PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED

In Vitro Macromolecular Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage

This document contains Performance Standards (PS) for determining the reliability and relevance of similar and modified test methods for ocular hazard that are structurally and mechanistically similar to the *in vitro* macromolecular test method described in TG XXX (OECD 20xx) and the DBAlm protocol n. 157 (1), in accordance with the principles of Guidance Document No. 34 (2). These PS include Essential Test Methods Components by which to evaluate the structural, mechanistic and procedural similarity of a new similar or modified proposed test method, as well as a list of 30 Reference Chemicals by which to evaluate the minimum reproducibility and predictive capacity necessary for the test method to be considered comparable to the validated reference method (VRM), in this PS the Ocular Irritection® assay.

INTRODUCTION

1. The purpose of Performance Standards (PS) is to provide the basis by which new or modified test methods, both proprietary (*i.e.* copyrighted, trademarked, registered) and non-proprietary, can be deemed to be structurally and mechanistically similar to a Validated Reference Method (VRM) and demonstrate to have sufficient reliability and relevance for specific testing purposes, in accordance with the principles of Guidance Document No. 34 (2). The PS, based on valid and accepted test method(s), can be used to evaluate the reliability and relevance of other similar test methods (colloquially referred to as “me-too” test methods) that are based on similar scientific principles and measure or predict the same biological or toxic effect (2). On the other hand, modified test methods, which propose potential improvements to an approved test method, should also be evaluated to determine the effect of the proposed changes on the test method’s performance and the extent to which such changes affect the information available for the other components of the validation process. Depending on the number and nature of the proposed changes, the generated data and supporting documentation for those changes, they should either be subjected to the same validation process as described for a new test method or, if appropriate, to a limited assessment of reliability and relevance using the established PS (2).

2. These PS are proposed for evaluating the validity of new or modified *in vitro* macromolecular test method for identifying i) chemicals inducing serious eye damage (UN GHS Category 1) and ii) chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Category). As defined by the OECD GD 35 (2) the PS consists of: (i) Essential Test Method Components; (ii) Recommended Reference Chemicals, and; (iii) Defined Reliability and Accuracy Values that the proposed similar or modified test method should meet or exceed. The VRM used to develop the present PS is the Ocular Irritection® assay, as described in OECD TG XXX (OECD 20XX) and the DBAlm protocol n. 157 (1). Definitions are provided in Annex I.

3. Similar (me-too) or modified test methods proposed for the use(s) specified in the OECD TG XXX (X) should be evaluated to determine their reliability and relevance using Reference Chemicals (table 2) representing the full range of the OECD TG 405 *in vivo* ocular hazard, *i.e.*, serious eye damage (UN GHS Cat. 1), eye irritation (UN GHS Cat. 2) and non-classified chemicals (UN GHS No Category) (3) (4). The

proposed similar or modified test methods should have reproducibility, sensitivity, specificity and accuracy values which are comparable or better than those derived from the VRM and as described in paragraphs 18 to 26 of these PS (5). The reliability of the similar or modified test method, as well as its ability to correctly identify UN GHS No Category and UN GHS Cat. 1 chemicals, should be determined prior to its use for testing chemicals. Where possible, the classes or types of chemicals that are consistently over - or under-predicted should be defined.

ESSENTIAL TEST METHOD COMPONENTS

4. The Essential Test Method Components consist of essential structural, functional, and procedural elements of the Validated Reference Method (VRM) that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding VRM (2). The essential test method components to be considered for similar or modified test methods related to the VRM Ocular Irritection® assay are described in detail in the following paragraphs 5 to 15). For specific parameters (*e.g.*, for Table 1) or modified procedures, adequate values or procedures should be provided for the proposed similar or modified test method. These specific values or procedures may vary depending on the specific test method and/or its modification.

Test Method Components

5. The the VRM Ocular Irritection® assay for identifying i) chemicals inducing serious eye damage and ii) chemicals not requiring classification for eye irritation or serious eye damage, consists of two essential components i.e., a macromolecular matrix and a membrane disc for the controlled delivery of the test chemical to the macromolecular matrix. The macromolecular matrix serves as the target for the test chemical and is composed of a mixture of proteins, glycoproteins, carbohydrates, lipids and low molecular weight components When hydrated, the protein oligomers within the macromolecular matrix tend to self-associate and form larger fibrils that are held together by non-covalent into a highly ordered and transparent structure presumably similar to the transparent cornea.

