OECD/OCDE TG 431

Adopted:
26 September 2014

OECD GUIDELINES FOR THE TESTING OF CHEMICALS

IN VITRO SKIN CORROSION: RECONSTRUCTED HUMAN EPIDERMIS (RHE) TEST METHOD

INTRODUCTION

1. Skin corrosion refers to the production of irreversible damage to the skin manifested as visible necrosis through the *epidermis* and into the *dermis*, following the application of a test chemical [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS)] (1). This updated Test Guideline 431 provides an *in vitro* procedure allowing the identification of non-corrosive and corrosive substances and mixtures in accordance with UN GHS (1). It also allows a partial sub-categorization of corrosives.

2. The assessment of skin corrosivity has typically involved the use of laboratory animals (OECD Test Guideline 404 (TG 404); adopted in 1981 and revised in 1992 and 2002) (2). In relation to animal welfare concerns, TG 404 recommends the use of a tiered testing strategy for the determination of skin corrosion and irritation which includes the use of validated *in vitro* or *ex vivo* test methods avoiding pain and suffering of animals. In addition to TG 431 (originally adopted in 2004) (3), two other *in vitro* test methods for testing of corrosivity have been validated and adopted as OECD Test Guidelines 430 (4) and 435 (5). Three validated *in vitro* test methods have been adopted as OECD TG 439 (6), to be used for the skin irritation part of the tiered testing strategy of TG 404 (2).

3. This Test Guideline addresses the human health endpoint skin corrosion. It makes use of reconstructed human *epidermis* (RhE) (obtained from human derived non-transformed epidermal keratinocytes) which closely mimics the histological, morphological, biochemical and physiological properties of the upper parts of the human skin, *i.e.* the *epidermis*. This Test Guideline was originally adopted in 2004 and updated in 2013 and 2014 to include a set of Performance Standards (PS) (Annex 1) for the assessment of similar and modified RhE-based test methods (7), in accordance with the principles of Guidance Document No. 34 (8). Other updates comprise the addition of two test methods using the RhE

models SkinEthic™ RHE[1] and epiCS® (previously named EST-1000), and the possibility to use the methods to support the sub-categorisation of corrosive chemicals.

4. Four validated test methods using commercially available RhE models are included in this Test Guideline. Prevalidation studies (9), followed by a formal validation study for assessing skin corrosion (10)(11) (12) have been conducted (13) (14) for two of these commercially available test methods, EpiSkin™ Standard Model (SM) and EpiDerm™ Skin Corrosivity Test (SCT) (EPI-200) (referred to in the following text as the Validated Reference Methods – VRMs). The outcome of these studies led to the recommendation that the two VRMs mentioned above could be used for regulatory purposes for distinguishing corrosive (C) from non-corrosive (NC) substances, and that the EpiSkin™ could moreover be used to support sub-categorization of corrosive substances (15) (16) (17). Two other commercially available *in vitro* skin corrosion RhE test methods have shown similar results to the EpiDerm™ VRM according to PS-based validation (18) (19) (20). These are the SkinEthic™ RHE and epiCS® (previously named EST-1000) that can also be used for regulatory purposes for distinguishing corrosive from non-corrosive substances (21) (22). Post validation studies performed by the RhE model producers in the years 2012 to 2014 with a refined protocol correcting interferences of unspecific MTT reduction by the test chemicals improved the performance of both discrimination of C/NC as well as supporting sub-categorisation of corrosives (23) (24).

5. Before a proposed similar or modified *in vitro* RhE test method for skin corrosion other than the VRMs can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure its similarity to the VRMs, in accordance with the requirements of the PS set out in this Test Guideline (Annex 1). The Mutual Acceptance of Data will only be guaranteed after any proposed new or updated test method following the PS of this Test Guideline have been reviewed and included in this Test Guideline. The test methods included in this Test Guideline can be used to address countries' requirements for test results on *in vitro* test method for skin corrosion,while benefiting from the Mutual Acceptance of Data.

