Organisation for Economic Co-operation and Development

Unclassified English - Or. English
19 September 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Peer Review Report- SkinEthicTM Human Corneal Epithelium Time-to-Toxicity Test for Eye irritation

Series on Testing and Assessment
No. 355

JT03502646

OECD Environment, Health and Safety Publications
Series on Testing & Assessment
No. 355

Peer Review Report- SkinEthic™ Human Corneal Epithelium Time-to-Toxicity Test for Eye irritation

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents;** and **Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment: link

For this and many other Environment,

Health and Safety publications, consult the OECD's

World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,

Environment, Health and Safety Division

2 rue André-Pascal

75775 Paris Cedex 16

France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

Foreword

This document encloses the Peer review report of the validation study for SkinEthic™ Human Corneal Epithelium (HCE) Time-to-Toxicity Test (TTT) on liquids (TTL) and on solids (TTS), which is contained in OECD Test Guideline 492B on Reconstructed Human Cornea-like Epithelium (RHCE) Test Method for Eye Hazard Identification. The review was conducted by an independent international Peer review panel from May to December 2020, and the report was made available to the OECD Expert Group for Eye Irritation for review in January 2021. The validation report for SkinEthic™ HCE TTT is available in a peer-reviewed publication and is referenced in the Peer review report. The WNT approved the new Test Guideline 492B and endorsed the Peer review report at its 34th meeting in April 2022, on the basis of a project led by France. This report is published under the responsibility of the Chemicals and Biotechnology Committee.

Peer Review Panel Report

on the scientific validity of the

SkinEthic™ Human Corneal Epithelium (HCE) Time-to-Toxicity Test (TTT) on liquids (TTL) and on solids (TTS)

January 12, 2021

Table of contents

1. Summary

The SkinEthic™ Human Corneal Epithelium (HCE) Time-to-Toxicity Test (TTT) on liquids (TTL) and on solids (TTS) has been developed by L'Oréal to determine the eye irritation and serious eye damage potential of chemicals according to the three categories of the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS), i.e. Cat. 1 (serious eye damage), Cat. 2 (eye irritation) and No Cat. (not classified for serious eye damage/eye irritation). A validation study according to the modular approach to validation (Hartung et al., 2004) was conducted to demonstrate the transferability, within- and between-laboratory reproducibility and predictive capacity of the SkinEthic HCE™ TTT. The study was sponsored by the test method developer L'Oréal R&I and coordinated by L'Oréal R&I and Adriaens Consulting BVBA.

The SkinEthic™ HCE TTT was peer-reviewed by an independent international peer-review panel (PRP) from May to December 2020 in order to assess its scientific validity for the intended purpose. The PRP agreed on 15 evaluation criteria to guide the review. Despite some shortcomings identified in the study management and conduct, the PRP concludes that the validation study and its conclusions are of sufficient quality and completeness to allow an assessment of the scientific validity of the SkinEthic™ HCE TTT. The PRP considers the SkinEthic™ HCE TTT to be highly reproducible with both liquids (TTL) and solids (TTS). For the SkinEthic™ HCE TTL, the within-laboratory reproducibility (WLR) ranged from 85% to 95%, and the between-laboratory reproducibility (BLR) was of 90% for the 20 coded liquid chemicals tested. For the SkinEthic™ HCE TTS, the WLR and BLR were of 100% for the 20 coded solid chemicals tested.

The PRP also considers that the predictive capacity of the SkinEthic™ HCE TTT for the purpose of discriminating the three UN GHS categories for serious eye damage/eye irritation is sufficient and adequate. The SkinEthic™ HCE TTT correctly identified 79% UN GHS Cat. 1 (n = 50), 69% UN GHS Cat. 2 (n = 44) and 75% UN GHS No Cat. (n = 57) of the 151 chemicals tested for assessing predictive capacity. None of the UN GHS Cat. 1 were underpredicted as UN GHS No Cat., and only 2% of UN GHS No Cat. were overpredicted as UN GHS Cat. 1. Furthermore, although the SkinEthic™ HCE TTT was not developed to account for all mechanisms associated with serious eye damage/eye irritation, the observed performances demonstrate that the SkinEthic™ HCE TTT method is able to predict serious eye damage/eye irritation with sufficient accuracy independent of the types of ocular effects observed *in vivo* (i.e., corneal, iridal and conjunctival injuries; including severity and persistence of effects).

On the basis of all available data, the PRP concludes that the relevance and reliability of the SkinEthic™ HCE TTT have been satisfactorily demonstrated for the intended purpose of discriminating the three UN GHS categories for serious eye damage/eye irritation, i.e. UN GHS Cat. 1, Cat. 2 and No Cat. Therefore, the PRP considers the test method to be scientifically valid for regulatory use as a full replacement to the *in vivo* Draize acute eye irritation test for classification of chemicals according to UN GHS. The SkinEthic™ HCE TTT is, however, not intended to differentiate between eye irritants (optional Cat. 2A) and mild eye irritants (optional Cat. 2B), as defined by UN GHS.

2. Introduction

The SkinEthic™ Human Corneal Epithelium (HCE) Time-to-Toxicity Test (TTT) on liquids (TTL) and on solids (TTS) has been developed by L'Oréal to determine the eye irritation and serious eye damage potential of chemicals according to the three categories of the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS), i.e. Cat. 1 (serious eye damage), Cat. 2 (eye irritation) and No Cat. (not classified for serious eye damage/eye irritation). A validation study according to the modular approach to validation (Hartung et al., 2004) was conducted to demonstrate the transferability, within- and between laboratory reproducibility and predictive capacity of the SkinEthic HCE™ TTT. The study was sponsored by the test method developer L'Oréal R&I and coordinated by L'Oréal R&I and Adriaens Consulting BVBA. The study was initiated in the first half of 2019 by training technicians of three laboratories at the test developer site. After successful transfer to the laboratories was demonstrated, a multi-laboratory trial was conducted to assess the reproducibility of the SkinEthic™ HCE TTT method by testing 20 coded liquid and 20 coded solid chemicals, each in three independent experiments per laboratory, in the respective protocols until January 2020. The predictive capacity of the SkinEthic™ HCE TTT method was assessed with retrospective data on 126 chemicals tested non-coded by the test method developer plus the prospective data generated in the multi-laboratory trial for a total of 151 unique chemicals (126 + 25). The datasets were analysed by Adriaens Consulting BVBA.