6. Test chemicals that can cause serious eye damage/eye irritation should be able to promote protein denaturation, protein unfolding and changes in conformation which result in the disruption and disaggregation of the highly organized macromolecular matrix, and produce turbidity. Such phenomena should be quantified, by e.g. measuring the changes in light scattering. In the case of the the VRM Ocular Irritection® assay, changes in optical density at 405 nm (OD_{405}) are measured and used to determine Irritection Draize Equivalent (IDE) score for each dose of the tested chemical as described in paragraph 11 of this document and in paragraph 19 of the OECD TG XXX (OECD 20XX). The highest IDE score from a qualified test run, named the Maximal Qualified Score (MQS), is used for identification of the UN GHS ocular hazard category (3) for the tested chemical based on a prediction model described in paragraph 15.

Procedural Conditions

Applicability of the test method

7. The applicability of the macromolecular test method to specific chemical classes and physico-chemical properties should be well characterized, as well as the exposure procedures for specific test chemicals. In the case of the VRM Ocular Irritection® assay, most critical physico-chemical property is the pH and only test chemicals with $4 \leq pH \leq 9$ fall within its applicability domain. Some chemicals may cause

interference with the Ocular Irritection® test system. However, these can be identified by the quality controls and acceptance criteria inbuilt in the VRM Ocular Irritection® assay, described in detail in paragraph 20 to 21 of the OECD TG XXX (OECD, 201X) and in paragraph 14 of this document.

Reagent preparation and activation

8. The preparation of the macromolecular matrix and the necessary quality controls should be well defined. In the case of the VRM Ocular Irritection® assay, the macromolecular matrix powder is hydrated and activated with a pH lowering activating reagent to form the macromolecular matrix. The pH of the macromolecular matrix pre and post activation should fall within pre-established ranges of 7.9-8.2 and 6.4-6.7, respectively. Activator solution is also added to the blanking buffer used as a control for each test chemical dose.

Application of Test Chemicals

9. An appropriate number of doses/concentrations of the test chemical should be tested according to the pre-defined conditions. In the case of the VRM Ocular Irritection® assay, solids and non-surfactant liquids are applied over a cellulose membrane placed on top of the macromolecular matrix, and surfactants and waxy solids are applied directly to the macromolecular matrix. A series of five doses of each test chemical are applied, as specified in OECD TG XXXX and DBALM protocol (1).

10. The macromolecular matrix should be exposed to the test chemicals and concurrent controls for a well defined duration and temperature conditions. In the case of Ocular Irritection® incubation is performed at 24.0±0.5 hours at 25±1oC in an incubator.

Determination of Irritation score

11. Determination of an appropriate irritation score for use in a prediction model for identification of UN GHS Classification categories should be defined. In the case of the VRM Ocular Irritection® assay Irritection Draize Equivalent (IDE) score is determined for each tested dose/concentration based on the analysis of the OD_{405} measured for the test chemical and analysed against the standard curve established using a set of 4 calibrating chemicals with well defined ranges of OD_{405} response (paragraph 12) tested in parallel. The IDE scores are calculated by a software incorporating formulas described in the OECD TG xxx (OECD 201X). In case of the VRM OI® the highest IDE Score of five doses/concentrations obtained for the test chemical in a qualified test run (see paragraph 14), namely the Maximal Qualified Score (MQS), is selected for determination of the UN GHS ocular hazard category (3) based on pre-defined cut-off values described in Table 1.

Control Substances

12. Appropriate controls should be tested in parallel to the test chemical, and should comprise a series of calibrators and quality control chemicals. In the case of the VRM Ocular Irritection®, these encompass four calibrating chemicals (Cal_{0-3}) included in the assay kit covering the range of OD responses (Cal_0: 0.062 - 0.262; Cal_1: 0.089 – 0.315; Cal_2: 0.351 - 0.945; Cal_3: 1.277 – 2.127) for derivation of the standard curve. Furthermore, the VRM Ocular Irritection® commercial kit includes two quality control chemicals (QC) that should result in well defined ranges of IDE scores within the lower (7.2-20.8) and mid-upper (23.6-35.6) IDE range of the VRM test.

