Supporting Document for Evaluation and Review of Draft Guideline (GL) For Defined Approaches (DAs) for Serious Eye Damage / Eye Irritation for Solids

Table of contents

Tables

Figures

List of acronyms

BCOP: Bovine Corneal Opacity and Permeability
CASRN: 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
CC: conjunctival chemosis
CO: corneal opacity
CON4EI: CONsortium for *in vitro* Eye Irritation testing strategy
Conj: conjunctival effects
CR: conjunctival redness
DA: Defined approach
DAL: Defined approach liquids
DAS: Defined approach solids
DIP: data interpretation procedure
DRD: Draize eye test Reference Database
ECHA: European Chemicals Agency
EITS: Eye Irritation Test Solids
EURL ECVAM: European Union Reference Laboratory on Alternatives to Animal Testing
FN: false negative
FP: false positive
GL: guideline
HCE: Human Corneal Epithelium
IATA: Integrated Approaches to Testing and Assessment
IR: iritis
LLBO: laser light-based opacitometer
MoA: Modes of Action
MSDS: material safety data sheet
MW: molecular weight
No Cat.: Not requiring UN GHS Classification
OECD: Organisation for Economic Co-operation and Development
OFG: Organic Functional Groups
RhCE: Reconstructed human Cornea-like Epithelium
SPSF: Standard Project Submission Form
UN GHS: United Nations Globally Harmonized System
WNT: Working Group of National Co-ordinators of the Test Guidelines programme

1. Introduction

1. In June 2022 two defined approaches (DAs) for non-surfactant liquids were accepted (DAL-1 and DAL-2) and were integrated in a new OECD test guideline (TG) for serious eye damage/eye irritation i.e., discrimination between the three United Nations Globally Harmonized System of Classification (UN GHS) categories (Part I and Part II, OECD TG 467 2022). Recently, a DA has been developed to specifically address the eye hazard potential of solids across the 3 UN GHS categories. In November 2022, a Standard Project Submission Form (SPSF) on the DA for solids (DAS) was submitted by France to the OECD secretariat. The SPSF submitters updated the draft SPSF based on recommendations from the Expert Group on Eye/Skin Irritation/Corrosion & Phototoxicity and this version was shared with the WNT. In April 2023, the WNT (Working Group of National Co-ordinators of the Test Guidelines programme) accepted the SPSF on the DAS. The first steps for developing the DA were taken by Cosmetics Europe, after which the DA for solids was further developed by L'Oréal and supported by L'Oréal only. The new DAS is proposed to be integrated in TG 467 as part III.

2. According to the UN GHS classification system, Category 1 (serious eye damage) refers to the production of tissue damage in the eye, or serious physical decay of vision, which is not fully reversible, occurring after exposure of the eye to a substance or mixture. Category 2 (eye irritation) refers to the production of changes in the eye, which are fully reversible, occurring after the exposure of the eye to a substance or mixture. Based on this definition, the hazard potential of a test chemical is determined in the Draize eye test (OECD TG 405, 2023) based on its effect on corneal opacity (CO), iritis (IR), conjunctival redness (CR), and conjunctival chemosis (CC). Based on the severity of effects and/or the timing of their reversibility, classifications are derived by the UN GHS (UN 2023). Effects not fully reversed at the end of the 21 day observation period of the Draize test are considered irreversible (Category 1) or not (Category 2). Cat. 2 may be divided into the optional Categories 2A (effects fully reversible within 21 days) and 2B (effects fully reversible within 7 days). When none of the Cat. 1 or Cat. 2 classification criteria are met, the chemical does not require classification which corresponds with No Category (No Cat.). Note that every time reference is made to in vivo Cat. 1, Cat. 2, and No Cat. in this background review document, those classifications have been derived from testing in albino rabbits according to the Draize eye test method (OECD TG 405). The main data source of the historical data was

the Draize eye test Reference Database (DRD) published by Cosmetics Europe (Barroso et al., 2017; see §15).

3. A comprehensive analysis to address the main *in vivo* ocular tissue effects that drive UN GHS classification was conducted and the outcomes were used to evaluate the performance of the DAS described in the present document. The analyses identified nine different criteria from the four *in vivo* tissue effects (CO, IR, CR, and CC) that can each independently drive the classification of a chemical (Barroso et al., 2017). Of note, CR and CC were not reported separately but were reported together as conjunctival effects (Conj) because previous analyses revealed that CC rarely drives the classification of chemicals in the absence of CR effects (Adriaens et al., 2014; Barroso and Norman, 2014). Chemicals classified as Cat. 1 were grouped based on (i) severity (mean scores of days 1–3); (ii) persistence of any ocular effect on day 21 in the absence of severity; or (iii) CO = 4 (at any observation time during the study) in the absence of both severity and persistence (or if unknown). Chemicals classified as Cat. 2 were allocated to one of the three following groups based on the main endpoint leading to Cat. 2 classification, i.e., "CO", "Conj", and "IR". Studies with chemicals not requiring classification for serious eye damage/eye irritation (No Cat.) were distributed in four different groups depending on whether they showed CO scores equal to 0 in all animals and all observed time points (CO = 0 and CO = 0**) or not (CO > 0 and CO > 0**). No Cat. studies for which at least one animal had a mean of the scores of days 1–3 above the classification cut-off for at least one endpoint but not in enough animals to generate a classification (borderline cases) were marked with ** (CO = 0**, CO > 0**). A detailed description of the drivers of classification and use of the terms CO, IR and Conj to describe key effects is provided in the paper of Barroso and co-workers (2017).

1.1. 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.

Barroso J, Norman K (2014). REACHing for alternatives to animal testing. A webinar series on modern testing strategies for REACH. Webinar 3 of 6 on "Serious Eye Damage and Eye Irritation", December 4, 2014. http://www.piscltd.org.uk/reaching-alternatives-animal-testing/ (accessed 17.06.2016).

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 (2017) 91:521–547.

OECD (2023). Test No. 405: Acute Eye Irritation/Corrosion. In: OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264185333-en

UN (2023). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30/Rev.10, Tenth Revised Edition, New York and Geneva: United Nations. Available at [https://unece.org/sites/default/files/2023-07/GHS%20Rev10e.pdf]

2. Presentation of the DA analysed

2.1. Introduction

4. This document supports the integration of a DA for solids into OECD TG 467, Defined Approaches for Serious Eye Damage and Eye Irritation as Part III. The DA for neat solids (hereinafter referred to as DAS) is based on a combination of a Reconstructed human Cornea-like Epithelium (RhCE) test method (OECD TG 492, 2023a), and the Bovine Corneal Opacity and Permeability (BCOP) test method using the laser light-based opacitometer (LLBO) (OECD TG 437, 2023b).

5. The applicability domain of the DAS is restricted to neat solids (i.e., not pipettable test substances). The DAS that is proposed in the current document is a refinement of the initially proposed defined approaches that resulted from the CONsortium for *in vitro* Eye Irritation testing strategy (CON4EI) project (Adriaens et al., 2018). During the CON4EI project, 80 chemicals (liquids and solids) were tested with 8 different alternative test methods (including OECD TG 437 and TG 492 test methods). The chemicals were chosen in collaboration with Cosmetics Europe from the Draize eye test Reference Database (DRD) developed by Cosmetics Europe (Barroso et al., 2017). Additional analyses performed by L'Oréal on a larger set of solids showed that the combination of the SkinEthic™ Human Corneal Epithelium (HCE) Eye Irritation Test (EITS) applying the solids protocol and the BCOP LLBO for solids could distinguish between the three UN GHS categories.

6. The DAS has proven useful in making predictions across the whole range of UN GHS categories i.e., Category 1 (Cat. 1) on "serious eye damage"; Category 2 (Cat. 2) on "eye irritation" and No Category (No Cat.) for chemicals "not requiring classification and labelling" for eye irritation or serious eye damage (UN GHS, 2023). Whilst the components of the DAS are accepted stand-alone OECD TGs which can be used to identify Cat. 1 (OECD TG 437, 2023b) and chemicals that do not require classification for eye irritation or serious eye damage (No Cat.; OECD TG 492, 2023a), the DAS also allows the classification of Cat. 2. However, the DAS is not designed to distinguish between Categories 2A and 2B.

