± 3.7	C	Yes
18	No Cat	97.8 ± 12.3	NC	82.8 ± 6.2	NC	84.5 ± 14.5	NC	Yes
19	No Cat	93.5 ± 1.6	NC	102.2 ± 5.6	NC	99.7 ± 4.9	NC	Yes
20	No Cat	97.7 ± 1.8	NC	99.3 ± 1.6	NC	92.5 ± 11.9	NC	Yes
21	No Cat	59.3 ± 6.2	C	67.0 ± 22.0	NC	76.1 ± 16.1	NC	No
22	No Cat	90.9 ± 2.3	NC	90.3 ± 6.8	NC	94.9 ± 16.4	NC	Yes
23	No Cat	86.8 ± 5.4	NC	88.1 ± 4.6	NC	92.3 ± 21.1	NC	Yes
24	No Cat	93.3 ± 3.9	NC	89.2 ± 7.1	NC	96.9 ± 4.0	NC	Yes
25	No Cat	82.4 ± 3.5	NC	99.2 ± 13.6	NC	86.0 ± 9.0	NC	Yes
26	No Cat	89.9 ± 16.9	NC	88.7 ± 3.5	NC	95.8 ± 7.2	NC	Yes
27	No Cat	91.1 ± 3.1	NC	83.6 ± 5.4	NC	94.1 ± 7.8	NC	Yes
28	No Cat	88.3 ± 2.6	NC	87.4 ± 5.1	NC	78.7 ± 3.1	NC	Yes
29	Cat 2B	1.2 ± 0.1	C	28.7 ± 47.6	C	1.6 ± 0.4	C	Yes
30	Cat 2B	2.7 ± 0.5	C	2.3 ± 0.4	C	9.5 ± 5.4	C	Yes
31	Cat 2B	1.3 ± 0.2	C	1.8 ± 1.2	C	4.5 ± 4.5	C	Yes
32	Cat 2B	1.7 ± 0.2	C	3.2 ± 1.9	C	3.5 ± 0.1	C	Yes
33	Cat 2B	27.8 ± 4.2	C	39.3 ± 2.6	C	57.4 ± 9.2	C	Yes
34	Cat 2B	2.1 ± 0.1	C	2.7 ± 0.2	C	2.5 ± 0.2	C	Yes
35	Cat 2B	0.9 ± 0.1	C	3.1 ± 2.0	C	1.7 ± 0.9	C	Yes
36	Cat 2B	8.0 ± 3.2	C	18.2 ± 7.8	C	25.5 ± 7.4	C	Yes
37	Cat 2A	4.0 ± 1.3	C	3.7 ± 0.8	C	8.0 ± 1.4	C	Yes
38	Cat 2A	3.4 ± 0.4	C	4.3 ± 0.8	C	9.8 ± 2.2	C	Yes
39	Cat 2A	1.9 ± 0.4	C	0.9 ± 0.1	C	0.9 ± 0.1	C	Yes
40	Cat 2A	0.9 ± 0.1	C	1.1 ± 0.3	C	2.0 ± 1.4	C	Yes
41	Cat 2A	1.6 ± 0.0	C	1.5 ± 0.1	C	1.5 ± 0.1	C	Yes
42	Cat 2A	2.9 ± 1.0	C	4.8 ± 1.9	C	8.8 ± 1.5	C	Yes
43	Cat 2A	1.4 ± 0.2	C	2.0 ± 0.6	C	1.7 ± 0.2	C	Yes
44	Cat 2A	0.7 ± 0.1	C	2.5 ± 1.7	C	5.8 ± 4.7	C	Yes
45	Cat 1	17.7 ± 5.7	C	38.6 ± 9.7	C	52.0 ± 23.3	C	Yes
46	Cat 1	1.8 ± 0.3	C	3.4 ± 2.9	C	2.1 ± 0.7	C	Yes
47	Cat 1	1.9 ± 0.2	C	1.7 ± 0.2	C	1.7 ± 0.1	C	Yes
48	Cat 1	0.0 ± 0.0	C	0.0 ± 0.0	C	4.6 ± 3.0	C	Yes
49	Cat 1	1.8 ± 0.1	C	2.1 ± 0.2	C	2.6 ± 0.4	C	Yes
50	Cat 1	1.6 ± 0.2	C	2.0 ± 0.4	C	1.9 ± 0.5	C	Yes
51	Cat 1	3.8 ± 0.9	C	3.1 ± 0.9	C	3.3 ± 0.7	C	Yes
52	Cat 1	1.9 ± 0.4	C	1.0 ± 0.9	C	0.6 ± 0.2	C	Yes
53	Cat 1	2.0 ± 0.3	C	2.5 ± 1.0	C	2.6 ± 0.3	C	Yes
54	Cat 1	2.1 ± 0.1	C	2.8 ± 0.1	C	3.1 ± 0.9	C	Yes
55	Cat 1	1.9 ± 0.4	C	2.0 ± 0.3	C	2.6 ± 1.4	C	Yes
56	Cat 1	0.6 ± 0.1	C	2.6 ± 1.3	C	1.5 ± 1.1	C	Yes
57	Cat 1	2.5 ± 0.3	C	2.8 ± 1.1	C	2.5 ± 0.2	C	Yes
58	Cat 1	0.0 ± 0.0	C	2.5 ± 1.2	C	13.9 ± 3.1	C	Yes
59	Cat 1	21.9 ± 6.2	C	47.2 ± 17.5	C	55.2 ± 3.8	C	Yes
60	Cat 1	0.9 ± 0.1	C	0.8 ± 0.1	C	0.7 ± 0.1	C	Yes