L'Oréal contracted seh consulting + services (Sebastian Hoffmann) to organise and manage an independent peer-review of the SkinEthic™ HCE TTT. An international peer-review panel (PRP) was convened. The PRP, who provided Declarations of Interests before the first panel meeting, included the following experts:

- João Barroso, European Commission, Joint Research Centre, Italy
- Chantra Eskes, SeCAM, Switzerland
- Dori Germolec, NICEATM/NTP/NIEHS, USA
- Bae-Hwan Kim, Keimyung University, Korea
- Hajime Kojima, JaCVAM, NIHS, Japan

The PRP members acted on their own personal capacity and not in representation of the institutions to which they are affiliated. The PRP met twelve times virtually in the period from May to December 2020. A representative of the test developer (Nathalie Alépée) attended the first two meetings to present the SkinEthic™ HCE TTT and to provide answer to PRP questions. Minutes are available for the first five meetings. In the remaining seven meetings, the PRP agreed on 15 evaluation criteria and reviewed the test method according to these criteria. The progress made in the meetings was reflected in various draft versions of this peer review report.

3. Peer-review panel evaluation

3.1 Study objective and test method purpose (Evaluation criterion 1)

Were the study objective and the purpose of the test method adequately described?

The PRP considers that the objective is clearly stated in the validation report. It understands that the objective of the validation study of the SkinEthic™ HCE TTT was to assess the transferability, the reliability in terms of within- and between-laboratory reproducibility and the relevance in terms of predictive capacity for the purpose of discriminating the three UN GHS categories for serious eye damage/eye irritation (Cat. 1, Cat. 2, No Cat.), in order to facilitate international acceptance for UN GHS hazard classifications of chemicals.

3.2 Need and benefits in comparison to existing test methods (Evaluation criterion 2)

Were the need and benefits in comparison to existing test methods appropriately addressed in terms of (regulatory) purpose, IP rights, geographical availability, animal welfare, costs, analysis time, sample amount and other relevant aspects?

The PRP considers that the main advantage of the SkinEthic™ HCE TTT as compared to currently accepted test methods is the fact that it is the first individual test method formally considered by the OECD for discriminating on its own the three UN GHS ocular hazard categories, i.e., Cat. 1, Cat. 2 and No Cat. Other approaches proposed for this purpose include Defined Approaches[1] that combine different information sources and are currently being considered for adoption by the OECD (Alépée et al., 2019a, b). In addition, the EpiOcular Reconstructed human Cornea-like Epithelium (RhCE) ET_{50} test method is accepted by the US EPA to be used in combination with the Bovine Corneal Opacity and Permeability (BCOP) assay and the Cytosensor Microphysiometer (CM) assay for classification of eye irritation potential of EPA-regulated pesticide products (EPA, 2015). Finally, the PRP observes that other test methods for discriminating multiple categories of ocular hazard classification have been described in the scientific literature and in OECD GD 263 (2019). These, however, have not been formally validated.

The PRP observes that the name SkinEthic™ is protected by a trademark, but SkinEthic™ HCE tissues are commercially available. The tissues can be shipped globally from the main production site in France without any loss of functionality, as demonstrated for the SkinEthic™ HCE Eye Irritation Test (EIT). In addition, the tissues are also produced in Brazil and China, so that they are readily available in those regions with shorter shipment routes.

Regarding animal welfare, the PRP agreed that the SkinEthic™ HCE TTT would, when implemented into regulations, reduce the need for animal testing, as currently some chemicals might need to be tested *in vivo* to identify eye irritating effects, especially those falling into UN GHS Cat. 2. The PRP was informed by the test method developer that none of the materials used to conduct the SkinEthic™ HCE TTT are of animal origin.

[1] A Defined Approach consists of a fixed data interpretation procedure (DIP) applied to data generated with a defined set of information sources to derive a result that can either be used on its own, or together with other information sources within an overall weight of evidence approach, to satisfy a specific regulatory need (OECD, 2016).

The PRP did not have all the necessarily information to appropriately compare the costs of the SkinEthic™ HCE TTT to other test methods and approaches addressing the same purpose. However, the PRP agrees that the sample amount required for testing is in the same range as required for the Draize acute eye irritation Test (OECD TG 405), the only OECD-accepted individual test method for the same purpose.

3.3 Comparison of the test method with the essential test method components as described in the Performance Standards of OECD GD 216 (Evaluation criterion 3)

The SkinEthic™ HCE TTT complies with the structural and functional essential test method components of OECD GD 216 (OECD, 2017). The SkinEthic™ HCE tissues closely mimic the histological, morphological, biochemical and physiological properties of the human corneal epithelium. It also complies with the functional essential test method components of OECD GD 216, such as viability, barrier function, morphology, reproducibility (of controls) and quality control.

In addition, the SkinEthic™ HCE TTT complies with most of the procedural conditions of OECD GD 216, with the following exceptions:

- the SkinEthic™ HCE TTT uses more than one exposure time (i.e., three for TTL and two for TTS);
- liquid chemicals are tested in the TTL protocol at 20% (w/v) for the exposure times of 16 and 120 min (neat for 5 min exposure), while chemicals are tested neat in EIT;
- a different 'Interpretation of results and prediction model' is used.

These differences are due to the fact that the SkinEthic™ HCE TTT is proposed to identify the three UN GHS ocular hazard categories, and not only UN GHS No Cat., as currently covered by OECD GD 216. The PRP notes that SkinEthic™ HCE TTT cannot be assessed based on the Performance Standards of OECD GD 216 as these were developed for test methods identifying only UN GHS No Cat. Indeed, the SkinEthic™ HCE TTT underwent a full validation study rather than a performance-based validation study based on the reference chemicals defined in OECD GD 216.