13. In case that modified or me-too macromolecular assays involve the use of a vehicle or solvent other than water with the test chemical, the vehicle or solvent should fall within the applicability domain of the macromolecular test method, and should not alter the eye irritation potential of the test chemical. When applicable, solvent (or vehicle) controls should be tested concurrently with the test chemical to demonstrate the compatibility of the solvent with the macromolecular matrix system. In the case of the VRM Ocular Irritection®, only distilled water is used as solvent for test chemicals with surfactant properties.

Study Acceptance Criteria

14. The conditions upon which the test result is determined to be acceptable or unacceptable should be clearly defined. In the case of the VRM Ocular Irritection® assay the set of acceptance criteria are analysed automatically by the software included in the assay kit and they include: (A) predefined ranges of OD for the calibrating chemicals and of IDE scores for the of the Quality controls (see paragraph 12); (B) and (C) optimal Net OD for the test chemicals in relation to a minimal value (should be > -0.015), and in relation to the Cal_2 value (Net OD for a test chemical lower than OD_{Cal2} triggers a check for interference with the proper response of the macromolecular matrix using an inhibition check solution; (D) Blank OD value higher than 1.2, is an indicator of interference by intensely coloured test chemicals, and (E) fitness of the dose response curves for the test chemicals in relation to the optimal dose response curve and typical dose response curves for known irritant and non-irritant chemicals. These criteria and their application for determination of a qualified result is described in more detail in paragraph 20 and 21 of the OECD TG xxx (OECD 20XXX).

Interpretation of Results and Prediction Model

15. The methodology for interpretation of the results and the prediction model should be clearly defined. In the case of the VRM Ocular Irritection® assay, the highest IDE score obtained in a Qualified test run, named the Maximal Qualified Score (MQS) is used to predict the ocular hazard potential of the test chemical according to the UN GHS classification system (3), based on the Prediction Model described in table 1.

Table 1. Ocular Irritection® prediction model

Maximal Qualified Score (MQS)	**Predicted UN GHS classification**
0 – 12.5	No Category
> 12.5 – 30.0	No Prediction Can be Made
> 30.0	Category 1

MINIMUM LIST OF REFERENCE CHEMICALS

16. Reference Chemicals are used to determine if the reliability and relevance of a proposed similar or modified test method, proven to be structurally and functionally sufficiently similar to the *in vitro* macromolecular test method VRM, or representing a minor modification of the VRM, are comparable or better to those of the VRM (5). The 30 Reference Chemicals listed in Table 2 include chemicals

representing different chemical classes of interest and are representative of the full range of TG 405 *in vivo* ocular hazard, *i.e.*, serious eye damage (UN GHS Cat. 1), eye irritation (UN GHS Cat. 2) and non-classified chemicals (UN GHS No Category) (3) (4). The distribution of chemicals in this list comprise 10 UN GHS Cat. 1 chemicals, 10 Cat. 2 chemicals and 10 No Category test chemicals. The Reference Chemicals were selected from the test chemicals used in the validation study of the VRM (5) using the selection criteria as described in Table 2 (foot-note A).

17. The 30 Reference Chemicals listed in Table 2 represent the minimum number of chemicals that should be used to evaluate the reliability and relevance of a proposed similar or modified *in vitro* macromolecular test method for identifying i) chemicals inducing serious eye damage (Category 1) and ii) chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Category), in accordance with the UN GHS (3). The use of these Reference Chemicals for the development/optimization of new similar test methods should be avoided to the extent possible. In situations where a listed chemical is unavailable, or where justifiable, another chemicals fulfilling the selection criteria as described in Table 2 (foot-note A) and for which adequate *in vivo* reference data are available, could be used, e.g. primarily from the test chemicals used in the validation study of the VRM (5). Additional chemicals representing other chemical or product classes and for which adequate *in vivo* reference data are available are recommended to be tested in addition to the minimum list of Reference Chemicals to further evaluate the accuracy of the proposed test method.