LO: L'Oréal; CRL: Charles River Laboratories; concord.: concordance. Values correspond with mean ± SD of 3 independent runs.

Table 5
Predictive capacity for the set of 60 chemicals based on individual laboratory predictions: overall and for each laboratory.

In vivo UN GHS	Cumulative		L'Oréal		Charles River Laboratories		VITO	
	C	NC	C	NC	C	NC	C	NC
Classified (n)	283	5	96	0	94	2	93	3
No Category (n)	72	180	25	59	22	62	25	59
Total (n)	540		180		180		180	
Sensitivity (%)	98.3		100		97.9		96.9	
Specificity (%)	71.4		70.2		73.8		70.2	
Accuracy (%)	85.7		86.1		86.7		84.4	

additional results were unqualified, three based on the difference of viability of the replicate tissues which was >20% (chemicals No. 21 (two runs) and 45). Two other results were unqualified because of technical issues (chemicals No. 4 and 47). VITO performed 16 runs and obtained two unqualified results, the difference of viability of the replicate tissues was >20% for chemical Nos. 10 and 28. In total 54 valid independent runs 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 NC was within the acceptance limit (between 1.4 and 2.5) (data not shown).

Among the chemicals, two colourants that were also MTT reducers, were identified (No. 48 and 52, Table 3), requiring the use of adapted controls for the determination of non-specific colouration and MTT reduction. Eight chemicals (No. 9, 12, 20, 21, 45, 58, and 59) were identified as MTT reducers by all the three laboratories. Five chemicals (No. 14, 27, 33, 36, and 44) were identified as MTT reducers by L'Oréal and CRL. Four chemicals were identified as MTT reducer by one laboratory only (CRL No. 10, 35, and 37, and L'Oréal No. 7). Both uncorrected and corrected (final) viabilities were reported in the Table 3.

3.1.1. Within laboratory reproducibility

The reliability of the SkinEthic™ HCE EITL protocol was assessed in terms of concordance in predictions for three independent runs. The results for each laboratory are presented in Table 3. 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. Chemicals No. 3, 5, 11, and 21 resulted in discordant results in two laboratories. Chemical Nos. 1, 16, 29, 33, 45, and 59 resulted in a discordant prediction in one laboratory. The discordant predictions obtained for chemical No. 1 can be attributed to the viability which fluctuated around the cut-off value of 60% (between 53% and 66.5%), except for one run that resulted in a higher viability (71.2%). For chemical No. 16, 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 chemical No. 29, low viabilities were reported for all runs (<2%), except one run (viability: 83.6%). Chemical Nos. 33, 45, and 59 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.