However, on the basis of the compliance of the test method with the structural and functional essential test method components as well as of the compliance with most of the procedural conditions stated in GD 216, the PRP considers that the biological and mechanistic relevance of the test method is similar to those of the test methods covered by OECD TG 492.

3.4 Biological and mechanistic relevance (Evaluation criterion 4)

Are the toxicological mechanisms and the relationship between the test method endpoint(s) with the biological effect as well as the toxicity of interest adequately addressed?

On the basis of the nearly complete compliance of the test method with the essential test method components as described in criterion 3, the PRP considers that the biological and mechanistic relevance of the test method has been properly established. The biological relevance of the SkinEthic™ HCE TTT can be inferred from the fact that the same tissues (SkinEthic™ HCE) have already been evaluated and included in OECD TG 492. Moreover, the test method assesses cytotoxicity, which is considered as a biologically relevant endpoint in OECD TG 492 (paragraph 9). The PRP believes that the

available documentation provides sufficient evidence and reasoned arguments in favour of this RhCE test method and its measured endpoint being as biologically and mechanistically relevant as the currently validated *in vitro* methods described in OECD TG 492.

3.5 Test method protocol (Evaluation criterion 5)

Is a detailed protocol for the test method available?

The PRP considers that detailed protocols are available for both the SkinEthic™ HCE TTL and the SkinEthic™ HCE TTS. However, there are no instructions in either protocol or in any other document provided to the PRP on how to decide which protocol to use. It is important to define a clear procedure for the selection of the correct protocol, especially for example, for pastes and for chemicals having a low melting point and/or high viscosity.

Furthermore, the PRP notes that no instructions are provided for the testing of poorly water-soluble liquids at 20% (w/v) exposed for 16 or 120 minutes in the TTL protocol. Liquids not forming a stable suspension in water for the duration of the exposure may lead to increased variability and to overprediction if denser than water or underpredictions if less dense than water. The PRP did not conduct a detailed analysis to confirm this hypothesis. However, the PRP notes that the validation data, which included 16 liquids with a water solubility of less than 0.1 g/L (corresponding to four UN GHS Cat. 1, two UN GHS Cat. 2 and ten UN GHS No Cat.), do not indicate that this was a major issue regarding mispredictions, as only one UN GHS Cat.1 chemical was underpredicted as UN GHS Cat. 2 and one UN GHS No Cat. chemical was overpredicted as UN GHS Cat. 2.

The PRP did not see a value in including Positive Control 1 (neat Lactic acid in both TTL and TTS) in the SkinEthic™ HCE TTT protocols because it kills the tissues (viabilities < 7% obtained in the multi-laboratory trial). Therefore, the PRP recommend that only Positive Control 2 (neat Methyl acetate in TTL and 1% (w/v) Lactic acid in TTS) is retained in the protocols. The test developer followed this recommendation and removed the Positive Control 1 from both protocols in revised versions, which were provided to the PRP during the peer review. Moreover, the PRP acknowledges that the use of a liquid positive control (Lactic acid) in a protocol for solids is common practice (e.g. in the BCOP) and acceptable. However, the PRP suggests considering the use of a solid positive control in the future.

Finally, the PRP suggests simplifying the prediction model by specifying only the criteria used for predicting UN GHS Cat. 1 and UN GHS No Cat. chemicals, and stating that all other combinations correspond to UN GHS Cat. 2 without specifying them.

3.6 Appropriateness of the validation study management and conduct (Evaluation criterion 6)

The validation report states that the study was conducted according to the principles and criteria documented in OECD GD 34 (OECD, 2005) and the modular approach to validation (Hartung et al., 2004). The PRP agrees that the study followed the modular approach. However, the PRP notes that the independence of the study management and of the selection of chemicals as recommended in OECD GD 34 was unclear.

The test method developer who owns the company producing and commercialising the tissues was the sponsor of the study. In addition, this same company was involved in the study communication, study coordination and test chemical selection, acquisition and distribution. These are tasks that, in order to ensure full independence of the study, should ideally have been conducted by a third party with no vested interests and/or by individuals having no conflicts of interest with the test method or the outcome of the validation study. In the case of the SkinEthic™ HCE TTT validation study, the PRP notes that, while a different department within the sponsoring organisation (chemistry library department) was responsible for purchasing and shipment of chemicals and for providing safety data sheets to the participating laboratories, the selection of chemicals was conducted by the department that developed the test method. Chemical coding and data analysis were conducted by a contracted third party (Adriaens Consulting BVBA), which had to provide the codes to the sponsor (i.e. to the chemistry library department), in order for the latter to be able to distribute the chemicals to the participating laboratories. The PRP was informed that the third party also contributed to the chemicals selection, coordinated the study and was the single point of contact for all parties during the multi-laboratory trial. However, the distribution of tasks between the sponsor and the third party, in terms of coordination and chemicals selection, was not adequately described in the validation report and is not clear to the PRP. Therefore, the PRP cannot determine if full independence was ensured according to the provisions of OECD GD 34.

In addition, the PRP could not confirm from the information provided whether the prediction model of the SkinEthic™ HCE TTT was defined before the initiation of the experimental phase of the multi-laboratory trial. The PRP notes that the transfer reports that describe the prediction models were signed during or after the experimental phase of the validation and that, within the peer-review package, there is no other document preceding the transfer reports that describes the prediction model.

The PRP agrees that the data analysis was conducted by an independent statistician, and the results are presented in the validation report. However, a detailed statistical report was not provided to the PRP.

Finally, the acceptance criteria for the study against which to evaluate the validity status of the test method in terms of reproducibility and predictive capacity were not described in the original validation report nor in any other document provided to the PRP. Upon request, the test method developer provided these in a revised version of the validation report, with a statement that they had been discussed with the independent statistician and were presented during the training phase to the participating laboratories. However, they were not described in the training report, nor in the study project plan and it remains thus unclear when these criteria were defined. Nevertheless, the PRP considers that the rationale justifying their definition as discussed under criterion 12 is more important than the timepoint when these criteria were defined.