7. This supporting document provides information on the evaluation of the proposed DAS for identification of GHS Cat. 1, Cat. 2, and No Cat., that is proposed to include as Part III in OECD TG 467 on DAs for serious eye damage/eye irritation. Information resulting from the application of the DAS contained in the final TG will be used, either on its own or in conjunction with other information, to meet regulatory data requirements for serious eye

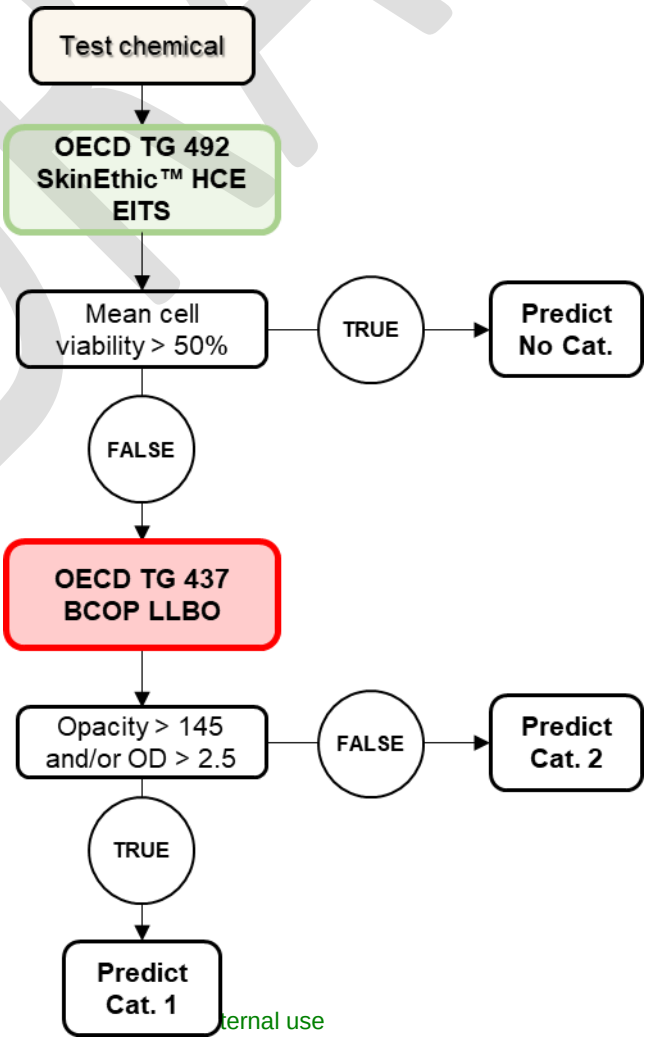
damage/eye irritation and will be covered under the agreement on the mutual acceptance of data (MAD).

2.2. DAS

8. The DAS presented in this document describes the combination of two *in vitro* test methods (RhCE and BCOP LLBO) for the identification of the eye hazard potential for neat solids primarily for the purposes of classification and labelling without the use of animal testing. The RhCE model that is part of DAS is the SkinEthic™ Human Corneal Epithelium (HCE) EITS as described in OECD TG 492 (2023a).

9. The data interpretation procedure (DIP) applied uses the readout of the prediction models of each of the individual test method as defined by the Test Guidelines. A scheme of DAS is presented in Figure 2.1. Solids are evaluated based on the SkinEthic™ HCE EITS test method in Step 1. Solids that result in a tissue viability > 50% are classified No Cat. Solids that result in a tissue viability ≤ 50% are evaluated based on the BCOP LLBO test method in a second step. Solids that result in an opacity > 145 and/or OD > 2.5 are predicted Cat. 1 and the remaining solids are assigned Cat. 2.

Figure 2.1. Scheme of the DAS; step 1 SkinEthic™ HCE EITS test method used to identify No Cat., and step 2 BCOP LLBO used to identify Cat. 1



2.3. References

Adriaens E, Verstraelen S, Alépée N, Kandarova H, Drzewiecka A, Gruszka K, Guest R, Willoughby JA, Van Rompay AR (2018). CON4EI: Development of testing strategies for hazard identification and labelling for serious eye damage and eye irritation of chemicals. Toxicol. in Vitro 49, 99–115. https://doi.org/10.1016/j.tiv.2017.09.008.

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 (2017) 91:521–547.

OECD (2023a), Test No. 492: Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264242548-en.

OECD (2023b), Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264203846-en.

UN (2023). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30/Rev.10, Revised Edition, New York and Geneva: United Nations. Available at [https://read.un-ilibrary.org/environment-and-climate-change/globally- https://unece.org/sites/default/files/2023-07/GHS%20Rev10e.pdf

3. *In vivo* reference data (Draize eye test)

10. The data source was the Draize eye test Reference Database (DRD) published by Cosmetics Europe (Barroso et al., 2017). The DRD contains 681 independent Draize rabbit eye test studies and was compiled using various sources of historical Draize eye test data, produced according to OECD TG 405, which were created to support previous validation activities (Barroso et al., 2017). Detailed information on the UN GHS categories, the drivers of classification, the organic functional groups present, and the identification of chemicals that should not be used for the evaluation of alternative methods and/or testing strategies can be retrieved from the DRD Supplementary Material 1.

3.1. Criteria applied for the selection of the reference chemicals for the DAS

11. The following criteria, as identified by Barroso and co-workers (2017), were considered when selecting the reference chemicals: (1) the expected applicability of the DAS in terms of UN GHS prediction (No. Cat., Cat. 2, and Cat. 1; §6), (2) important drivers of classification, (3) purity of the chemicals, and (4) relevance of the chemicals in terms of their representative functional and chemical classes and industrial use.

3.1.1. Drivers of classification

12. The chemical selection was performed by taking into account several key criteria that were identified by Barroso and co-workers (2017). One of the key criteria is that the pool of reference chemicals needs to address the main ocular tissue effects that drive classification. In the Draize rabbit eye test, the hazard potential of a test chemical is determined based on its effect on corneal opacity (CO), iritis (IR), conjunctival redness (CR), and conjunctival chemosis (CC). Based on the severity of effects and/or the timing of their reversibility, classifications are derived according to the serious eye damage/eye irritation classification criteria defined by UN GHS (UN 2021).

As described by Barroso and co-workers (2017), there are nine different criteria derived from the four tissue effects (CO, IR, CR, and CC) that can each independently drive the classification of a chemical (**Table 3-1**).

Table 3-1. Drivers of UN GHS classification

Category 1 Irreversible effects on the eye/serious eye damage						Category 2 Reversible effects on the eye/eye irritation		
Severity (Mean scores of Days 1-3) [a]		Persistence on Day 21			Severe CO	Severity (Mean scores of Days 1-3) [a]		
CO mean ≥ 3	**IR mean > 1.5**	**CO**	**Conj**	**IR**	**CO=4**	**CO mean ≥ 1**	**Conj mean ≥ 2**	**IR mean ≥ 1**
in ≥ 60% of the animals	in ≥ 60% of the animals	in at least one animal	in at least one animal	in at least one animal	in at least one animal	in ≥ 60% of the animals	in ≥ 60% of the animals	in ≥ 60% of the animals

CO: corneal opacity; IR: iritis; Conj: conjunctival redness (CR) and/or conjunctival chemosis (CC)

Drivers with a greyed background correspond with the most important drivers.

[a] Mean scores are calculated from gradings at 24, 48, and 72 hours after instillation of the test chemical

3.1.2. Key criteria considered when selecting reference chemicals

13. According to Barroso and co-authors (2017), corneal opacity is the most important endpoint driving Cat. 1 classification, corneal opacity and conjunctival effects are the most important endpoints driving Cat. 2 classification. The most important drivers of Cat. 1 (3 different criteria) and Cat. 2 (2 different criteria) classification are shown with a greyed background in **Table 3-1** and are listed below. Note that for chemicals that were classified based on the driver CO persistence on day 21 or CO = 4, this effect should be present in at least 60% of the animals as advised by Barroso and co-authors (2017).