3.1.2. Between laboratory reproducibility

In order to assess the transferability 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 Table 4. 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% (57/60 chemicals) for L'Oréal and VITO, and 98.3% (59/60 chemicals) for CRL and VITO. Chemicals No. 1, 5, 11, and 21 resulted in discordant predictions. The BLR of the SkinEthic™ HCE EITL test method was higher than the defined minimum value of 80% set by the VMG (Barroso et al., 2015a).

3.1.3. Predictive capacity

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 5). The calculations were based on the individual predictions derived from the qualified tests 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 70.2% (L'Oréal and VITO), and 73.8% (CRL). An accuracy of 86.1%, 86.7%, 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 10,000 resamplings of size 1 per chemical for the set of 60 chemicals. The distribution of the bootstrap samples is presented in Fig. 2. This resulted in an overall sensitivity of 98.2% (95% CI: 93.8% to 100%), a specificity of 73.4% (95% CI: 67.9% to 82.1%), and an accuracy of 86.6% (95% CI: 81.7% to 90.0%). In conclusion, the SkinEthic™ HCE EITL test method exceeds the defined values for sensitivity (≥90%), specificity (≥60%) and accuracy (≥75%) that were set by the VMG.

3.2. Additional data

The lead laboratory (L'Oréal) tested 45 additional chemicals (Table 2) in three independent runs. Twenty two chemicals did not require classification in vivo and 23 chemicals (including 11 Cat 1 and 12 Cat 2) were classified. Concordant prediction was obtained for 41 of the 45 chemicals, resulting in a WLR of 91.1% (Table 6). 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). This resulted in an accuracy of 84.8% with a 100% sensitivity and 68% specificity for L'Oréal only (Table 7). Overall, the sensitivity, specificity, and accuracy based on the individual predictions of the three laboratories were 98.6%, 70.1%, and 85.2% respectively. The bootstrap estimates for

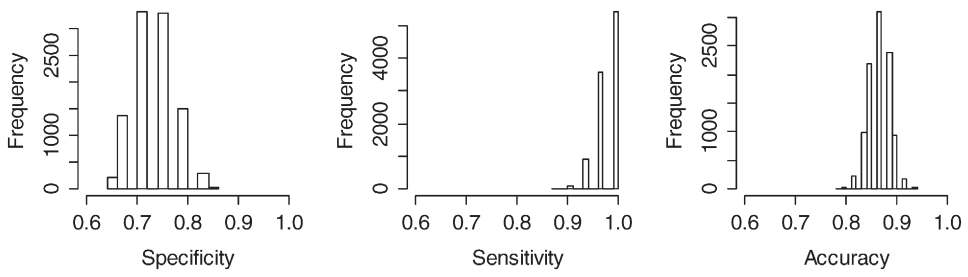

Fig. 2. Distribution of the bootstrap sample representing 10,000 resamplings of size 1 per chemical for the set of 60 chemicals (multicentre study).

Table 6
WLR for SkinEthic™ HCE EITL protocol. Mean cell viability (n = 2) for 45 additional chemicals and agreement of prediction within L'Oréal (LO).