Despite the shortcomings identified above, the PRP concludes that the validation study and its conclusions are of sufficient quality and completeness to allow an assessment of the scientific validity of the SkinEthic™ HCE TTT.

3.7 Adequacy of chemicals selection for the validation study objective (Evaluation criterion 7)

A prospective multi-laboratory trial with 40 coded chemicals was conducted to assess the reproducibility of the SkinEthic™ HCE TTL (20 chemicals) and TTS (20 chemicals). The PRP considers that the criteria for selecting these chemicals were appropriate and well defined. In particular, chemicals were selected from published reference databases following the recommendations given in those publications (Barroso et al. 2017). Of these 40 chemicals, 15 had been previously tested by the test method developer to inform within-laboratory reproducibility and predictive capacity and 25 chemicals had not been tested before. It remains unclear to the PRP why so many chemicals previously tested by the test method developer (15 out of 40) were selected for the multi-laboratory trial. Moreover, the PRP notes that 14 of these 15 chemicals had been correctly classified by the test method developer. Six liquids were tested in one run, including a Cat. 1 chemical classified as Cat. 2 and two Cat. 2 chemicals with borderline viability results, but correctly classified. Six solids were tested in two concordant runs and three in one run with viabilities far off cut-off values. In order to evaluate the impact of the inclusion of 15 previously tested chemicals in the assessment of test method performance, the PRP conducted an independent analysis for the 25 new chemicals and concluded that the performance of the method was better for these 25 additional chemicals than for the full chemical set (see evaluation criteria 11 and 12).

The predictive capacity was assessed with retrospective data on 126 chemicals tested non-coded by the test method developer plus the prospective data generated in the multi-laboratory trial for a total of 151 unique chemicals (126 + 25). Out of these 151 chemicals, 50 were UN GHS Cat. 1, 44 UN GHS Cat. 2 and 57 UN GHS No Cat., which the PRP considers to be an appropriate distribution. In addition, these chemicals included different *in vivo* drivers of classification and organic functional groups as described by Barroso et al. (2017).

The PRP considers the number of selected chemicals adequate for assessing the within- and between-laboratory reproducibility as well as the predictive capacity of the SkinEthic™ HCE TTT. The PRP notes, however, that justifications for the number of selected chemicals, especially those used in the multi-laboratory trial, were not provided.

In addition, the PRP considers that there was a potential lack of independence in the selection of chemicals, due to the involvement of the test developer, as described in evaluation criterion 6. However, due to the adequacy of the chemicals distribution and representation, the PRP concludes that the chemicals selection was of sufficient quality and completeness to allow an assessment of the scientific validity of the SkinEthic™ HCE TTT.

3.8 Quality of the reference data used for the evaluation of relevance (Evaluation criterion 8)

The test method developer tried to minimise uncertainty associated with the *in vivo* reference data by selecting chemicals from the Draize acute eye irritation test Reference Database (DRD) and following the recommendations for the selection of chemicals for validation studies by Barroso et al. (2017). Chemicals with highly uncertain *in vivo* reference data were not included in the study. However, due

to the known reproducibility issues associated with the Draize acute eye irritation test (Prinsen, 2006; Luechtefeld et al., 2016; Barroso et al., 2017) a certain level of uncertainty remains, which is reflected in the PRP acceptance criteria for predictive capacity (see evaluation criterion 12). As observed with other test methods, the SkinEthic™ HCE TTT showed higher predictive capacity with liquids than with solids. This is an indicator for higher uncertainty associated with Draize acute eye irritation test reference data for solid chemicals due to more variable exposure conditions *in vivo* (e.g., uncontrolled exposure time, mechanical abrasion, secondary effects), especially for studies performed before the revision of OECD TG 405 in 2002, which introduced a washing step after one hour of exposure for solids.

3.9 Training of naïve laboratories and transferability (Evaluation criterion 9)

Three laboratories participated in the validation study of the SkinEthic™ HCE TTT with EPISKIN SA (Lyon, France) acting as lead laboratory, and EUROSAFE (Saint Grégoire, France), and VITROSCREEN Srl (Milano, Italia) acting as naïve laboratories. All laboratories in the validation study received 3-4 days training by the test method developer for each of the two protocols (TTL in March and April 2019; TTS in June 2019). The training covered all aspects of the protocols, in particular pre-checking for colour interference and direct MTT reduction, and included testing of three chemicals as well as the negative and positive controls. Issues identified during training were resolved and all laboratories achieved results indicating successful training. Training reports are available for each laboratory and protocol.

The transferability phase addressed demonstrating proficiency with the protocols and completion of the EXCEL spreadsheet for data analysis. Three coded chemicals as well as the negative and positive controls were tested by each laboratory in at least two runs. For the TTL, one chemical showed discordant results at 5 min in two (of the three) laboratories due to viabilities close to the cut-off of 50%. However, all laboratories predicted the chemical correctly in at least one run. For the TTS, all runs produced by the three laboratories were concordant and accurate.

On the basis of the results of the training and transfer phases, the PRP believes that the test method is easily learnable and transferable. Moreover, the PRP considers that the technicians from the participating laboratories demonstrated adequate proficiency in performing the SkinEthic™ HCE TTT and readiness to enter the formal validation study.

3.10 Use of quality assurance system(s) during data generation (Evaluation criterion 10)

The three laboratories participating in the validation study, EPISKIN SA, EUROSAFE, and VITROSCREEN Srl are GLP-certified. However, the training, transferability and multi-laboratory trial were all conducted in the spirit of Good Laboratory Practices (GLP). No internal audits were conducted due to the non-GLP compliance of the studies.

The test developer's laboratory is not GLP-certified and the acquisition of the data for the assessment of the predictive capacity did not follow any formal quality assurance system. However, the following safeguards recommended by Balls et al. (1995) 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.

The PRP considers this to be sufficient and appropriate.

3.11 Within- and between-laboratory reproducibility (Evaluation criterion 11)

Have the within- and between-laboratory reproducibility of the test method been demonstrated? How do they compare to the proposed acceptance criteria? Were these sufficiently high for the intended purpose?