Drivers of classification for Cat. 1, by order of importance:

(1) CO mean ≥ 3 (days 1 – 3) in ≥ 60% of the animals;
(2) CO persistence on day 21 (D21) in ≥ 60% [a] of the animals (with CO mean < 3);
(3) CO = 4 in ≥ 60 % [b] of the animals in the absence of both severity and persistence or if unknown).

[a] Note that the 60% criterion applied for selection of the reference chemicals differs from the GHS criteria for the Draize test, in which a substance that produces in at least one animal effects on the cornea, iris or conjunctiva that are not expected to reverse or have not fully reversed within an observation period of normally 21 days.

[b] Cat. 1 is also adopted for substances that result in grade 4 cornea lesions and other severe reactions (e.g. destruction of cornea) observed at any time during the test.

Drivers of classification for Cat. 2, by order of importance:

(4) CO mean ≥ 1;
(5) CR mean ≥ 2 (with CO mean < 1)

Subgroups for chemicals that do not require classification (No Cat.):

(6) CO > 0 (minor effects on CO observed)
(7) CO = 0 (clear negative results)
(8) CO = 0 ** and CO > 0 ** (only a few chemicals should be included)

CO = 0: CO scores equal to 0 in all animals and all observed time points

CO > 0: in at least one observation time in at least one animal and all animals showing mean scores of days 1–3 below the classification cut-offs for all endpoints.

** Indicates at least one animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint

3.1.3. Purity of the chemicals

14. According to OECD GD 34, the reference chemicals should have a well-defined chemical structure and purity (OECD, 2005). The chemicals tested should, where possible, be of the highest available purity, or be of known composition.

15. The set of reference chemicals to support the review of the DAS was composed of 109 neat solids. The purity of the chemicals should be as high as possible and ideally ≥ 95% (Barroso et al., 2017). The purity reported in the DRD supplementary Material 1 applies to available purity for the commercial source as indicated in the DRD, in fact the commercial source in the DRD is provided as an example (Barroso et al., 2017). Annex A.1 includes the purity of the chemical as tested in the *in vivo* study in case this was known. For the *in vitro* methods, the highest purity that was commercially available, was tested.

3.1.4. Chemical class, functional groups and uses

16. The set of solids covers a broad range of uses and chemical classes, containing small and large molecules, and hydrophobic and hydrophilic chemicals, with a wide range of organic functional groups represented (112 different OFGs) defined according to OECD QSAR Toolbox analysis version 3.2; https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm). The most common OFGs are listed in chapter 7.3.

3.1.5. Criteria used for chemicals that should not be selected according to Barroso and co-authors (2017)

17. Health and safety issues relating to transport and handling of chemicals were also taken into account. The chemical selection avoided substances known to have critical toxicological and/or unstable physical properties (*e.g.*, carcinogens, mutagens, lethal by inhalation) evident from official classifications and material safety data sheet (MSDS) information.

18. In general, chemicals that were classified based on one of the following criteria only (based on Draize eye test) were not included in the reference set of chemicals since they were identified as *"should not be used"* in either prospective studies or retrospective evaluations:

- Chemicals classified as Cat. 1 based only on persistence of CR and/or CC equal to 1 on day 21.

The reasoning behind this is that in terms of biological relevance, persistence of low-level conjunctival effects (CR/CC = 1 on day 21) in the absence of any other Cat. 1 triggering effects should not have resulted in a Cat. 1 classification. A Cat. 1 classification in case of a repeat study is highly unlikely, especially when the effect is observed in only 1 out of 6 animals and where the other 5 animals are completely recovered by day 21. The DRD contains in total 3 liquids where the Cat. 1 classification was driven by conjunctival persistence on day 21 in 1/6 animals. Those substances are identified as "*should not be used*" in the DRD. No solids with those criteria are reported in the DRD.

- Chemicals that are classified Cat. 1 in the absence of any other Cat. 1 triggering effect (none of the Cat. 1 drivers listed in **Table 3-1** could be assigned, those chemicals are listed in the DRD supplementary Material 1 https://static-content.springer.com/esm/art%3A10.1007%2Fs00204-016-1679-x/MediaObjects/204_2016_1679_MOESM1_ESM.pdf with the label "other observations" in the column "Specific observations") should in general not be selected as this accounts for a very limited number of studies in the DRD (9 substances in total – 5.5% of the Cat. 1 substances in the DRD, including 8 solids with CASRNs: 537-21-3, 54029-45-7, 3248-91-7, 6443-90-9, 74578-10-2, 14866-33-2, 35695-36-4, 13977-28-1).

3.2. Key elements for evaluation of the DAS versus the Draize eye test

19. It was recognized that determination of the most relevant *in vivo* endpoint(s), in particular the effects on cornea, iris or conjunctiva, is extremely important for the development of adequate *in vitro* methods and will allow better understanding of the relationship between the *in vitro* and the *in vivo* data (Scott et al., 2010). A comprehensive in-depth analysis of historical *in vivo* rabbit eye data provided insight into which of the observed *in vivo* effects are important in driving the classification of chemicals for serious eye damage/eye irritation according to the UN GHS, concluding that full replacement of *in vivo* testing for eye hazard will require accounting for the impact of the *in vivo* tissue effects which drive classification (Adriaens et al. 2014). Further, the uncertainty (variability) of the *in vivo* reference data is also recognized as a challenging factor that may hinder the successful development of non-animal approaches and should be considered when evaluating/validating *in vitro* test methods and strategies. For example, it has been shown that, for the rabbit eye test, the likelihood of achieving the same classification upon repeat testing is <50% for substances which fall into the mild to moderate irritation range (Luechtefeld et al., 2016). It is therefore challenging to align the results from *in vitro* methods to the *in vivo* rabbit test for the middle category (UN GHS Cat 2). Next, a database of Draize data was compiled (Cosmetics Europe Draize eye test Reference Database, DRD) and an evaluation of the various *in vivo* drivers of classification compiled in the database was

performed to establish which of these are most important from a regulatory point of view (Barroso et al., 2017). These analyses established the most important drivers for Cat. 1 and Cat. 2 classification and the distribution in different groups for the chemicals that do not require classification. Further, a number of key criteria were identified that should be taken into consideration when selecting reference chemicals for the development, evaluation and/or validation of alternative methods and/or strategies for serious eye damage/eye irritation testing.

20. In November (Nov 3, 2020) a teleconference was held with a subgroup of the Expert Group on Skin and Eye irritation to discuss the issue regarding the Modes of Action (MoA). It was concluded that the MoA are unknowable for the majority of the chemicals and most test substances would fall into multiple chemical classes. As such an analysis based on the MoA will not provide additional insight in explaining the performance of test methods and defined approaches and it is not possible to assess whether the DAs cover all relevant mode of actions. In addition to ensuring that the key *in vivo* drivers of classification have been covered by the selected reference chemicals, analysis of the OFGs present across the reference test chemicals show that a wide range of functionality has been covered over the UN GHS Cat. 1, Cat. 2 and No Cat. classified chemicals.

21. In conclusion, the assessment of the performance of the individual test methods and the DAS against the Draize eye test has been conducted based on reference chemicals selected according to key criteria, as defined by Barroso and co-workers (2017), such that the important drivers of classification for each UN GHS category and a wide range of organic functional groups are represented.

3.3. 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.

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 91:521–547.

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, 123-134. doi:10.14573/altex.1510053.

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

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 BL, Meloni M, McNamee P, Osborne R, Pape W, Pfannenbecker U, Prinsen M, Seaman C, Spielmann H, Stokes W, Trouba K, Berghe CV, Goethem FV, 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. doi:10.1016/j.tiv.2009.05.019

4. Evaluation of the Draize eye test uncertainty and reproducibility

22. The Draize eye test is the *in vivo* animal reference test used for benchmarking the predictive performance of serious eye damage/eye irritation DA.