No.	in vivo UN GHS/ EU CLP	LO			
		Run 1	Run 2	Run 3	Concordance
61	No Cat	78.4	78.0	71.3	Yes
62	No Cat	73.2	81.8	76.1	Yes
62[a]		(73.2)	(81.8)	(76.1)	
63	No Cat	36.5	57.5	30.1	Yes
63[a]		(36.5)	(57.6)	(30.2)	
64	No Cat	93.0	81.1	72.8	Yes
64[a]		(93.0)	(81.1)	(72.8)	
65	No Cat	104.6	98.6	103.2	Yes
65[a]		(104.6)	(98.7)	(103.2)	
66	No Cat	71.4	64.8	80.4	Yes
67	No Cat	90.3	88.8	99.0	Yes
68	No Cat	23.2	10.0	14.7	Yes
69	No Cat	60.5	56.9	77.4	No
70	No Cat	101.7	106.0	86.0	Yes
71	No Cat	3.9	1.5	1.3	Yes
72	No Cat	2.4	6.0	3.3	Yes
73	No Cat	69.9	64.7	61.5	Yes
74	No Cat	86.3	81.2	88.8	Yes
75	No Cat	2.1	2.3	1.7	Yes
76	No Cat	84.0	99.7	96.2	Yes
76[b]		(94.9)	(112.1)	(106.9)	
77	No Cat	75.7	89.4	72.5	Yes
78	No Cat	89.6	91.4	94.9	Yes
79	No Cat	73.6	65.4	59.2	No
79[a]		(73.7)	(65.5)	(59.3)	
80	No Cat	57.6	73.9	58.2	No
81	No Cat	88.7	75.0	59.0	No
82	No Cat	10.1	1.2	5.8	Yes
83	Cat 2B	10.5	0.0	7.7	Yes
83[a]		(13.3)	(0.0)	(10.7)	
84	Cat 2A	25.9	17.6	40.0	Yes
85	Cat 2A	1.2	1.5	1.4	Yes
86	Cat 2A	1.9	1.5	1.5	Yes
87	Cat 2A	1.5	1.8	1.9	Yes
88	Cat 2A	9.5	14.3	40.4	Yes
89	Cat 2A	1.6	2.2	1.7	Yes
90	Cat 2A	1.7	1.6	2.2	Yes
91	Cat 2A	3.0	5.6	1.6	Yes
92	Cat 2A	23.9	23.6	9.3	Yes
93	Cat 2A	1.5	1.5	1.3	Yes
93[a]		(1.6)	(1.5)	(1.3)	
94	Cat 2A[c]	3.9	10.7	0.9	Yes
94[a]		(3.9)	(10.7)	(0.9)	
95	Cat 1	2.2	2.8	2.3	Yes
96	SCNM[d]	17.9	3.7	6.5	Yes
97	Cat 1	1.8	1.9	5.7	Yes
98	Cat 1	1.4	2.4	2.2	Yes
99	Cat 1	2.1	1.6	1.6	Yes
99[a]		(2.1)	(1.6)	(1.6)	
100	Cat 1	22.3	32.0	34.8	Yes
101	Cat 1	1.7	1.6	1.3	Yes
102	Cat 1	0.1	0.3	0.1	Yes
103	Cat 1	3.9	5.2	6.9	Yes
104	Cat 1	21.7	29.8	12.2	Yes
105	Cat 1	2.8	3.2	2.8	Yes

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 Fig. 3. In conclusion, also for the extended set of 105 liquid chemicals, the SkinEthic™ HCE EITL test method met all the acceptance criteria set by the VMG.

3.3. Misclassified chemicals

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. A comprehensive in depth analysis of historical *in vivo* Draize eye data co-sponsored by Cosmetics Europe and the European Commission, provided more insight in the importance of understanding of individual *in vivo* tissue effects that drive classification of chemicals for serious eye damage/eye irritation (Adriaens et al., 2014). Recently, Cosmetics Europe has compiled a database of Draize eye test data, the so-called Draize eye test Reference Database (DRD) (Barroso et al., 2015b). The DRD contains a full description of all the ocular effects observed in vivo from 681 independent *in vivo* Draize eye studies on 634 individual chemicals. In order to evaluate the predictive capacity of the SkinEthic™ HCE EITL test method and its limitations, the misclassified chemicals were correlated with the *in vivo* drivers of classification as presented in the DRD published by Barroso et al. (2015b).