Table 2. Minimum list of Reference Chemicals for determination of Reproducibility and Predictive Capacity of similar or modified *In Vitro* macromolecular test method for identifying i) chemicals inducing serious eye damage (UN GHS Category 1) and ii) chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Category)

Chemical name [A]	CASRN	*In vivo* UN GHS	Physical state	pH [B]	Organic functional group [C]	VRM Prediction (5)
2-benzyl-4-chlorophenol	120-32-1	Category 1	Solid	5.9	Aryl halide, Benzyl, Phenol	NPCM
2-methylresorcinol	608-25-3	Category 1	Solid	5.8	Benzyl, Phenol	Cat. 1
4-(1,1,3,3-tetramethylbutyl)phenol	140-66-9	Category 1	Solid	5.2	Alkane branched with quaternary carbon, Phenol	NPCM
4-tert-butylcatechol	98-29-3	Category 1	Solid	5.5	Phenol	Cat. 1
Benzalkonium chloride (5%)	63449-41-2	Category 1	Liquid	6.5	Ammonium quaternary (salt), Benzyl	Cat. 1*
Cetylpyridinium bromide (6%)	140-72-7	Category 1	Liquid	4.4	Heterocyclic fragment, Pyridine	Cat. 1*
Lauric acid	143-07-7	Category 1	Solid	4.5	Carboxylic acid	NPCM*
Promethazine hydrochloride	58-33-3	Category 1	Solid	4.5	Aliphatic Amine, Arene, Heterocyclic fragment, Sulfide	Cat. 1
p-tert-butylphenol	98-54-4	Category 1	Solid	7.7	Phenol	NPCM
Sodium oxalate	62-76-0	Category 1	Solid	7.0	Carboxylic acid	NPCM
2,4,11,13-tetraazatetradecanediamidine, N,N"-bis(4-chlorophenyl)-3,12-	18472-51-0	Category 2A	Liquid	6.3	Alcohol, Aliphatic Amine secondary, Amidine, Arene, Aryl halide, Carboxylic acid, Imidine (substituted)	Cat. 1

Chemical name [A]	CASRN	*In vivo* UN GHS	Physical state	pH [B]	Organic functional group [C]	VRM Prediction (5)
diimino-, di-D-gluconic acid (20% aqueous)						
Ammonium nitrate	6484-52-2	Category 2A	Solid	4.8	n.a.	NPCM
Cetylpyridinium bromide (1%)	140-72-7	Category 2A	Liquid	4.7	Heterocyclic fragment, Pyridine	NPCM*
Methyl acetate	79-20-9	Category 2A	Liquid	6.8	Acetoxy	NPCM
Methyl cyanoacetate	105-34-0	Category 2A	Liquid	5.7	Carboxylic acid ester, Nitrile	Cat. 1*
Naphthalene-1,5-diol	83-56-7	Category 2A	Solid	5.8	Fused polyciclic aromatic, Phenol	No Cat.
Sodium lauryl glucose carboxylate (and) lauryl glucoside	383178-66-3 (110615-47-9)	Category 2A	Liquid	5.7	n.a.	Cat. 1
Propasol solvent P	1569-01-3	Category 2A	Liquid (viscous)	6.2	Alcohol, Ether	Cat. 1
Sodium benzoate	532-32-1	Category 2A	Solid	8.2	Arene, Carboxylic acid	NPCM
Sodium chloroacetate	3926-62-3	Category 2B	Solid	6.1	Alkyl halide, Carboxylic acid	NPCM
1,5-dibromopentane	111-24-0	No category	Liquid	5.7	Alkyl halide	No Cat.
2,2-dimethyl-3-pentanol	3970-62-5	No category	Liquid	5.3	Alcohol, Alkane branched with quaternary carbon	NPCM
2-(2-ethoxyethoxy)ethanol	111-90-0	No category	Liquid (viscous)	5.6	Alcohol, Ether	NPCM
Cetyl pyridinium bromide 0.1%	140-72-7	No category	Liquid	7.1	Heterocyclic fragment, Pyridine	No Cat.
Di-n-propyl disulphide	629-19-6	No category	Liquid	6.1	Disulfide	No Cat.
Dioctyl ether	629-82-3	No category	Liquid	7.1	Ether	No Cat.
Myristyl myristate	3234-85-3	No category	Solid	6.3	Carboxylic acid ester	No Cat.
n,n-dimethylguanidine sulphate	598-65-2	No category	Solid	6.8	Alphatic Amine, tertiary, Amidine	NPCM
Potassium tetrafluoroborate	14075-53-7	No category	Solid	4.5	n.a.	No Cat.
Sodium lauryl sulfate (3%)	151-21-3	No category	Liquid	6.8	Sulfate	NPCM

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; Cat.: category; n.a. = not available; NPCM: No Prediction Can be Made; UN GHS = United Nations Globally Harmonized System of Classification and Labelling of Chemicals (3).