For the TTL, a within-laboratory reproducibility (WLR) of 80% concordance for three categories was obtained by the test method developer's laboratory for 50 chemicals tested in three independent runs. In the multi-laboratory trial, the three participating laboratories obtained a WLR of 85%, 90% and 95% and a between-laboratory reproducibility (BLR) of 90% for the 20 coded chemicals tested. For the TTS, a WLR of 95% concordance for three categories was obtained by the test method developer's laboratory for 59 chemicals tested in three independent runs. In the multi-laboratory trial, all three participating laboratories obtained a WLR of 100% and a BLR of 100% for the 20 coded chemicals tested. These values are equal or higher than what has been obtained with OECD-accepted *in vitro* methods for discriminating two categories in the area of serious eye damage/eye irritation. The PRP, therefore, concludes that the SkinEthic™ HCE TTT is highly reproducible with both liquids (TTL) and solids (TTS).

The PRP considers that the higher WLR obtained in the multi-laboratory trial as compared to the WLR obtained by the test developer's laboratory may suggest that more reproducible chemicals might have been selected for the multi-laboratory trial (see evaluation criterion 7). In order to explore this hypothesis in detail, the PRP conducted an analysis of the WLR based on the 25 new chemicals that had not been tested before, i.e. excluding 15 previously tested chemicals (6 liquids and 9 solids). The TTL WLR of the three laboratories for the 14 liquids not tested before were 100%, 92.8% and 100%. The TTS WLR of 100% in all laboratories for the 20 chemicals used in the multi-laboratory trial did, of course, not change for the subset of 11 solids not tested before. As the WLR was equal or higher compared to the full set of 40 chemicals, the PRP concludes that the inclusion of 15 previously tested chemicals in the multi-laboratory trial probably did not bias the reproducibility estimates.

The PRP observes that the reproducibility with liquids appears to be slightly lower as compared to solids. The test developer did not provide any hypothesis that could possibly explain this observation.

The PRP explored the data in this regard and concludes that it is not related to the shorter exposure time used in TTL (5 min), but could be explained by the higher number of exposure times of this protocol (5, 16 and 120 min,) as compared to TTS (30 and 120 min), which increases the likelihood of threshold effects. The PRP also notes that poorly water-soluble liquids, which are tested at 20% (w/v) for the 16 min and 120 min exposures in TTL, do not seem to explain the slightly higher variability observed with TTL, since the provided data shows that these chemicals have not resulted in increased variability.

3.12 Predictive capacity (Evaluation criterion 12)

Has the predictive capacity of the test method been demonstrated? How does it compare to the proposed acceptance criteria? Was it sufficiently high for the intended purpose?

As the SkinEthic™ HCE TTT is the first test method to be assessed for discriminating the three UN GHS categories for ocular hazard, (agreed) acceptance criteria against which to evaluate the validity status of the test method specifically for this purpose did not yet exist at the time of the validation study. In fact, the PRP also notes that acceptance criteria were not described in the original validation report nor in any other document provided to the PRP. Upon request, the test method developer provided these in a revised version of the validation report, with a statement that they had been discussed with the independent statistician and were presented during the training phase to the participating laboratories. However, they were not described in the training report, nor in the study project plan and it remains thus unclear when these criteria were defined. The acceptance criteria proposed by the test method developer in the revised validation report were primarily based on established acceptance criteria for test methods identifying chemicals inducing serious eye damage (UN GHS Cat. 1) or not requiring classification (UN GHS No Cat.) and are summarised in Table 1. The PRP considered that these criteria were not the most appropriate to take into account the implications of discriminating three categories and decided therefore to define its own acceptance criteria.

Table 1: Acceptance criteria proposed by the test method developer

Test method developer acceptance criteria		SkinEthic™ HCE TTT data		
		UN GHS Cat. 1	UN GHS Cat. 2	UN GHS No Cat.
Reference data	UN GHS Cat. 1	≥ 70%	≤ 30%	≤ 0%*
	UN GHS Cat. 2	≤ 30/> 40%**	≥ 50%	≤ 20/> 30%**
	UN GHS No Cat.	≤ 10%	≤ 40/> 50%**	≥ 60/< 50%**

* 'Ideally, no Category 1 should be underpredicted as No Category'.

** considered as definitively acceptable/unacceptable.

When discussing and defining its acceptance criteria (Table 2), the PRP considered the prevalence of categories, as well as the variability in the *in vivo* reference data as described by Luechtefeld et al. (2016) and Barroso et al. (2017). The PRP agreed to set as criteria that at least 75% of *in vivo* UN GHS Cat. 1 chemicals should be predicted as such by the SkinEthic™ HCE TTT, and that at most 25% should be underpredicted as UN GHS Cat. 2. Ideally, no UN GHS Cat. 1 should be underpredicted as No Cat. However, the PRP recognises that the uncertainty associated with the *in vivo* data is high, even if following the recommendations of Barroso et al. (2017) for selecting chemicals for validation studies. Therefore, the PRP considers that an underprediction rate of *in vivo* UN GHS Cat. 1 chemicals as UN

GHS No Cat. of ≤ 5% is acceptable. The PRP also agreed to set as criteria that at least 70% of *in vivo* UN GHS No Cat. chemicals should be predicted as such by the SkinEthic™ HCE TTT, allowing for a maximum of 30% overpredictions as UN GHS Cat.2 and a maximum of 5% overprediction as UN GHS Cat. 1. The PRP considered it necessary to set a higher target of acceptance for the proportion of correctly identified UN GHS No Cat. than used in the validation of EpiOcular™ EIT and SkinEthic™ HCE EIT (described in TG 492), where ≥ 60% was considered as definitely acceptable. The reason for this is that, while for the RhCE-based methods described in TG 492 a false positive would need further testing with other methods to confirm the definitive UN GHS category, in SkinEthic™ HCE TTT a false positive result for a UN GHS No Cat. chemical will immediately lead to an overclassification as UN GHS Cat. 2 or Cat. 1, and this should be avoided whenever possible. Regarding the identification of *in vivo* UN GHS Cat. 2 chemicals, the PRP took the high variability of Draize acute eye irritation data for the chemicals in this category into account. For example, Luechtefeld et al. (2016) reported conditional probabilities for 491 chemicals with at least two Draize acute eye irritation study records submitted in REACH registrations. For the 224 chemicals falling into UN GHS Cat. 2 (including 2A and 2B), the likelihood of being a UN GHS No Cat. when re-tested was 67.4%. A similar observation was reported with a smaller dataset by Barroso et al. (2017). Therefore, the PRP agreed that at least 50% of the *in vivo* UN GHS Cat. 2 chemicals should be predicted as such by the SkinEthic™ HCE TTT, and that both the proportion of UN GHS Cat. 2 chemicals mispredicted as UN GHS Cat. 1 or as UN GHS No Cat. should not exceed 30%.