In total, out of 55 classified chemicals that were tested, four false negative results were obtained. Chemicals No. 29, 33 and 59 were predicted NC in 1 out of 9 runs and chemical No. 45 was predicted NC in 2 out of 9 runs (Table 3). Furthermore, the four false negatives represented four different functional groups (alcohol; ester, ketone; amine-silane; and polyether-acrylate). Two chemicals (No. 29: 2-methyl-1-pentanol and No. 33: ethyl-2-methylacetoacetate) correspond with an *in vivo* UN GHS Cat 2B classification. No. 29 resulted in an abnormal high viability (83.6%) in one run performed by Charles River in comparison with all other runs (viability ≤2%). A single false negative result was obtained for chemical No. 33 with a viability of 67.4% (VITO). Two other false negatives (No. 45: [3-(2-aminoethylamino)propyl]-trimethoxysilane and No. 59: tetraethylene glycol diacrylate) correspond with an *in vivo* Cat 1 classification. No. 45 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 chemical No. 45 ([3-(2-aminoethyl amino)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 chemical No. 45 were performed within a period of less than 30 days. Chemical No. 59 resulted in a NC prediction for one experiment performed by CRL, the viability of 62%

Notes to Table
Cells with a grey background correspond to irritant prediction (mean cell viability ≤60%).
[a]Chemical was identified as an MTT reducer by at least one laboratory. Values in brackets correspond to uncorrected viabilities, i.e. before subtraction of viability measured in killed epithelial tissue.
[b]Chemical corresponds with a colourant. Values in brackets correspond to uncorrected viabilities, i.e. before subtraction of non-specific colouring.
[c]Study criteria were not met for the *in vivo* Draize rabbit eye test, the UN GHS/EU CLP classifications corresponds with at least a Cat 2A
[d]Study criteria were not met for the *in vivo* Draize rabbit eye test, the UN GHS/EU CLP classifications correspond with classified. The summary results of this study were published in the DRD (Barroso et al., 2015b). The study was terminated on day 7 with CO = 2, CC = 2 and CC = 2 in 1/1 animal.

Table 7
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]	
	C	NC	C	NC	C	NC	C	NC
Classified (n)	**352**	5	**165**	0	**94**	2	**93**	3
No Category (n)	95	**223**	48	**102**	22	**62**	25	**59**
Total (n)	675		315		180		180	
Sensitivity (%)	98.6		100		97.9		96.9	
Specificity (%)	70.1		68.0		73.8		70.2	
Accuracy (%)	85.2		84.8		86.7		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.

was just above the classification cut-off of 60%. In terms of the *in vivo* drivers of classification, chemicals No. 29 and 33 were classified Cat 2B based on corneal opacity in the Draize eye test. Chemical No. 45, classified Cat 1 in the Draize eye test based on persistent conjunctival and corneal effects on day 21 in the majority of the animals. Chemical No. 59, 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 drivers 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 run. Twelve *in vivo* UN GHS No Cat chemicals (No. 6, 8, 9, 12, 14, 17, 63, 68, 71, 72, 75, and 82) were consistently predicted C (mean viability < 60%) in all runs. Among them were three esters, Ethyl acetate (No. 71), ethyl trimethyl acetate (No. 72), n-butyl acetate (No. 75) which resulted in a mean viability < 10% (Table 2). These esters also resulted in a false positive prediction in the BCOP (Balls et al., 1995) 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 (No. 1) was seven times predicted C. Dipropyl disulphide (No. 11) was six times predicted C and ethoxydiglycol (No. 5) was four times predicted C, p-methyl thiobenzaldehyde (No. 21) was three times predicted C, and 1-bromo-4-chlorobutane (No. 3) was predicted C twice. The false positive results for dimethyl sulfoxide (No. 69), triethanolamine orthovanadate 30 (No. 79), triethylene glycol monomethyl ether (No. 80), and triethylene glycol (No. 81) all resulted in a mean viability of ≥ 56.9%. The single false positive result obtained for iso-propyl myristate (No. 16, 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., 2015b). 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., 2015b). Seven out of those 10 chemicals were consistently predicted C with the SkinEthic™ HCE EITL method 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 irritant 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 70.1% (60% cut-off) to 76.1% with a slight decrease in sensitivity from 98.6% to 97.2%. However, one Cat 1 chemical (No. 59) 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, 2015b). However, with the 60% cut-off value, the SkinEthic™ HCE EITL method resulted in a similar sensitivity (98.6%), specificity (70.1%) and accuracy (85.2%) 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). Considering the data obtained in the validation study, 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) (EC EURL ECVAM (2014). The SkinEthic™ HCE EITL test method is currently in the work plan 2015 programme of the OECD for identifying No Cat chemicals.