23. This document reports an assessment of the Draize eye test reproducibility that was based on two published comprehensive analyses (Adriaens et al. 2014; Barroso et al. 2017) on the inherent variability of the Draize eye test. The variability of the animal data has to be considered in the evaluation of the uncertainties when comparing DAs' predictions to the benchmark animal predictions.

4.1. Within-test variability

24. The impact of the uncertainty of *in vivo* reference data on the evaluation/validation of alternative methods was illustrated by the resampling analysis (within-test variability using individual rabbit data) presented by Adriaens et al. (2014). In total, 2089 studies were used for this analysis.

25. The resampling probabilities were estimated based on the individual rabbit data. Only studies with individual data on at least three rabbits were taken into account. In the resampling approach used in this study, simulated chemicals were created by randomly grouping together three animals that may have been tested with different chemicals.

- First, the different studies were pooled according to UN GHS classification of the tested chemicals. In this way, it was assured that the rabbits used in the various resampling always came from studies with chemicals classified with the same UN GHS category (i.e. No Cat., Cat. 2, or Cat. 1).
- Next, separate resampling analyses were then performed on each of the three individual data pools (the pool of studies within each UN GHS category). Data on 10,000 simulated chemicals were generated, i.e. a random sample of three rabbits was drawn 10,000 times from the data pool without replacement. This means that each animal entered a simulated chemical only once.
- Finally, the UN GHS classification criteria were applied for these simulated chemicals and the predictive capacity (correct classification) was calculated by comparing the theoretical classification (resulting from the resampling approach) with the observed classification.

26. This analysis strongly suggests an over-predictive power of the Draize eye test. The resampling analyses based on the simulated chemicals demonstrated an overall probability of

- at least **8%** of solids classified as Cat. 1 by the Draize eye test could be equally identified as Cat. 2 and none of them were identified as No Cat.
- about **13%** of Cat. 2 solids could be equally identified as No Cat.
- the over-classification error for No Cat. and Cat. 2 solids was negligible **(<1 %)**

4.2. Between-test variability

27. Cosmetics Europe has compiled a database of Draize data (Draize eye test Reference Database, DRD) from external lists that were created to support past validation activities (Barroso et al. 2017). This database contains 681 independent *in vivo* studies on 634 individual chemicals representing a wide range of chemical classes.

28. For the purpose of this document, an evaluation of the Draize eye test between-test variability was considered. Such analyses were based on solids for which more than one independent study was performed by different laboratories. However, one must take into account the low number of repeat studies and therefore, generalization of the reproducibility is not possible.

- The reproducibility of the repeat studies (set of 6 solids) evaluated in terms of agreement of classifications showed a 100% concordance outcome of the repeat studies (5 Cat. 1 solids and 1 No Cat. solid).

4.3. Performance metrics

29. The performance of the DAS was assessed against the performance metrics that were agreed by the OECD experts for the DA's for non-surfactant liquids (OECD SD 354, 2022). The values are reported in **Table 4-1**.

Table 4-1. Performance metrics for assessment of the predictivity of a DA of non-surfactant liquid test substances for eye hazard identification

UN GHS	Defined Approach		
	Cat. 1	Cat. 2	No Cat.
Cat. 1	≥ 75%	≤ 25%	≤ 5%
Cat. 2	≤ 30%	≥ 50%	≤ 30%
No Cat.	≤ 5%	≤ 30%	≥ 70%

4.4. 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.

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

525 OECD (2022), GD No. 354: Supporting Document for Evaluation and Review of Test
526 Guideline 467 on Defined Approaches (DAs) for Serious Eye Damage / Eye Irritation,
527 OECD Environment, Health and Safety Publications, Series on testing and assessment,
528 OECD Publishing, Paris. Available at
529 [https://www.oecd.org/chemicalsafety/testing/seriesontestingandassessmenttestingforhu
530 manhealth.htm]

531

5. Analyses of the DAS performance

30. Chapter 5 of this document includes information on the DAS that was supported by the OECD Expert Group to be considered in their programme and was presented at the OECD Expert Group on Eye/Skin Irritation/Corrosion & Phototoxicity (2023). The information in this chapter was organised according to the evaluation framework proposed for the Defined Approaches for liquids (DAL) for eye hazard identification (OECD, 2022). Further, chapter 3 provides details on the criteria applied for the selection of the reference chemicals used to assess the DAS performance, and as agreed during the OECD Expert Group on Eye/Skin Irritation/Corrosion and Phototoxicity meeting of 2019.

5.1. DAS

5.1.1. Development of the DIP

31. The DAS was developed based on the results of 71 neat solids (training set) that were available for the different components of the DAS.

32. In a next step, the performance of the DAS was assessed for the test set. No changes were made to the DIP after assessing the performance of the test set since no further improvement to the DIP was possible based on the performance of the training and test set results. The identification of the substances that were used in the training set and the test set is available in Annex A (spreadsheet Annex_A.2) of the current background review document. The distribution of the solids by UN GHS category and chemical set is provided in **Table 5-1**.

Table 5-1. Distribution of the reference chemicals: number of chemicals tested

UN GHS	Training set	Test set	Total
Cat. 1	20	11	31
Cat. 2	12	6	18
No Cat.	39	21	60
Total	71	38	**109**

33. The full set of substances evaluated with the DAS (total 109 different substances) is reported in Annex A. Summary statistics describing the chemical space of the chemicals tested using the DAS are provided in **Table 5-2**.

Table 5-2. Summary of the physicochemical property ranges that describe the chemical space of the chemicals tested using DAS

UN GHS	**MW**	**Melting point** (C°)	**Water solubility** (mg/mL)	**LogP**	**Vapour Pressure** (mmHg)
	Min – Max	**Min – Max**	**Min – Max**	**Min – Max**	**Min – Max**
Cat. 1	68.1 – 985.1	43 – 237	0.001 – 578	-1.79 – 8.0	0 – 2.96
Cat. 2	80.0 – 480.4	46 – 233	< 0.001 – 1000	-1.56 – 4.64	0 – 2.53
No Cat.	78.0 – 823.1	42 – 357	< 0.001 – 589	-3.36 – 9.12	0 – 1.91
Overall	**68.1 – 985.1**	**42 – 357**	**< 0.001 – 1000**	**-3.36 – 9.12**	**0 – 2.96**

34. For the set of 109 substances, high quality Draize eye test data were described in Supplementary Material 1 of the Cosmetics Europe Draize eye test Reference Database (Barroso et al., 2017).

5.1.2. Predictive capacity for the overall set

35. The predictive performance considering the three UN GHS categories (Cat. 1, Cat. 2, No Cat.) of DAS is reported in **Table 5-3**.

Table 5-3. Performance of the DAS based on SkinEthic™ HCE EITS and BCOP LLBO (N = 109 solids)

UN GHS	DAS		
	Cat 1	Cat 2	No Cat
Cat. 1 (N=31), %[a] (n/N)	*≥ 75%*[b] **77.4% (24.0/31.0)**	*≤ 25%*[b] 22.6% (7.0/31.0)	*≤ 5%*[b] 0.0% (0.0/31.0)
Cat. 2 (N=18), %[a] (n/N)	*≤ 30%*[b] 29.5% (5.3/18.0)	*≥ 50%*[b] **52.3% (9.4/18.0)**	*≤ 30%*[b] 18.2% (3.3/18.0)
No Cat. (N=60), %[a] (n/N)	*≤ 5%*[b] 1.7% (1.0/60.0)	*≤ 30%*[b] 28.3% (17.0/60.0)	*≥ 70%*[b] **70.0% (42.0/60.0)**

66.7 % balanced accuracy

[a] The proportion given is based on a weighted calculation which takes into account (where they exist) multiple results from an individual information source for a given chemical, and applying a correction factor so that all chemicals have a weight of 1. To improve the readability of the numbers in the table, the numbers n/N have been rounded, so they may deviate slightly from the percentage corresponding to the weighted calculation.