Table 2: Acceptance criteria defined by the PRP

PRP acceptance criteria		SkinEthic™ HCE TTT data		
		UN GHS Cat. 1	UN GHS Cat. 2	UN GHS No Cat.
Reference data	UN GHS Cat. 1	≥ 75%	≤ 25%	≤ 5%*
	UN GHS Cat. 2**	≤ 30%	≥ 50%	≤ 30%
	UN GHS No Cat.	≤ 5%	≤ 30%	≥ 70%

* recognising that the uncertainty associated with the *in vivo* data is high, even if following the recommendations of Barroso et al. (2017) for selecting chemicals for validation studies.

** considering *in vivo* reference data variability described by Luechtefeld et al. (2016) and Barroso et al. (2017).

The SkinEthic™ HCE TTL correctly identified 85.4 % UN GHS Cat. 1 (n = 21), 79.8% UN GHS Cat. 2 (n = 25) and 79.2% UN GHS No Cat. (n = 24) of the 70 liquid chemicals tested for assessing predictive capacity. None of the UN GHS Cat. 1 and UN GHS Cat. 2 liquids were predicted as UN GHS No Cat. As shown in Table 3, all of the PRP acceptance criteria for predictive capacity described in Table 2 were met.

Regarding the SkinEthic™ HCE TTS, the PRP assessed the predictive capacity including the chemical 2,5-hexanediol, 2,5-dimethyl-, since it was selected, even though Barroso et al. (2017) recommended not to use it for validation purposes. The SkinEthic™ HCE TTS correctly identified 74.7 % UN GHS Cat. 1 (n = 29), 55.3% UN GHS Cat. 2 (n = 19) and 71.7% UN GHS No Cat. (n = 33) of the 81 solid chemicals tested for assessing predictive capacity. None of the UN GHS Cat. 1 solids were predicted as UN GHS No Cat. As shown in Table 3, all of the PRP acceptance criteria for predictive capacity described in Table 2 were met.

Regarding the combined predictive capacity of the SkinEthic™ HCE TTT (TTL and TTS), the PRP concludes that all of the PRP acceptance criteria for predictive capacity described in Table 2 were met.

In addition, the PRP notes that the proportion of correctly predicted UN GHS Cat. 1 is comparable to OECD accepted methods for identifying UN GHS Cat. 1 chemicals (vs. the rest) and that the proportion of correctly predicted UN GHS No Cat. is comparable to OECD accepted methods for identifying UN GHS No Cat. chemicals (vs. the rest).

Table 3: Performance of the SkinEthic™ HCE TTL and TTS protocols as well as of the TTT test method

A. SkinEthic™ HCE TTL (n = 70)		UN GHS Cat. 1	UN GHS Cat. 2	UN GHS No Cat.
Reference data	UN GHS Cat. 1 (n = 21)	85.4%	14.6%	0%
	UN GHS Cat. 2 (n = 25)	20.2%	79.8%	0%
	UN GHS No Cat. (n = 24)	0%	20.8%	79.2%

B. SkinEthic™ HCE TTS (n = 81)		UN GHS Cat. 1	UN GHS Cat. 2	UN GHS No Cat.
Reference data	UN GHS Cat. 1 (n = 29)	74.7%	25.3%	0%
	UN GHS Cat. 2 (n = 19)	15.8%	55.3%	28.9%
	UN GHS No Cat. (n = 33)	3.0%	25.3%	71.7%

C. SkinEthic™ HCE TTT (n = 151)		UN GHS Cat. 1	UN GHS Cat. 2	UN GHS No Cat.
Reference data	UN GHS Cat. 1 (n = 50)	79.2%	20.8%	0%
	UN GHS Cat. 2 (n = 44)	18.3%	69.2%	12.5%
	UN GHS No Cat. (n = 57)	1.8%	23.4%	74.9%

The SkinEthic™ HCE TTT showed higher predictive capacity with liquids than with solids. As discussed in evaluation criterion 8, this is likely the result of the higher uncertainty associated with Draize acute eye irritation test reference data for solid chemicals due to more variable exposure conditions *in vivo* (e.g., uncontrolled exposure time, mechanical abrasion, secondary effects). This is especially true for studies performed before the revision of OECD TG 405 in 2002, which introduced a washing step after one hour of exposure for solids. Focusing on the predictivity of chemicals grouped by drivers of classification, the PRP observes that:

a) the underprediction rate of UN GHS Cat. 1 chemicals was higher for solids having CO mean ≥ 3 than for those classified based only on persistence of effects at day 21, but the opposite was observed for liquids. This could again be due to uncontrolled exposure in many *in vivo* studies with solids leading to mechanical abrasion (especially those conducted before 2002);
b) the overprediction of UN GHS No Cat. chemicals with CO scores greater than 0 in at least one animal and at least one observed time point (CO > 0) was higher (38.9% for TTL, 66.7% for TTS and 52.8% overall) than the overprediction rate (14.8% for TTL, 19.7% for TTS and 17.8% overall) for UN GHS No Cat. chemicals with CO scores equal to 0 in all animals and all observed time points (CO = 0). The PRP considers this as acceptable because the prevalence of CO > 0 in the DRD (22%) is much lower than that of CO = 0 (78 %) for UN GHS No Cat. chemicals. In fact, the PRP observes that the distribution of the subgroups CO > 0 (12 out of 57 chemicals; 21%) and CO = 0 (45 out of 57 chemicals; 79%) in the test chemicals used to assess the predictive capacity of SkinEthic™ HCE TTT is representative of the respective distribution in the DRD. Chemicals with CO > 0 induce mild irritation in the Draize acute eye irritation test that does