4. Conclusions

The present work assessed the reliability (WLR and BLR) and relevance (predictive capacity) of the SkinEthic™ HCE EITL 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). Following successful transfer to two naïve laboratories, good WLR was achieved of at least 88.3% (53/60). The WLR for the extended data set (L'Oréal only) was 92.4% (97/105). Furthermore, the overall concordance between the laboratories was 93.3% (56/60). The accuracy of the SkinEthic™ HCE EITL method for the extended dataset, based on bootstrap, was 84.4% (95% CI: 81.9% to 87.6%) with a sensitivity of 99.0% (95% CI: 96.4% to 100%) and specificity of 68.5% (95% CI: 64.0% to 74.0%).

Conflict of interest statement

The authors declare there are no conflicts of interest.

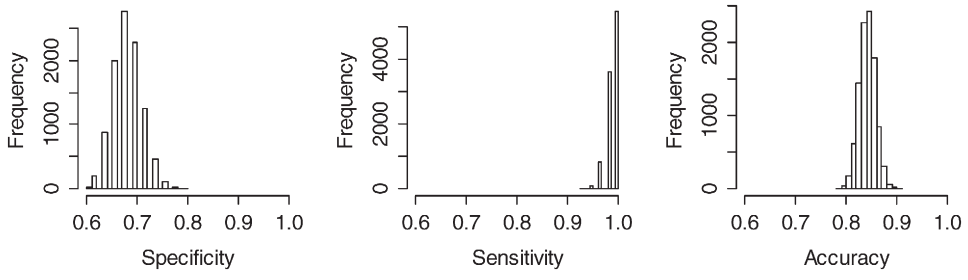

Fig. 3. Distribution of the bootstrap sample representing 10,000 resamplings of size 1 per chemical for the extended data set of 105 chemicals.

Funding

No external funding was received for this study. The organisation of the study and evaluation of the data was funded by L'Oréal Research & Innovation.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

The authors gratefully acknowledge Tadeja Kerin and Iwona Wilinska (Charles River Laboratories, Edinburgh, United Kingdom), An Jacobs and Karen Hollanders (VITO NV, Mol, Belgium) for their technical assistance and fruitful discussions.

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. http://dx.doi.org/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, 1476–1488. http://dx.doi.org/10.1016/j.tiv.2013.02.009.

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

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., 2015b. 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 August, 2015.

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

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., 2015a. 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. http://dx.doi.org/10.1016/j.toxlet.2015.08.507 (Available from: Accessed on 05/10/2015).

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, C., Alépée, N., Tinois-Tessonneaud, E., De Brugerolle de Fraissinette, A., Meuniera, 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, 523–537. http://dx.doi.org/10.1016/j.tiv.2009.11.010.

Cotovio, J., Grandidier, M.-H., Lelièvre, D., Bremond, C., Flamand, N., Loisel-Joubert, S., Van Der Lee, A., Capallere, C., Meunier, J.-R., Leclaire, J., 2007. The use of the reconstructed human corneal model (HCE) to assess in vitro eye irritancy of chemicals. ALTEX 14 (special issue), 343–350.

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).

ENV/JM/MONO, 2015, 23, 2015. OECD Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Reconstructed Human Cornea-like Epithelium (RhCE) Test Methods for Eye Hazard, OECD Environment, Health and Safety Publications. Series on Testing & Assessment 216, p. 21.

ESAC, 2009. Statement on the scientific validity of cytotoxicity/cell-function based *in vitro* assay for eye irritation testing. Issued by the ECVAM Scientific Advisory Committee (ESAC) at its 31st Meeting, 10th July 2009. European Commission Joint Research Centre. Institute for Health and Consumer Protection. European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) (Available from: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-eye-irritation/ESAC31_CBA_eye-irritation_20091005.pdf. (accessed 31.07.2015)).

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), 261–266.

Hartung, T., Bruner, L., Curren, R., Eskes, C., Goldberg, A., McNamee, P., Scott, L., Zuang, V., 2010. First alternative method validated by a retrospective weight-of-evidence approach to replace the Draize eye test for the identification of non-irritant substances for a defined applicability domain. ALTEX 27, 43–51.