[A] The 30 Reference Chemicals comprise a representative selection from the 88 chemicals that were used to validate the reference test method (Ocular Irritection®) (5). The goal of the selection process was to include, to the extent possible, chemicals that: (i) cover the full range of *in vivo* serious eye damage/eye irritation responses based on the UN GHS classification system (i.e., Categories 1, 2A, 2B or No Category); (ii) are based on high quality results obtained in the reference *in vivo* rabbit eye test (OECD TG 405) (4) (6); (iii) cover different physical states; (iv) cover a broad range of the chemical classes and organic functional groups, representative of those used in the validation study (5); (v) reflect the overall performance characteristics of the reference test method; (vi) cover the full range of *in vitro* responses based on high quality Ocular Irritection® data (0 to 51 MQS); (vii) produced reproducible results in the VRM; (viii) are commercially available; and (ix) are not associated with prohibitive disposal costs.

[B] The pH values are rounded to one decimal point, and values were obtained from the original sources as indicated in (5).

[C] The organic functional groups were characterized using the OECD QSAR toolbox (version 2.3) as described in (5).

* Test chemicals having limited data in within- and between- laboratory reproducibility but included as representing relevant chemistries and/or outcome. These chemicals were however not taken into account for establishing the minimum performance standards for reproducibility as described in paragraphs 23 to 28.

DEFINED RELIABILITY AND ACCURACY VALUES

18. For purposes of establishing the reliability and relevance of proposed similar or modified *in vitro* macromolecular test methods to be used by several independent laboratories, all 30 Reference Chemicals listed in Table 2 should be tested in at least three laboratories. However, an assessment of between-laboratory reproducibility is not essential if the proposed test method is to be used in one laboratory only. In each laboratory, all Reference Chemicals should be tested in three independent experiments performed at sufficiently spaced time points. Each experiment should consist of at least five concurrently tested doses/concentrations for each test chemical, and appropriate controls: blanking samples, four calibrating chemicals and two Quality Control chemicals..

19. The calculation of the within-laboratory reproducibility, between-laboratory reproducibility, accuracy, sensitivity and specificity values of the proposed test method should be done according to the rules described below to ensure that a predefined and consistent approach is used:

- *Within-laboratory reproducibility:* for chemicals having MQS within the 0-51 range of responses, the standard deviation obtained for the three independent experiments should be calculated, and then the overall mean of the 25 Reference Chemicals having WLR data should be calculated for each participating laboratory. In addition, the concordance of the predictions of the three independent experiments of the 25 Reference Chemicals having WLR data should be calculated for both cut-offs (12.5 and 30.0) and for each participating laboratory (possible non-qualified and excluded test results should be included in this analysis as considered integral part of the result spectrum of the VRM).

- *Between-laboratory reproducibility:* for chemicals having MQS within the 0-51 range of responses, the mean MQS value is calculated for each laboratory (obtained from the three independent experiments), the standard deviation obtained for the results from at least three laboratories is then calculated, and the overall mean SD for the 25 Reference Chemicals having BLR data should be then calculated. In addition, the concordance of the predictions between at least three laboratories should be calculated for both cut-offs (12.5 and 30.0) based on the majority laboratory classification (possible non-qualified and excluded test results should be included in this analysis as considered integral part of the result spectrum of the VRM).

- *Predictive capacity:* should be assessed based on the majority of predictions available per chemical for all 30 Reference Chemicals. The concordance of these predictions with the expected result as defined by the *in vivo* UN GHS classifications (dichotomized into UN GHS Category 1 vs. non-Category 1 chemicals and into UN GHS No Category vs. classified chemicals) should then be used to calculate the specificity, sensitivity and concordance of results.

Within-laboratory reproducibility

20. An assessment of within-laboratory reproducibility (WLR) for similar or modified test method proposed should show in terms of MQS, an overall mean standard deviation (SD) for the 25 Reference Chemicals having WLR data from three independent experiments that is smaller or equal (≤) than 3 within each laboratory (actual for Ocular Irritection®: 1.7, 2.2, 1.5 in each laboratory for the 25 Reference Chemicals having WLR data, and 2.0, 2.6, 2.2 for the overall validation dataset (5)).