not warrant classification, but the SkinEthic™ HCE TTT is very sensitive in detecting such *in vivo* effects and the PRP does not consider this to be a limitation of the method;

c) the UN GHS No Cat. chemicals with CO ≥ 0 and at least one animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint (CO = 0 ** and CO > 0**) as described in the DRD are underrepresented in the chemicals selection for this validation study. To be representative of the distribution of the DRD in the chemicals selection, the selection of 57 UN GHS No Cat. chemicals should have had 44 CO = 0; 1 CO = 0 **; 7 CO > 0 and 5 CO > 0 **. Instead, it has 45 CO = 0; 11 CO > 0; and 1 CO > 0 **. However, the PRP considers that the inclusion of more chemicals from the subgroups CO = 0** and/or CO > 0** in the validation study probably would have had only a minor effect, if any, on the predictive capacity. This is because the proportion of CO = 0** UN GHS No Cat. chemicals in the DRD is very low (1.7%) and the overprediction rate of CO > 0 chemicals is already high (52.8%). The proportions of CO > 0 and CO > 0** UN GHS No Cat. chemicals in the DRD are 13.1% and 8.7%, respectively.

The PRP also noted that the potential lack of independence in the selection of chemicals may have led to an underrepresentation of more challenging chemicals, possibly resulting in a bias in the estimation of the predictive capacity. In order to explore this hypothesis in detail, the PRP conducted an analysis of the predictive capacity based on the 25 chemicals that had not been tested before, i.e. excluding 15 previously tested chemicals (6 liquids and 9 solids). Thirteen of the 14 liquids not tested before were correctly predicted by the three laboratories in all runs, while one UN GHS Cat. 2 liquid was correctly classified in seven runs and overpredicted as UN GHS Cat. 1 in two runs. Nine of the 11 solids not tested before were correctly predicted by the three laboratories in all runs, while one UN GHS Cat. 2 solid was consistently overpredicted as UN GHS Cat. 1 and one UN GHS No Cat. solid was consistently overpredicted as UN GHS Cat. 2. As the predictive capacity for this subset of 25 new chemicals was similar to the predictive capacity for the full set of 40 chemicals used in the multi-laboratory trial and similar or higher when compared to all 126 chemicals previously tested by the developing laboratory, the PRP concludes that the inclusion of 15 previously tested chemicals in the multi-laboratory trial probably did not bias the predictive capacity estimates.

Therefore, the PRP concludes that the predictive capacity of the SkinEthic™ HCE TTT for the purpose of discriminating the three UN GHS categories for serious eye damage/eye irritation is sufficient and adequate. Furthermore, although the SkinEthic™ HCE TTT was not developed to account for all mechanisms associated with serious eye damage/eye irritation, the above performances demonstrate that the SkinEthic™ HCE TTT method is able to predict serious eye damage/eye irritation with sufficient accuracy independently of the types of ocular effects observed *in vivo* (i.e., corneal, iridal and conjunctival injuries; including severity and persistence of effects).

3.13 Applicability domain and limitations (Evaluation criterion 13)

The PRP considers that the applicability domain and limitations of the test method were properly defined in the validation report. These were established using 70 liquids and 81 solids representing a broad range of chemical classes, physicochemical properties, organic functional groups and *in vivo* drivers of classification (see evaluation criterion 7). In terms of UN GHS Categories, the 151 validation chemicals were distributed as follows: 50 UN GHS Cat. 1 (21 liquids and 29 solids), 44 UN GHS Cat. 2 (25 liquids and 19 solids) and 57 UN GHS No Cat. (24 liquids and 33 solids) chemicals. Furthermore, a

wide range of chemical types, chemical classes, molecular weights, Log P, chemical structures, organic functional groups, *in vivo* drivers of classification, etc., were tested.

Solid chemicals with poor water solubility were frequently underpredicted by SkinEthic™ HCE TTS in the validation study (6 out of 9 chemicals; 67%). Since solid chemicals are tested neat in SkinEthic™ HCE TTS (in both the 30 min and 120 min exposure times), the PRP considers that poor water solubility should not negatively impact the bioavailability of the chemicals to the tissues as it would if chemicals were tested in dilution. Therefore, the PRP hypothesises that the *in vivo* positive results obtained for these chemicals are due to mechanical damage or secondary effects that are not related to the chemical itself since this type of chemical would be more difficult to clear from the rabbit conjunctival cul-de-sac by tearing. No other relationships between the SkinEthic™ HCE TTT mispredictions obtained in the validation study and specific physicochemical properties, organic functional groups or *in vivo* drivers of classification could be identified by the test method developer.

The majority of the chemicals tested in the validation of SkinEthic™ HCE TTT represented mono-constituent chemicals (n=132), tested neat (52 liquids, 80 solids) or in dilution (0.01 to 30%), and therefore, there is limited information on test method performance with mixtures. Nevertheless, the PRP considers that the applicability domain and limitations of the SkinEthic™ HCE TTT are similar to those reported in OECD TG 492 since SkinEthic™ HCE TTT uses the same tissue constructs and the same principles as SkinEthic™ HCE EIT described in OECD TG 492 (see evaluation criterion 3). Indeed, based on the results obtained in the validation study, no additional limitations regarding applicability could be identified beyond those already identified in TG 492 for other RhCE-based test methods. Furthermore, the PRP acknowledges that including a wide range of chemical mixtures in validation studies currently raises several problems, e.g. availability of *in vivo* data, selection of test mixtures, and continuity of supply.

Thus, the PRP agrees with the statement in the validation report that the SkinEthic™ HCE TTT is applicable to chemicals 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. Chemicals in the form of gases and aerosols were also not assessed in the validation study. While it is conceivable that these can be tested using RhCE technology, they are currently considered to be out of the applicability domain of RhCE-based test methods. The SkinEthic™ HCE TTT is not intended to differentiate between eye irritants (optional Category 2A) and mild eye irritants (optional Category 2B), as defined by UN GHS.