[b] Performance metrics for assessment of the predictivity of a DA of non-surfactant liquid test substances for eye hazard identification

5.1.3. Limitations of individual sources of information

36. The strengths and limitations on individual test methods are described in the corresponding OECD Test Guidelines (OECD TG 437 and TG 492, 2023a and 2023b).

37. It is important to separate these limitations into:

- technical limitations
- limitations in the predictivity for UN GHS categories

38. The technical limitations may make a chemical not testable in one or more component methods of DAS and may thus limit its applicability domain.

39. The predictivity limitations of some individual test methods for UN GHS categories do not necessarily limit the predictivity of an overarching DA; one of the advantages of DAs is that they are designed to overcome predictivity limitations of single test methods, i.e. the DAs can predict Cat. 2.

5.2. References

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, 521–547.

OECD (2022), GD No. 354: Supporting Document for Evaluation and Review of Test Guideline 467 on Defined Approaches (DAs) for Serious Eye Damage / Eye Irritation, OECD Environment, Health and Safety Publications, Series on testing and assessment, OECD Publishing, Paris. Available at [https://www.oecd.org/chemicalsafety/testing/seriesontestingandassessmenttestingforhumanhealth.htm]

OECD (2023a), Test No. 492: Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264242548-en.

OECD (2023b), Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264203846-en.

6. Analyses of the DAS uncertainty and reproducibility

40. The objective of this analysis is to evaluate the uncertainty associated with the performance of the individual test methods (SkinEthic™ HCE EITS and BCOP LLBO) and the DAS. The aim was to assess the reproducibility of each information source, and how that propagates to the DAS overall.

41. The evaluation is based on 46 chemicals (**Table 6-1**) for which multiple results are available for each test method and are therefore suitable for reproducibility analysis. The reference benchmark is the UN GHS classification based on the Draize eye test (UN GHS column in **Table 6-1**).

42. A recognised method for incorporating uncertainty assessment into performance evaluation is also to apply a bootstrap approach. Bootstrapping is a resampling procedure by which a single dataset is randomly resampled over a high number of times. Each random resample is obtained from the original dataset (i.e., resampling with replacement) creating many simulated samples. Here, bootstrapping was used to produce a distribution of DA predictions based on 100,000 replicates and compared to benchmark reference classification (Draize eye reference data).

43. Example: Multiple predictions obtained by individual methods for m-dinitrobenzene (No. 32) are reported in **Table 6-1**. For SkinEthic™ HCE EITS there are 11 classifications available based on existing data, i.e. 1 positive (Cat. 1/ Cat. 2) and 10 negatives (No Cat.). Bootstrap allows generation of "new" data either Cat. 1/Cat. 2 or No Cat. The probability of generating Cat. 1/Cat. 2 is proportional to the occurrence of Cat. 1/Cat. 2 in existing data, i.e. 1/11. Therefore, bootstrap can be used to generate an arbitrarily large number of "new" classifications where the frequency of Cat. 1/Cat. 2 will be 1/11 (9.1%) whereas No Cat. 10/11 (90.9%) (similar to weighted calculation approach).

44. Bootstrap is used to generate a full matrix of classifications for two single methods by resampling the data for all chemicals (N=46). Resampling is repeated 100,000 times and resulting performance measures are averaged across 100,000 bootstrap replicates.

The resulting performance values are shown in **Table 6-2** and are based on predictions reported in **Table 6-1**. In addition, the performance values for the set of solids that were tested during the SkinEthic™ HCE EITS validation study and the BCOP LLBO evaluation study are reported in

Table 6-3.

Table 6-1. Prediction for the individual test methods (proportion of correct predictions, TRUE pred. %). DAS predictions are derived by applying the data interpretation procedure (DIP) to predictions from a single method. [TRUE pred., proportion of correctly predicted results: SkinEthic™ HCE EITS = No Cat. versus Cat. 1 + Cat. 2 and BCOP LLBO = Cat. 1 versus Cat. 2 + No Cat.]. The last column corresponds with the proportions of correct predictions within each UN GHS Category (Cat. 1, Cat. 2 and No Cat.) for the DAS.

				SINGLE METHODS						DAS
	Chemicals	**CAS#**	**UN GHS**	**SkinEthic™ HCE EITS**			**BCOP LLBO**			
				Cat. 1 + Cat. 2	**No Cat.**	**TRUE pred %**	**Cat. 1**	**Cat. 2 + No Cat.**	**TRUE pred %**	**TRUE pred %**
1	2-Benzyl-4-chlorophenol	120-32-1	Cat 1	11	0	100	2	0	100	100
2	2-Hydroxy iso-butyric acid	594-61-6	Cat 1	2	0	100	2	0	100	100
3	4,4'-(4,5,6,7-Tetrabromo-3H-2,1-benzoxathiol-3-ylidene)bis[2,6-dibromophenol] S,S-dioxide	4430-25-5	Cat 1	11	0	100	2	0	100	100
4	4-(1,1,3,3-Tetramethylbutyl)phenol	140-66-9	Cat 1	11	0	100	0	2	0	0
5	alpha-Ketoglutaric acid	328-50-7	Cat 1	11	0	100	2	0	100	100
6	Dibenzoyl-L-tartaric acid	2743-38-6	Cat 1	2	0	100	5	0	100	100
7	1-Naphthalene acetic acid Na salt	61-31-4	Cat 1	2	0	100	2	0	100	100
8	Captan 90-concentrate	133-06-2	Cat 1	2	0	100	0	2	0	0
9	Lauric acid	143-07-7	Cat 1	11	0	100	2	0	100	100
10	m-Phenylene diamine	108-45-2	Cat 1	2	0	100	1	1	50	50
11	p-tert-Butylphenol	98-54-4	Cat 1	2	0	100	2	0	100	100
12	Sodium perborate tetrahydrate	10486-00-7	Cat 1	2	0	100	2	0	100	100
13	Sodium salicylate	54-21-7	Cat 1	2	0	100	1	1	50	50
14	Benzoic acid	65-85-0	Cat 1	11	0	100	2	0	100	100
15	1,2-Benzisothiazol-3(2H)-one	2634-33-5	Cat 1	11	0	100	2	0	100	100
16	Chlorhexidine	55-56-1	Cat 1	9	0	100	3	0	100	100
17	Paraformaldehyde	30525-89-4	Cat 1	2	0	100	0	2	0	0
18	N-(2-Methylphenyl)-iminodicarbonimidic diamide (1-(o-Tolyl)biguanide)	93-69-6	Cat 1	2	0	100	2	0	100	100
19	N-Acetyl-DL-methionine	1115-47-5	Cat 1	11	0	100	2	0	100	100
20	Triethanolamine orthovanadate	13476-99-8	Cat 1	2	0	100	0	2	0	0
21	(2R,3R)-3-((R)-1-(Tert-butyldimethylsiloxy)ethyl)-4-oxoazetidin-2-yl acetate	76855-69-1	Cat 2	0	11	0	0	2	100	0