ICCVAM, 2010. ICCVAM Test Method Evaluation Report: Current Validation Status of *In Vitro* Test Methods Proposed for Identifying Eye Injury Hazard Potential of Chemicals and Products. NIH Publication No. 10-7553. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA (Available from: http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/ocular/in-vitro-test-methods/tmer/index.html. (accessed 31.07.2015)).

Kaluzhny, Y., Kandarova, H., Handa, Y., DeLuca, J., Truong, T., Hunter, A., Kearney, P., d'Argembeau-Thornton, L., Klausner, M., 2015. The EpiOcular Eye Irritation Test (EIT) for hazard identification and labelling of eye irritating chemicals: protocol optimisation for solid materials and the results after extended shipment. ATLA 43, 101–127.

Kaluzhny, Y., Kandarova, H., Handa, P., Kunilus, J., d'Argembeau-Thornton, L., Klausner, M., 2011. DEvelopment of the EpiOcular™ 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.

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

OECD, 2012. Test No. 460: Fluorescein Leakage test method for identifying ocular corrosives and severe irritants. OECD Guideline for the Testing of Chemicals, Section 4. OECD publishing, Paris http://dx.doi.org/10.1787/9789264185401-en (16 pp., OECD publishing).

OECD, 2013a. Test No. 437: Bovine Corneal Opacity and Permeability test method for identifying i) chemicals inducing serious eye damage and ii) chemicals not requiring classification for eye irritation or serious eye damage. OECD Guideline for the Testing of Chemicals, Section 4. OECD publishing, Paris http://dx.doi.org/10.1787/9789264203846-en (27 pp. OECD publishing).

OECD, 2013b. Test No. 438: Isolated Chicken Eye test method for identifying i) chemicals inducing serious eye damage and ii) chemicals not requiring classification for eye irritation or serious eye damage. OECD Guideline for the Testing of Chemicals, Section 4. OECD publishing, Paris http://dx.doi.org/10.1787/9789264203860-en (20 pp. OECD publishing).

OECD, 2015a. Test No. 491: The Short Time Exposure *in vitro* test method for identifying i) chemicals inducing serious eye damage and ii) chemicals not requiring classification for eye irritation or serious eye damage. OECD Guideline for the Testing of Chemicals, Section 4. OECD publishing http://dx.doi.org/10.1787/9789264242432-en (14 pp.).

OECD, 2015b. Test No. 492: Reconstructed Human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage. OECD Guideline for the Testing of Chemicals, Section 4. OECD publishing http://dx.doi.org/10.1787/9789264242548-en (27 pp.).

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-Fairly, M., 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 http://dx.doi.org/10.1016/j.tiv.2012.11.007.

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/).

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., Varlet, B.L., Meloni, M., McNamee, P., Osborne, R., Pape, W., Pfannenbecker, U., Prinsen, M., Seaman, C., Spielmann, H., Stokes, W., Trouba, K., Berghe, C.V., Van Goethem, F., Vassallo, M., Vinardell, P., Zuang, V., 2010. A proposed eye irritation testing strategy to reduce and replace *in vivo* studies using Bottom-Up and Top-Down approaches. Toxicol. In Vitro 24, 1–9. http://dx.doi.org/10.1016/j.tiv.2009.05.019.

UN, 2013. United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30/Rev.5, Part 3 Health Hazards—Chapter 3.3 Serious Eye Damage/Eye Irritation. United Nations, New York, USA, and Geneva, Switzerland, pp. 135–147 (Available from: http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf. (accessed 31.07.2015)).

Van Goethem, F., Adriaens, E., Alépée, N., Straube, F., De Wever, B., Cappadoro, M., 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. http://dx.doi.org/10.1016/j.tiv.2005.05.002.

Wilson, S.L., Ahearne, M., Hopkinson, A., 2015. An overview of current techniques for ocular toxicity testing. Toxicol. In Vitro 327, 32–46. http://dx.doi.org/10.1016/j.tox.2014.11.003.