The only difference between SkinEthic™ HCE TTT and SkinEthic™ HCE EIT that may affect the applicability domain and limitations of one method versus the other is that liquid chemicals are tested in the TTL protocol at 20% (w/v) for the exposure times of 16 and 120 min (neat for 5 min exposure), while all chemicals are tested neat in EIT. Liquids not forming a stable suspension in water for the duration of the exposure in TTL may lead to overprediction if denser than water (higher apparent concentration in contact with the tissue) or underpredictions if less dense than water (lower apparent concentration in contact with the tissue). The validation data are too limited to be able to do a robust analysis of the effect of poorly soluble chemicals on the performance of the method, but based on the information available, the PRP notes that this does not appear to be a major issue (see evaluation criterion 5).

3.14 Completeness of data and documentation (Evaluation criterion 14)

Were all data and documents supporting the assessment of the validity of the test method available and complete?

The data and documents provided to the PRP were well structured and allowed to readily locate specific information. However, in order to perform a fully comprehensive review, the PRP sought to obtain further information from the test method developer. While the additional information provided by the test developer was generally helpful, some relevant material was not provided, including:

- a detailed statistical report including the chemical codes used in the multi-laboratory trial;
- justifications of the number of chemicals used in all phases, but especially in the multi-laboratory trial;
- a justification for why a substantial proportion (15 of 40) of the chemicals selected for the multi-laboratory trial had been previously tested by the test method developer;
- a statement from Adriaens Consulting describing her involvement in the chemicals selection and sample size determination;
- a clear description of the distribution of tasks between the sponsor and Adriaens Consulting in terms of coordination and chemicals selection;
- a clear timeline for the testing of chemicals by different laboratories covering all test method development and validation phases;
- demonstration that the acceptance criteria for the study against which to evaluate the validity status of the test method in terms of reproducibility and predictive capacity were defined before the completion of the study;
- the point in time when the prediction model of the SkinEthic™ HCE TTT was defined;
- additional information on the training and tests phases before the multi-laboratory trial, especially regarding protocol optimisation aspects, such as the prediction model development and exposure time determination;
- instructions on how to decide which protocol to use.

However, the PRP considers the documentation provided to be sufficiently complete to assess the validity of the SkinEthic™ HCE TTT.

3.15 Other relevant aspects (Evaluation criterion 15)

The PRP notes that a quality audit of the tissue production has been conducted on the French production site in the context of the validation of the SkinEthic™ HCE EIT, which uses the same tissue.

4. Conclusion on the scientific validity for the intended purpose

On the basis of all available data, the PRP concludes that the relevance and reliability of the SkinEthic™ HCE TTT have been satisfactorily demonstrated for the intended purpose of discriminating the three UN GHS categories for serious eye damage/eye irritation, i.e. UN GHS Cat. 1, Cat. 2 and No Cat. Therefore, despite the shortcomings noted in the study management and conduct as described in evaluation criteria 6, 7 and 14, the PRP considers the test method to be scientifically valid for regulatory use as a full replacement to the *in vivo* Draize acute eye irritation test for classification of chemicals according to UN GHS. The SkinEthic™ HCE TTT is, however, not intended to differentiate between eye irritants (optional Cat. 2A) and mild eye irritants (optional Cat. 2B), as defined by UN GHS.

References

Alépée N, Adriaens E, Abo T, Bagley D, Desprez B, Hibatallah J, Mewes K, Pfannenbecker U, Sala À, Van Rompay AR, Verstraelen S, McNamee P (2019a). Development of a Defined Approach for eye irritation or serious eye damage for liquids, neat and in dilution, based on Cosmetics Europe analysis of *in vitro* STE and BCOP test methods. *Toxicol In Vitro*. 57:154-163.

Alépée N, Adriaens E, Abo T, Bagley D, Desprez B, Hibatallah J, Mewes K, Pfannenbecker U, Sala À, Van Rompay AR, Verstraelen S, McNamee P (2019b). Development of a Defined Approach for eye irritation or serious eye damage for neat liquids based on Cosmetics Europe analysis of *in vitro* RhCE and BCOP test methods. *Toxicol In Vitro*. 59:100-114.

Barroso J, Pfannenbecker U, Adriaens E, Alépée N, Cluzel M, De Smedt A, Hibatallah J, Klaric M, Mewes KR, Millet M, Templier M, McNamee P (2017). Cosmetics Europe compilation of historical serious eye damage/eye irritation in vivo data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: the Draize eye test Reference Database (DRD). *Arch Toxicol.* 91(2):521-547.

EPA (2015). Use of an alternate testing framework for classification of eye irritation potential of EPA pesticide products. Office of Pesticide Programs. U.S. Environmental Protection Agency. Available at: https://www.epa.gov/pesticide-registration/alternate-testing-framework-classification-eye-irritation-potential-epa.

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V (2004). A modular approach to the ECVAM principles on test validity. *Altern Lab Anim.* 32(5):467-472.

Luechtefeld T, Maertens A, Russo DP, Rovida C, Zhu H, Hartung T (2016). Analysis of Draize eye irritation testing and its prediction by mining publicly available 2008-2014 REACH data. *ALTEX.* 33(2):123-34.

OECD (2005). Series on Testing and Assessment No. 34: Guidance Document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Environment, Health and Safety Publications. Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm.

OECD (2016). Series on Testing and Assessment No. 255: Guidance Document on the reporting of Defined Approaches to be used within Integrated Approaches to Testing and Assessment. OECD Environment, Health and Safety Publications. Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm.

OECD (2017). Series on Testing and Assessment No. 216: Revised 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. Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm.

OECD (2019). Series on Testing and Assessment: Guidance Document No. 263 on Integrated Approaches to Testing and Assessment (IATA) for serious eye damage and eye irritation. OECD Environment, Health and Safety Publications. Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm.

Prinsen MK (2006). The Draize Eye Test and *in vitro* alternatives; a left-handed marriage? *Toxicol in Vitro.* 20:78–81.