22	3,3'-Dithiopropionic acid	1119-62-6	Cat 2	11	0	100	0	2	100	100
23	4-Carboxybenzaldehyde	619-66-9	Cat 2	11	0	100	2	0	0	0
24	Dibenzyl phosphate	1623-08-1	Cat 2	11	0	100	2	0	0	0
25	1,5-Naphthalenediol	83-56-7	Cat 2	7	4	63.6	1	1	50	31.8
26	2-Amino-3-hydroxy pyridine	16867-03-1	Cat 2	11	0	100	0	2	100	100
27	Ammonium nitrate	6484-52-2	Cat 2	11	0	100	0	2	100	100
28	Sodium benzoate	532-32-1	Cat 2	11	0	100	0	2	100	100
29	2,6-Dichloro-5-fluoro-beta-oxo-3-pyridinepropanoate	96568-04-6	Cat 2	5	0	100	0	2	100	100
30	2-Hydroxy-1,4-naphthoquinone	83-72-7	Cat 2	11	0	100	2	0	0	0
31	Camphene	79-92-5	Cat 2	11	0	100	0	2	100	100
32	m-Dinitrobenzene	99-65-0	Cat 2	1	10	9.1	0	2	100	9.1
33	p-Nitrobenzoic acid	62-23-7	Cat 2	11	0	100	0	2	100	100
34	Sodium monochloroacetate	3926-62-3	Cat 2	11	0	100	0	2	100	100
35	Ethylenediaminetetraacetic acid dipotassium salt	25102-12-9	No Cat	3	0	0	0	3	100	0
36	Propyl-4-hydroxybenzoate	94-13-3	No Cat	9	0	0	0	4	100	0
37	1-(9H-Carbazol-4-yloxy)-3-[[2-(2-methoxyphenoxy)ethyl]amino]propan-2-ol	72956-09-3	No Cat	0	10	100	0	2	100	100
38	2,5,6-Triamino-4-pyrimidinol sulphate	1603-02-7	No Cat	4	5	55.6	0	2	100	55.6
39	2-Mercaptopyrimidine	1450-85-7	No Cat	0	11	100	0	7	100	100
40	Anthracene	120-12-7	No Cat	0	9	100	0	2	100	100
41	Phenothiazine	92-84-2	No Cat	0	11	100	0	2	100	100
42	Phenylbutazone	50-33-9	No Cat	0	9	100	0	3	100	100
43	Potassium tetrafluoroborate	14075-53-7	No Cat	0	11	100	0	2	100	100
44	Silicic acid (neat)	1343-98-2	No Cat	0	11	100	0	2	100	100
45	Tetrabromobisphenol A	79-94-7	No Cat	0	2	100	0	2	100	100
46	Theobromine	83-67-0	No Cat	0	11	100	0	2	100	100

Table 6-2. Performance measures based on 100,000 Bootstrap replicates, of individual methods and DAS against UN GHS classifications (individual data **Table 6-1**)

	Reproducibility	Accuracy	Specificity	Sensitivity	
12 No Cat. / 34 Cat. 1 + Cat. 2					
SkinEthic™ HCE EITS	98.1%	91.5%	79.7%	95.7%	
20 Cat. 1 / 26 Cat. 2 + No Cat					
BCOP LLBO	96.7%	81.5%	86.5%	76.2%	
UN GHS (N=46)		Accuracy	TRUE No Cat. (N=12)	TRUE Cat. 2 (N=14)	TRUE Cat. 1 (N=20)
DAS	93.5%	71.8%	79.7%	60.5%	75.0%

Table 6-3. Performance values from the validation studies (solids only)

	Reproducibility WLR/BLR	Accuracy	Specificity	Sensitivity
No Cat. / Cat. 1 + Cat. 2	Mean values 3 independent runs in 3 laboratories			
SkinEthic™ HCE EITS [a]	95.0% / 96.7% (N=60)	81.0% (N=95)	73.6% (N=43)	90.5% (N=52)
Cat. 1 / Cat. 2 + No Cat.	Mean value 2 independent runs			
BCOP LLBO [b]	92.5% / NA (40/NA)	79.7% (N=64)	85.9% (N=39)	70.0% (N=25)

[a] Alépée et al., 2016 ; [b] Adriaens et al., 2020

6.1. References

Adriaens E, Verstraelen S, Desprez B, Alépée N, Abo T, Bagley D, Hibatallah J, Mewes KR, Pfannenbecker U, Van Rompay AR (2020). Overall performance of Bovine Corneal Opacity and Permeability (BCOP) Laser Light-Based Opacitometer (LLBO) test method with regard to solid and liquid chemicals testing. Toxicol. in Vitro 70, 105-044

Alépée N, Leblanc V, Adriaens E, Grandidier MH, Lelièvre D, Meloni M, Nardelli L, Roper CS, Santirocco E, Toner F, Van Rompay AR, Vinall J, Cotovio J (2016). Multi-laboratory validation of SkinEthic HCE test method for testing serious eye damage/eye irritation using liquid chemicals. Toxicol. in Vitro 31, 43-53. http://dx.doi.org/10.1016/j.tiv.2015.11.012

7. Detailed performance analysis of individual methods and the DAS against Draize eye test

45. This chapter analyses the performance of the individual methods (SkinEthic™ HCE EITS and BCOP LLBO) and that of DAS, against the curated Draize Eye test reference data. The following methods and DAs were analysed:

- SkinEthic™ HCE EITS
- BCOP LLBO
- DAS

46. The performance of these methods with respect to the whole dataset, by driver of classification, and by chemical class, is presented in the next chapters, with a specific focus on mispredictions. The analyses are meant to provide considerations and support recommendations regarding the use of the DAS based on the performance observed in this dataset. More details regarding the drivers of classification and the OFG are provided in section 3.2. (Key criteria for evaluation of the DAS versus the *in vivo* Draize eye test) and section 7.3 (Analysis of the performance for specific Organic Functional Groups (OFG) with DAS).

7.1. All chemicals

47. The full set of substances evaluated with DAS (total 109 different substances) is reported in Annex A (spreadsheet Annex_A.2).

48. The prevalence of *in vivo* classified solids (i.e., UN GHS Cat. 1 and Cat. 2) is 45.0% (49/109).

49. The performance of the SkinEthic™ HCE EITS for identifying chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Cat.) is shown in Table 7-1. The method has an accuracy of 80.5%, with 70.0% specificity and 93.3% sensitivity.

Table 7-1 Predictive Capacity of the SkinEthic™ HCE EITS test method for identifying chemicals not requiring classification for eye irritation or serious eye damage [UN GHS No Cat. versus Not No Cat. (Cat. 1 + Cat. 2)]

	Accuracy (Balanced accuracy)	UN GHS Cat. 1 + Cat. 2	UN GHS No Cat.

	N	% [a]	N	Sensitivity (%)	FN (%)	N	Specificity (%)	FP (%)
SkinEthic™ HCE EITS	109	80.5 (81.7)	49	93.3	6.7	60	70.0	30.0

[a] The proportion in the tables are based on weighted calculation. For each chemical, all results were taken into account and a correction factor was applied so that all chemicals had the same weight (weight of 1).

50. The performance of the BCOP LLBO test method for identifying chemicals inducing serious eye damage (UN GHS Cat. 1) is shown in Table 7-2.

Table 7-2 Predictive Capacity of BCOP LLBO test method for identifying chemicals inducing serious eye damage [UN GHS Cat. 1 versus Not Cat. 1 (Cat. 2 + No Cat.)]

	Accuracy (Balanced accuracy)		**UN GHS Cat. 1**			**UN GHS Cat. 2 + No Cat.**		
	N	% [a]	N	Sensitivity (%)	FN (%)	N	Specificity (%)	FP (%)
BCOP LLBO	103	85.0 (82.8)	31	77.4	22.6	72	88.2	11.8

[a] The proportion in the tables are based on weighted calculation. For each chemical, all results were taken into account and a correction factor was applied so that all chemicals had the same weight (weight of 1).

51. An overview of the solids which are mispredicted by the DAS are listed in Table 7-3. In *vivo* Cat. 1 solids which are under-predicted are the result of an under-prediction by the BCOP LLBO. *In vivo* Cat. 2 solids which are over-predicted are the result of an over-prediction by the BCOP LLBO. False negative *in vivo* Cat. 2 solids and false positive *in vivo* No Cat. solids are the result of a misprediction by SkinEthic™ HCE EITS.