DEFINITIONS

6. Definitions used are provided in Annex 2.

INITIAL CONSIDERATIONS

7. This Test Guideline addresses the *in vitro* skin corrosion component of the tiered testing strategy recommended within TG 404 for dermal corrosion/irritation assessment (2) (25). It allows the identification of non-corrosive and corrosive substances and mixtures in accordance with the UN GHS (1). This Test Guideline further supports the sub-categorization of corrosive substances and mixtures into optional Category 1A, in accordance with the UN GHS (1), as well as a combination of Categories 1B and 1C (23) (24). A limitation of this Test Guideline is that it does not allow discriminating between skin corrosive sub-categories 1B and 1C in accordance with the UN GHS (1) due to the limited set of well-

[1] Please note that the abbreviation RhE (=Reconstructed human Epidermis) is used for all models based on RhE technology. The abbreviation RHE as used in conjunction with the SkinEthic™ model means the same, but, as part of the name of this specific test method as marketed, is spelled all in capitals.

known *in vivo* corrosive Category 1C chemicals. EpiSkin™, EpiDerm™, SkinEthic™ and epiCS®test methods are able to sub-categorize (i.e. 1A versus 1B-and-1C versus NC) but differences are observed between EpiSkin™ and the three other test methods, EpiDerm™, SkinEthic™ and epiCS® in view of their capacity to provide information on sub-categorisation. Results from EpiSkin™ can be used as such; whereas results from EpiDerm™, SkinEthic™ and epiCS® generate high over-classification rates for a combination of categories 1B and 1C (see Annex 4). Therefore, for EpiDerm™, SkinEthic™ and epiCS®, chemicals that are classified as 1B-and-1C can be considered as 1B-and-1C, while chemicals for which cell viability at 3 minutes is below 50% should just be considered as Category 1, since the Category 1A predictions of these three test methods contain a high rate of over-predictions of chemicals of Categories 1B and 1C. The regulatory framework in member countries will decide how this Test Guideline will be used, e.g. acknowledging the significant probability of overclassification, a Category 1A classification may still be accepted or further testing may be conducted to confirm the result.

8. A wide range of chemicals representing mainly individual substances has been tested in the validation supporting the test methods included in this Test Guideline when they are used for identification of non-corrosives and corrosives; the empirical database of the validation study amounted to 60 chemicals covering a wide range of chemical classes (10) (11) (12). Testing to demonstrate sensitivity, specificity, accuracy and within-laboratory-reproducibility of the assay for sub-categorization was performed by the test method developers and results were reviewed by the OECD (23) (24). On the basis of the overall data available, the Test Guideline is applicable to a wide range of chemical classes and physical states including liquids, semi-solids, solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other prior treatment of the sample is required. In cases where evidence can be demonstrated on the non-applicability of test methods included in the Test Guideline to a specific category of test chemicals, these test methods should not be used for that specific category of test chemicals. In addition, this Test Guideline is assumed to be applicable to mixtures as an extension of its applicability to substances. However, due to the fact that mixtures cover a wide spectrum of categories and composition, and that only limited information is currently available in the public domain on the testing of mixtures, in cases where evidence can be demonstrated on the non-applicability of the Test Guideline to a specific category of mixtures (*e.g.* following a strategy as proposed in (26)), the Test Guideline should not be used for that specific category of mixtures. Gases and aerosols have not been assessed yet in validation studies (10) (11) (12). While it is conceivable that these can be tested using RhE technology, the current Test Guideline does not allow testing of gases and aerosols. It should also be noted that some chemicals may interfere with the cell viability measurements and need the use of adapted controls for corrections (see paragraphs 25 to 27).

9. While this Test Guideline does not provide adequate information on skin irritation, it should be noted that OECD TG 439 specifically addresses the health effect skin irritation *in vitro* and is based on the same RhE test system, though using another protocol (6). For a full evaluation of local skin effects after a single dermal exposure, it is recommended to follow the sequential testing strategy as appended to TG 404 (2) (25). This testing strategy includes the conduct of *in vitro* tests for skin corrosion (such as described in this Test Guideline) and skin irritation before considering testing in living animals. It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

10. This Test Guideline also includes a set of Performance Standards (PS) (Annex 1) for determining the validation status (reliability and relevance) of similar and modified skin corrosion test methods that are structurally and mechanistically similar to the VRMs (7), in accordance with the principles of Guidance Document No. 34 (8). These PS include a list of Reference Substances by which to evaluate assay performance, the essential test method components by which to evaluate the structural, mechanistic and procedural similarity of a new proposed test method, and the minimum reliability and accuracy values necessary for the test method to be considered comparable to the VRMs. Within the Reference Chemical list, a subset of 13 Proficiency Substances (Table 1) is provided to be used by laboratories to demonstrate proficiency in using *in vitro* human skin models (see paragraphs 13 and 14).