21. In addition, to discriminate UN GHS Category 1 from non-Cat. 1 chemicals, the assessment of within-laboratory reproducibility for similar or modified test method proposed should show in every laboratory, a concordance of predictions (UN GHS Cat. 1 versus non-Cat. 1 test chemicals) obtained for the 25 Reference Chemicals having WLR data from three independent experiments that is equal or higher (≥) than 85% (actual for Ocular Irritection®: 96%, 88%, 92% in each laboratory for the 25 Reference Chemicals having WLR data, and 91%, 88%, 84%for the overall validation dataset (5)).

22. Finally, to discriminate UN GHS No Category from UN GHS classified chemicals (but not to categorize classified chemicals), the assessment of within-laboratory reproducibility for similar or modified test method proposed should show in every laboratory, a concordance of predictions (UN GHS No Category versus classified test chemicals) for the 25 Reference Chemicals having WLR Reference Chemicals data obtained from three independent experiments that is equal or higher (≥) than 80% (actual for Ocular Irritection®: 96%, 88% and 84% in each laboratory for the 25 Reference Chemicals having WLR data, and 89%, 86%, 78% for the overall validation dataset (5)).

Between-laboratory reproducibility

23. An assessment of between-laboratory reproducibility (BLR) for similar or modified test method proposed should show in terms of MQS, an overall mean standard deviation (SD) for the 25 Reference Chemicals having BLR data that is smaller or equal (≤) than 3 for studies conducted in three different laboratories actual for Ocular Irritection®: 1.9 for the 25 Reference Chemicals having BLR data, and 2.5 for the overall validation dataset (5)).

24. In addition, for similar or modified test methods proposed to discriminate between to discriminate UN GHS Category 1 from non-Cat. 1 chemicals, the between-laboratory concordance of predictions (UN GHS Cat. 1 versus non-Cat. 1 test chemicals) obtained for the 25 Reference Chemicals having BLR data should be equal or higher (≥) than 80% (actual for Ocular Irritection®: 84% for the 25 Reference Chemicals having BLR data, and 84% for the overall validation dataset (5)).

25. Finally, for similar or modified test methods proposed to discriminate between to discriminate UN GHS No Category from UN GHS classified chemicals (but not to categorize classified chemicals), the between-laboratory concordance of predictions (UN GHS No Category versus classified test chemicals) obtained for the 25 Reference Chemicals having BLR data should be equal or higher (≥) than 80% (actual for Ocular Irritection®: 92% for the 25 Reference Chemicals having BLR data, and 82% for the overall validation dataset (5))

Predictive capacity

26. The predictive capacity (sensitivity, specificity, false negative rate, false positive rate, ability to correctly identify UN GHS Cat. 1 and UN GHS No Category chemicals of the proposed similar or modified *in vitro* macromolecular test methods should be comparable or better to that of the validated test method (5) (Table 3).

Table 3. Required sensitivity, specificity and accuracy for similar or modified *in vitro* macromolecular (based on the predictive capacity obtained with the Reference Test Method)[1]

Purpose	Accuracy	Sensitivity	Specificity
Identification of UN GHS Cat. 1 chemicals	70% (actual VRM 74%)	50% (actual VRM 50%)	80% (actual VRM 81%)
Identification of UN GHS No Cat. chemicals	83% (actual VRM 81%)	95% (actual VRM 98%)	60% (actual VRM 63%)

[1] Table 3 provides the accuracy of the reference test method in correctly identifying i) chemicals inducing serious eye damage (UN GHS Category 1) and ii) chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Category) based on the results of the 30 Reference Chemicals (Table 2).

LITERATURE

(1) DB-ALM (INVITTOX) (2013) Protocol 157: Ocular Irritection® Assay System, 25pp. Available: http://ecvam-dbalm.jrc.ec.europa.eu.

(2) OECD (2005). OECD Series on Testing and Assessment No. 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: [http://www.oecd.org/document/30/0,3343,en_2649_34377_1916638_1_1_1_1,00.html].

(3) United Nations (UN) (2017). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Seventh revised edition, UN New York and Geneva, 2017. Available at: https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev07/English/03e_part3.