Table 7-3. Mis-predicted solids in comparison with UN GHS categories

DRD No.	Chemical	CASRN	UN GHS	Driver	DAS prediction
23	4-(1,1,3,3-Tetramethylbutyl)phenol	140-66-9	Cat 1	CO mean ≥ 3	UP (1)
86	3,4-Dichlorophenyl isocyanate	102-36-3	Cat 1	CO pers D21	UP (1)
88	Captan 90-concentrate	133-06-2	Cat 1	CO pers D21	UP (1)
91	m-Phenylene diamine	108-45-2	Cat 1	CO pers D21	TP/UP (0.50/0.50)
101	Sodium salicylate	54-21-7	Cat 1	CO pers D21	TP/UP (0.50/0.50)
148	Chlorophenacyl	6305-04-0	Cat 1	CO = 4	UP (1)
149	Paraformaldehyde	30525-89-4	Cat 1	CO = 4	UP (1)
156	Triethanolamine orthovanadate	13476-99-8	Cat 1	CO = 4	UP (1)
190	(2R,3R)-3-((R)-1-(Tert-butyldimethylsiloxy)ethyl)-4-oxoazetidin-2-yl acetate	76855-69-1	Cat 2	CO mean ≥ 1	FN (1)
192	4-Carboxybenzaldehyde	619-66-9	Cat 2	CO mean ≥ 1	OP (1)
194	Dibenzyl phosphate	1623-08-1	Cat 2	CO mean ≥ 1	OP (1)
211	1,3-bis-(2,4-Diaminophenoxy) propane tetrachloride	74918-21-1	Cat 2	Conj mean ≥ 2	OP (1)

212	1,5-Naphthalenediol	83-56-7	Cat 2	Conj mean ≥ 2	OP/TP/FN (0.32/0.32/ 0.64)
214	4-Amino-3-nitrophenol	610-81-1	Cat 2	CO mean ≥ 1	OP (1)
238	1,4-Dibutoxybenzene	104-36-9	Cat 2	Conj mean ≥ 2	FN (1)
239	2-Hydroxy-1,4-naphthoquinone	83-72-7	Cat 2	Conj mean ≥ 2	OP (1)
241	m-Dinitrobenzene	99-65-0	Cat 2	Conj mean ≥ 2	TP/FN (0.09/0.91)
271	Sodium bisulphite	7631-90-5	No Cat	CO > 0 **	FP (1)
302	1-Phenyl-3-pyrazolidone	92-43-3	No Cat	CO > 0	FP (1)
312	DL-Glutamic acid	19285-83-7	No Cat	CO > 0	FP (1)
313	Ethylenediaminetetraacetic acid dipotassium salt	25102-12-9	No Cat	CO > 0	FP (1)
316	Propyl-4-hydroxybenzoate	94-13-3	No Cat	CO > 0	FP (1)
324	N,N-Dimethyl guanidine sulphate	598-65-2	No Cat	CO = 0 **	FP (1)
504	1,5-Di(2,4-dimethylphenyl)-3-methyl-1,3,5-triazapenta-1,4-diene	33089-61-1	No Cat	CO = 0	FP/TN (0.56/0.44)
506	1H-Indole-2,3-dione	91-56-5	No Cat	CO = 0	FP (1)
516	2,5,6-Triamino-4-pyrimidinol sulphate	1603-02-7	No Cat	CO = 0	FP/TN (0.44/0.56)
517	2,6-Dihydroxy-3,4-dimethylpyridine	84540-47-6	No Cat	CO = 0	FP (1)
525	3,4-Dimethoxybenzaldehyde	120-14-9	No Cat	CO = 0	FP (Cat 1) (1)
526	3,5-Dihydroxyacetophenone	51863-60-6	No Cat	CO = 0	FP (1)
558	Gluconolactone	90-80-2	No Cat	CO = 0	FP (1)
560	Hexamethylenetetraamine	100-97-0	No Cat	CO = 0	FP (1)
562	Methyl p-hydroxybenzoate	99-76-3	No Cat	CO = 0	FP (1)
573	Sodium tripolyphosphate	7758-29-4	No Cat	CO = 0	FP (1)
577	Theophylline sodium acetate	8002-89-9	No Cat	CO = 0	FP (1)
580	Trisodium mono-(5-(1,2-dihydroxyethyl)-4-oxido-2-oxo-2,5-dihydro-furan-3-yl) phosphate	66170-10-3	No Cat	CO = 0	FP (1)
581	Xanthinol nicotinate	437-74-1	No Cat	CO = 0	FP (1)

TN: True Negatives; TP: True Positives; FN: False Negatives (*in vivo* Cat. 2 predicted as No Cat.); UP: Under-Predicted chemicals (*in vivo* Cat. 1 predicted as Cat. 2); FP: False Positives (*in vivo* No Cat. predicted as Cat. 2); OP: Over-Predicted chemicals (*in vivo* Cat. 2 predicted as Cat. 1); the number in parentheses corresponds with the proportion of the prediction, for some chemicals two fractions are provided, this is because the predictions differ between the *in vitro* study result available.

CO > 0: in at least one observation time in at least one animal and all animals showing mean scores of days 1–3 below the classification cut-offs for all endpoints, ** Indicates at least one animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint (see $13).

7.2. Analysis of the performance by driver of classification

52. This section focuses on the performance of the individual test methods and DAS by driver of classification. Details on the driver of classification are presented in Chapter 3. The results of the individual test methods are presented in tables summarising the TP (sensitivity), TN (specificity) and accuracy as compared to the Draize eye test benchmark

data. The results of the DAS are presented in a table summarising TP (true Cat. 1 and true Cat. 2), TN (true No Cat.), over-predictions (OP, *in vivo* Cat. 2 predicted as Cat. 1), under-predictions (UP, *in vivo* Cat. 1 predicted as Cat. 2), FN (*in vivo* Cat. 1 or Cat. 2 predicted as No Cat.) and accuracy as compared to the Draize eye test benchmark data.

53. The SkinEthic™ HCE EITS test method can be used to identify chemicals that do not require classification for eye irritation or serious eye damage with DAS. The TN rate per No Cat. subgroup is shown in Table 7-4. No Cat. solids from the subgroup CO > 0 and CO > 0 ** resulted in 50.0-55.6% FPs. A lower FP rate (25%) was observed for the subgroup CO = 0. FN results were not observed for the most important drivers of Cat. 1 classification.

Table 7-4 Predictive Capacity of SkinEthic™ HCE EITS for identifying chemicals not requiring classification for eye irritation or serious eye damage [UN GHS No Cat. (True Negative, TN) versus Not No Cat. (Cat. 1 + Cat. 2 = True Positive, TP)]

Parameter	UN GHS Driver of classification	SkinEthic™ HCE EITS N	SkinEthic™ HCE EITS Correct prediction [a]
TP	**Cat. 1**	**31**	100
	CO mean ≥ 3	9	100
	CO pers D21	11	100
	CO = 4	11	100
	Cat. 2	**18**	81.8
	CO mean ≥ 1	7	85.7
	Conj mean ≥ 2	11	79.3
TN	**No Cat.**	**60**	70.0
	CO > 0 **	2	50.0
	CO > 0	9	55.6
	CO = 0 **	1	0
	CO = 0	48	75.0
Accuracy		**109**	80.5

[a] The proportion in the tables are based on weighted calculation. For each chemical, all results were taken into account and a correction factor was applied so that all chemicals had the same weight (weight of 1).

** Indicates at least one animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint.

54. The BCOP LLBO is used to identify chemicals requiring classification for serious eye damage with DAS. The TP rate and TN rate by driver of classification is shown in Table 7-5. The FN rate for the BCOP LLBO was low for the driver CO mean ≥ 3 (FN = 11.1%) and was 27.3% for the drivers CO = 4 and CO pers D21. The FP rate for Cat. 2 was 42.8% for CO mean ≥ 1 and 22.7% for Conj mean ≥ 2.