PRINCIPLE OF THE TEST

11. The test chemical is applied topically to a three-dimensional RhE model, comprised of non-transformed, human-derived epidermal keratinocytes, which have been cultured to form a multi-layered, highly differentiated model of the human *epidermis*. It consists of organized basal, spinous and granular layers, and a multi-layered *stratum corneum* containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo.*

12. The RhE test method is based on the premise that corrosive chemicals are able to penetrate the *stratum corneum* by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. Cell viability is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (27). Corrosive chemicals are identified by their ability to decrease cell viability below defined threshold levels (see paragraphs 31 and 32). The RhE-based skin corrosion test methods have shown to be predictive of *in vivo* skin corrosion effects assessed in rabbits according to the OECD guideline 404 (2).

DEMONSTRATION OF PROFICIENCY

13. Prior to routine use of any of the four validated RhE test methods that adhere to this Test Guideline, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances listed in Table 1. In case of the use of a method for sub-classification, also the correct sub- categorization should be demonstrated.

Table 1: List of Proficiency Substances

Chemical[1]	CASRN	Chemical Class[2]	UN GHS Cat. Based on *In Vivo* results [3]	VRM Cat. Based on *In Vitro* results[4]	MTT Reducer[5]	Physical State
Category 1A *In Vivo* Corrosives						
Bromoacetic acid	79-08-3	Organic acid	1A	(3) 1A	--	S
Boron trifluoride dihydrate	13319-75-0	Inorganic acid	1A	(3) 1A	--	L
Phenol	108-95-2	Phenol	1A	(3) 1A	--	S
Dichloroacetyl chloride	79-36-7	Electrophile	1A	(3) 1A	--	L
Category 1B/1C *In Vivo* Corrosives						
Glyoxylic acid monohydrate	563-96-2	Organic acid	1B/1C	(3) 1B/1C	--	S
Lactic acid	598-82-3	Organic acid	1B/1C	(3) 1B/1C	--	L
Ethanolamine	141-43-5	Organic base	1B	(3) 1B/1C	Y	Viscous
Hydrochloric acid (14.4%)	7647-01-0	Inorganic acid	1B/1C	(3) 1B/1C	--	L
***In Vivo* Non Corrosives**						
Phenethyl bromide	103-63-9	Electrophile	NC	(3) NC	Y	L
4-Amino-1,2,4-triazole	584-13-4	Organic base	NC	(3) NC	--	S
4-(methylthio)-benzaldehyde	3446-89-7	Electrophile	NC	(3) NC	Y	L
Lauric acid	143-07-7	Organic acid	NC	(3) NC	--	S

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; UN GHS = United Nations Globally Harmonized System (1); VRM = Validated Reference Method; NC = Not Corrosive

[1]These substances, sorted first by corrosives versus non-corrosives, then by corrosive sub-category and then by chemical class, were selected from the substances used in the ECVAM validation studies of EpiSkin™ and EpiDerm™ (10) (11) (12) and from post-validation studies based on data provided by EpiSkin™ (24), EpiDerm™, SkinEthic™ and epiCS® developers. Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (10) (12). The selection includes, to the extent possible, substances that: (i) are representative of the range of corrosivity responses (*e.g.* non-corrosives; weak to strong corrosives) that the VRMs are capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation studies; (iii) have chemical structures that are well-defined; (iv) induce reproducible results in the VRM; (v) induce definitive results in the *in vivo* reference test method; (vi) are commercially available; and (vii) are not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt *et al.* (1998) (10).

[3]The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS 1A, 1B and 1C.

[4]The VRM *in vitro* predictions reported in this table were obtained with the EpiSkin™ and the EpiDerm™ test methods (VRMs) during post-validation testing performed by the test method developers.

[5]The viability values obtained in the ECVAM Skin Corrosion Validation Studies were not corrected for direct MTT reduction (killed controls were not performed in the validation studies). However, the post-validation data generated by the test method developers that are presented in this table were acquired with adapted controls.

14. As part of the proficiency exercise, it is recommended that the user verifies the barrier properties of the tissues after receipt as specified by the RhE model manufacturer. This is particularly important if tissues are shipped over long distance/time periods. Once a test method has been successfully established and proficiency in its use has been demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties in regular intervals.