(4) OECD (2017). Test Guideline 405. OECD Guideline for Testing of Chemicals. Acute eye irritation/corrosion. Organisation for Economic Cooperation and Development, Paris, France. Available at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

(5) Eskes C, Hoffmann S, Facchini D, Ulmer R, Wang A, Flego M, Vassallo M, Bufo M, van Vliet E, d'Abrosca F, Wilt N (2014). Validation Study on the Ocular Irritection® Assay for Eye Irritation Testing. Toxicology In Vitro 28, 1046-1065.

(6) Barroso J., Pfannenbecker U., Adriaens E., Alépée N., Cluzel M., De Smedt A., Hibatallah J., Klaric M., Mewes K.R., 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). Archives of Toxicology 91, 521-547.

(7) Scott L, Eskes C, Hoffman S, Adriaens E, Alepee N, Bufo M, Clothier R, Facchini D, Faller C, Guest R, Hamernik K, Harbell J, Hartung T, Kamp H, Le Varlet B, Meloni M, Mcnamee P, Osborn R, Pape W, Pfannenbecker U, Prinsen M, Seaman C, Spielmann H, Stokes W, Trouba K, Vassallo M, Van den Berghe C, Van Goethem F, 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. *Toxicology In vitro* 24,1-9.

(8) OECD (2017). Guidance Document on an Integrated Approach on Testing and Assessment for Serious Eye Damage and Eye Irritation. Series on Testing and Assessment, No. 263. Environment, Health and Safety Publications, Organisation for Economic Cooperation and Development, Paris, France. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2017)15&doclanguage=en

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of a test method (2).

Activator: Solution employed to initiate formation of the ordered macromolecular matrix when the protein has been rehydrated.

Benchmark chemical: A chemical used as a standard for comparison to a test chemical. A benchmark chemical should have the following properties; (i), a consistent and reliable source(s); (ii), structural and functional similarity to the class of chemicals being tested; (iii), known physical/chemical characteristics; (iv) supporting data on known effects; and (v), known potency in the range of the desired response.

Between-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Between-laboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as inter-laboratory reproducibility (2).

Blank qualification: The blank OD for each sample is checked to fall in an appropriate range (i.e., pre-established minimum and maximum blank OD), and for flatness (i.e., OD variability between two consecutive doses/concentrations, and between the highest and lowest doses/concentrations in a group of 3 doses/concentrations).

Bottom-Up approach: Step-wise approach used for a test chemical suspected of not requiring classification and labelling for eye irritation or serious eye damage, which starts with the determination of chemicals not requiring classification and labelling (negative outcome) from other chemicals (positive outcome) (7) (8).

Calibrators: defined irritant solutions having well characterized IDE scores. The calibrators are used to derive a standard curve with which the results of the test method are compared to, and ensure optimal performance.

Cornea: The transparent part of the front of the eyeball that covers the iris and pupil and admits light to the interior.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (2).

CV: Coefficient of Variation.

Eye irritation: Production of changes in the eye, which are fully reversible, occurring after the exposure of the eye to a substance or mixture. Interchangeable with "Reversible effects on the Eye" and with "UN GHS Category 2" (3).

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

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

Foam test: employed to determine whether the unknown substance should be tested utilizing surfactant or non-surfactant application procedure (8).

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

Hydrating Solution: Solution employed to rehydrate the reagent powder and facilitate formation of the ordered protein matrix.

IATA: Integrated Approach on Testing and Assessment (8).

Inhibition check solution: An irritating substance known to quickly react with the macromolecular reagent and produce evident turbidity, which can be employed to verify the functionality of macromolecular reagent when the OD readings of qualified test chemical doses/concentrations are less than Calibrator 2. The inhibition check solution verifies that the macromolecular reagent in those wells is still able to produce evident turbidity (e.g., > OD Calibrator 2) and identifies inaccurate low OD reading (or inaccurate non-irritant) results when the turbidity is less than OD Calibrator 2.

Irreversible effects on the eye: See “Serious eye damage” and "UN GHS Category 1".

Irritection Draize Equivalent (IDE) Score: A numerical score derived from the optical density measurement of the Ocular Irritection® assay for a tested dose/concentration when compared to the curve obtained with the calibrators.

Maximal Qualified Score (MQS): Represents the highest IDE Score obtained from the different tested doses/concentrations of a test chemical. Ranging from 0 to 51 it is used to predict the irritation potential of the test chemical.