Table 7-5 Predictive Capacity of individual *in vitro test* methods for identifying chemicals inducing serious eye damage [UN GHS Cat. 1 (True Positive, TP) versus Not Cat. 1 (Cat. 2 + No Cat. = True Negative, TN)]

Parameter	UN GHS Driver of classification	BCOP LLBO	
		N	Correct prediction [a]
TP	**Cat 1**	**31**	77.4
	CO mean ≥ 3	9	88.9
	CO pers D21	11	72.7
	CO = 4	11	72.7
TN (Not Cat. 1)	**Cat 2**	**18**	69.4
	CO mean ≥ 1	7	57.2
	Conj mean ≥ 2	11	77.3
	No Cat	**54**	94.4
	CO > 0 **	2	100
	CO > 0	8	87.5
	CO = 0 **	1	100
	CO = 0	43	95.3
Accuracy		**103**	85.5

[a] The proportion in the tables are based on weighted calculation. For each chemical, all results were taken into account and a correction factor was applied so that all chemicals had the same weight (weight of 1).

** Indicates at least one animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint.

55. The performance by driver of classification (Cat. 1 and Cat. 2) or by subgroup (No Cat.) with the DAS is shown in **Table 7-6**. The UP rate for the Cat. 1 driver of classification CO mean ≥ 3 was low (11.1%). The UP for CO=4 and CO pers D21 was 27.3%. Cat. 2 solids that were classified based on CO mean ≥ 1 resulted in a higher over-prediction rate in comparison with those that were classified based on Conj mean ≥ 2. The FP rate for solids from the subgroup CO = 0 was lower in comparison to the FP rates that were observed for the subgroups CO > 0 ** and CO > 0.

Table 7-6 Predictive performance considering the three UN GHS categories (Cat. 1, Cat. 2, No Cat.) of DAS

Cat. 1	CO mean ≥ 3	CO pers D21	CO=4	**All**
N	**9**	**11**	**11**	**31**
TP (%) [a]	88.9	72.7	72.7	**77.4**
UP (%) [a]	11.1	27.3	27.3	**22.6**
FN (%) [a]	0	0	0	**0**

Cat. 2	CO mean ≥ 1	Conj mean ≥ 2	**All**
N	**7**	**11**	**18**
OP (%) [a]	42.9	21.1	**29.5**
TP (%) [a]	42.9	58.3	**52.3**
FN (%) [a]	14.3	20.7	**18.2**

No Cat.	CO > 0 **	CO > 0	CO = 0 **	CO = 0	**All**
N	**2**	**9**	**1**	**48**	**60**
FP (%) [a]	50.0	44.4	100	25.0	**30.0**
TN (%) [a]	50.0	55.6	0	75.0	**70.0**

[a] The proportion in the tables are based on weighted calculation. For each chemical, all results were taken into account and a correction factor was applied so that all chemicals had the same weight (weight of 1). To improve the readability of the numbers in the table, the numbers n/N have been rounded, so they may deviate slightly from the percentage corresponding to the weighted calculation.

** Indicates at least one animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint.

7.3. Analysis of the performance for specific Organic Functional Groups (OFG) with DAS

56. This section focuses on the performance of the DAS by organic functional group. The results of the DAS are presented in tables summarising TP (true Cat. 1 and true Cat. 2), TN (true No Cat.), over-predictions (OP, *in vivo* Cat. 2 predicted as Cat. 1), under-predictions (UP, *in vivo* Cat. 1 predicted as Cat. 2), FN (*in vivo* Cat. 1 or Cat. 2 predicted as No Cat.) and accuracy as compared to the Draize eye test benchmark data.

57. Only performance metrics for the most frequent OFG's, being at least 5 chemicals per allocated OFG are discussed. The set of 109 solids contained 12 inorganic compounds. The distribution according to the UN GHS category is shown in **Table 7-7**.

Table 7-7 Number of solids with a specific OFG according to the UN GHS category

OFG	% of total N (=97)	n	UN GHS (n)		
			Cat. 1	Cat. 2	No Cat.
Aryl	32.0	31	6	4	21
Carboxylic acid	18.6	18	8	5	5
Phenol	17.5	17	7	2	8
Aryl halide	16.5	16	5	1	10
Ether	10.3	10	0	2	8
Carboxylic acid ester	9.3	9	1	1	7
Alcohol	7.2	7	2	0	5
Benzyl	7.2	7	2	1	4
Fused carbocyclic aromatic	7.2	7	3	1	3
Ketone	5.2	5	2	1	2
tert-Butyl	5.2	5	1	1	3
Aromatic amine	5.2	5	0	0	5

Note that a single chemical may have more than one organic functional group

58. The performance of the DAS by OFG is shown in **Table 7-8**. Only OFGs for which at least 5 solids were evaluated for a specific UN GHS category are discussed. Solids with an aryl, carboxylic acid, phenol or aryl halide are the only *in vivo* Cat. 1 chemicals with at least 5 solids. The TP rate for Cat. 1 solids with an aryl function (N=6) was 100% and for the other OFGs (carboxylic acid, phenol or aryl halide) the majority (> 75%) of the solids was predicted correctly. Solids with a carboxylic acid are the only *in vivo* Cat. 2 chemicals with at least 5 solids being tested with a TP rate of 80% (N=5). The FP rate was 60% for carboxylic acids (N=5) and 37.5% for phenols OFG (N=8). For all other solids with at least 5 UN GHS No Cat. chemicals the FP rate was ≤ 20%. About 81% of the *in vivo* No Cat. solids with an aryl group (N=21) were correctly predicted.

59. Overall, the number of substances per OFG with results for DAS is limited and therefore it is not possible to draw conclusion on the applicability domain of the DAS.

Table 7-8 Predictive performance considering the three UN GHS categories (Cat. 1, Cat. 2, No Cat.) of DAS with SkinEthic™ HCE EITS and BCOP LLBO

UN GHS	Predicted class	Aryl	Carboxylic acid	Phenol	Aryl halide	Ether	Carboxylic acid ester
		N = 31	N = 18	N = 17	N = 16	N = 10	N = 9
Cat. 1	**Cat. 1**	6.0/6	7.5/8	5.5/7	4.0/5	NA	1.0/1
	Cat. 2	0.0/6	0.5/8	1.5/7	1.0/5	NA	0.0/1
	No Cat.	0.0/6	0.0/8	0.0/7	0.0/5	NA	0.0/1
Cat. 2	Cat. 1	1.0/4	1.0/5	1.3/2	0.0/1	1.0/2	0.0/1
	Cat. 2	2.0/4	4.0/5	0.3/2	1.0/1	0.0/2	1.0/1
	No Cat.	1.0/4	0.0/5	0.4/2	0.0/1	1.0/2	0.0/1
No Cat.	Cat. 1/Cat. 2	4.0/21	3.0/5	3.0/8	0.0/10	1.0/8	2.0/7
	No Cat.	17.0/21	2.0/5	5.0/8	10.0/10	7.0/8	5.0/7

UN GHS	Predicted class	Alcohol	Benzyl	Fused carbocyclic aromatic	Ketone	tert-Butyl	Aromatic amine
		N = 7	N = 7	N = 7	N = 5	N = 5	N = 5
Cat. 1	**Cat. 1**	1.0/2	2.0/2	3.0/3	1.0/2	1.0/1	NA
	Cat. 2	1.0/2	0.0/2	0.0/3	1.0/2	0.0/1	NA
	No Cat.	0.0/2	0.0/2	0.0/3	0.0/2	0.0/1	NA
Cat. 2	Cat. 1	NA	1.0/1	1.3/2	0.0/1	0.0/1	NA
	Cat. 2	NA	0.0/1	0.3/2	1.0/1	0.0/1	NA
	No Cat.	NA	0.0/1	0.4/2	0.0/1	1.0/1	NA
No Cat.	Cat. 1/Cat. 2	1.0/5	0.0/4	0.0/3	1.0/2	0.0/3	0.0/5
	No Cat.	4.0/5	4.0/4	3.0/3	1.0/2	3.0/3	5.0/5

Annex A: Spreadsheets

Annex A.1

This Annex includes the detailed Draize eye test data that were used for DAS.

Annex A.2

Spreadsheet with the identification of the substances that were used in the training set and the test set of DAS and predictions of the individual test methods and DAS.