Membrane Discs: A semi-permeable membrane that facilitates controlled delivery of the test chemical into the protein reagent.

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation (2). The term is interchangeably used with similar test method.

Mixture: a mixture or solution composed of two or more substances in which they do not react (3).

Net Optical Density Check: Controls the net optical density by measuring the OD of the activated protein reagent and subtracting the OD of the activated blanking buffer. The Net OD (ODreagent – ODblank = ODNet) should be > -15.

Not Classified: Test chemicals that are not classified for eye irritation (UN GHS Category 2, 2A, or 2B) or serious eye damage (UN GHS Category 1). Interchangeable with "UN GHS No Category".

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

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

Predictive Capacity: The predictive capacity reflects the test method performance in terms of correct and incorrect predictions in comparison to reference data. It gives quantitative information (e.g. correct prediction rate) on the relevance of the test method. It comprises amongst others, the sensitivity and specificity of the test method.

Quality Control chemicals: Two defined irritant solutions (QC1 and QC2) with well characterized IDE scores within the lower (7.2-20.8) and mid-upper range (23.6-35.6) of the Ocular Irritection® test method. The quality control check verifies that the method is functioning properly and can correctly detect eye irritation potency in the lower and mid/upper IDE ranges.

Reagent Powder: Consists of a mixture of proteins, glycoproteins, carbohydrates, lipids and low molecular weight components. When hydrated, the reagent powder forms a solution containing an ordered macromolecular matrix. Proteins in this solution undergo changes in conformation when exposed to an irritant test chemical.

Reference Chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses (2).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (2).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (2).

Reproducibility: The agreement among results obtained from testing the same substance using the same test protocol (2).

Reversible effects on the eye: See “Eye irritation” and "UN GHS Category 2".

SD: Standard Deviation.

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (2).

Serious eye damage: 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. Interchangeable with "Irreversible effects on the eye" and with "UN GHS Category 1" (3).

Solvent/vehicle control: An untreated sample containing all components of a test system, including the solvent or vehicle that is processed with the test chemical-treated and other control samples to establish the baseline response for the samples treated with the test chemical dissolved in the same solvent or vehicle. When tested with a concurrent negative control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (2).

Standard Operating Procedures (SOP): Formal, written procedures that describe in detail how specific routine, and test-specific, laboratory operations should be performed. They are required by GLP.

Substance: means chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition (4).

Surfactants: Also called surface-active agent, this is a substance and/or its dilution (in an appropriate solvent/vehicle), which consists of one or more hydrophilic and one or more hydrophobic groups, that is capable of reducing the surface tension of a liquid and of forming spreading or adsorption monolayers at the water-air interface, and/or of forming emulsions and/or microemulsions and/or micelles, and/or of adsorption at water-solid interfaces.

Test chemical: The term "test chemical" is used to refer to what is being tested.

Tiered testing strategy: A stepwise testing strategy, which uses test methods in a sequential manner. All existing information on a test chemical is reviewed at each tier, using a weight-of-evidence process, to determine if sufficient information is available for a hazard classification decision, prior to progression to

the next tier in the strategy. If the hazard potential/potency of a test chemical can be assigned based on the existing information at a given tier, no additional testing is required (7) (8).

Top-Down approach: Step-wise approach used for a chemical suspected of causing serious eye damage, which starts with the determination of chemicals inducing serious eye damage (positive outcome) from other chemicals (negative outcome) (7) (8).

UN GHS (United Nations Globally Harmonized System of Classification and Labelling of Chemicals): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (3).

UN GHS Category 1: See “Serious eye damage” and/or "Irreversible effects on the eye".

UN GHS Category 2: See “Eye irritation” and/or "Reversible effects to the eye".

UN GHS No Category: Chemicals that do not meet the requirements for classification as UN GHS Category 1 or 2 (2A or 2B). Interchangeable with “Not Classified”.

Validated Reference Method(s) (VRM(s)): one (or more) test method(s) that was(were) used to develop the related official Test Guidelines and Performance Standards (PS). The VRM(s) is(are) considered the reference test method(s) to compare new proposed similar or modified test methods in the framework of a PS-based validation study.

Within-laboratory reproducibility: determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times, also referred to as intra-laboratory reproducibility (2).