PROCEDURE

15. The following is a generic description of the components and procedures of the RhE test methods for skin corrosion assessment covered by this Test Guideline. The RhE models endorsed as scientifically valid for use within this Test Guideline, *i.e.* the EpiSkin™ (SM), EpiDerm™ (EPI-200), SkinEthic™ RHE and epiCS® models (18) (19) (20) (28) (29) (30) (31) (32) (33), can be obtained from commercial sources. Standard Operating Procedures (SOPs) for these four RhE models are available (34) (35) (36) (37), and their main test method components are summarized in Annex 3. It is recommended that the relevant SOP be consulted when implementing and using one of these methods in the laboratory. Testing with the four RhE test methods covered by this Test Guideline should comply with the following:

RHE TEST METHOD COMPONENTS

General Conditions

16. Non-transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum*, *stratum granulosum*) should be present under a functional *stratum corneum*. The *stratum corneum* should be multi-layered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic marker chemicals, *e.g.* sodium dodecyl sulfate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET_{50}) upon application of the marker chemical at a specified, fixed concentration (see paragraph 18). The containment properties of the RhE model should prevent the passage of material around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional Conditions

Viability

17. The assay used for determining the magnitude of viability is the MTT-assay (27). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control. The optical density (OD) of the extraction solvent alone should be sufficiently small, *i.e.* OD < 0.1. An acceptability range (upper and lower limit) for the negative control OD values is established by the RhE model developer/supplier, and the acceptability ranges for the four validated RhE test methods included in this Test Guideline are given in Table 2. It should be documented that the tissues treated with negative control are stable in culture (provide similar OD measurements) for the duration of the exposure period.

Table 2: Acceptability ranges for negative control OD values to control batch quality

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM)	≥ 0.6	≤ 1.5
EpiDerm™ SCT (EPI-200)	≥ 0.8	≤ 2.8
SkinEthic™ RHE	≥ 0.8	≤ 3.0
epiCS®	≥ 0.8	≤ 2.8

Barrier function

18. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of certain cytotoxic marker chemicals (*e.g.* SDS or Triton X-100), as estimated by IC_{50} or ET_{50} (see paragraph 21).

Morphology

19. Histological examination of the RhE model should be performed demonstrating multi-layered human *epidermis*-like structure (containing stratum basale, stratum spinosum, stratum granulosum and stratum corneum and exhibits lipid profile similar to lipid profile of human epidermis.

Reproducibility

20. Test method users should demonstrate reproducibility of the test methods over time with the positive and negative controls. Furthermore, the test method should only be used if the RhE model developer/supplier provides data demonstrating reproducibility over time with corrosive and non-corrosive chemicals from *e.g.* the list of Proficiency Substances (Table 1). In case of the use of a method for sub-categorization, also the reproducibility of sub-categorization should be demonstrated.

Quality control (QC)

21. The RhE model should only be used if the developer/supplier demonstrates that each batch of the RhE model used meets defined production release criteria, among which those for *viability* (paragraph 17), *barrier function* (paragraph 18) and *morphology* (paragraph 19) are the most relevant. These data are provided to the test method users, so that they are able to include this information in the test report. Only results produced with QC accepted tissue batches can be accepted for reliable prediction of corrosive classification. An acceptability range (upper and lower limit) for the IC_{50} or the ET_{50} is established by the RhE model developer/supplier. The acceptability ranges for the four validated test methods are given in Table 3.

Table 3:QC batch release criteria

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS)(35)	IC_{50} = 1.0 mg/mL	IC_{50} = 3.0 mg/mL
EpiDerm™ SCT (EPI-200) (1% Triton X-100)(36)	ET_{50} = 4.0 hours	ET_{50} = 8.7 hours
SkinEthic™ RHE (1% Triton X-100)(37)	ET_{50} = 4.0 hours	ET_{50} = 10.0 hours
epiCS®(1% Triton X-100)(38)	ET_{50} = 2.0 hours	ET_{50} = 7.0 hours

Application of the Test Chemical and Control Substance

22. At least two tissue replicates should be used for each test chemical and controls for each exposure time. For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the epidermis surface while avoiding an infinite dose, *i.e.* a minimum of 70 μL/cm^2 or 30 mg/cm^2 should be used. Depending on the methods, the epidermis surface should be moistened with deionized or distilled water before application of solid chemicals, to improve contact between the test chemical and the epidermis surface (34) (35) (36) (37). Whenever possible, solids should be tested as a fine powder. The application method should be appropriate for the test chemical (see *e.g.* references 12, 35-38). At the end of the exposure period, the test chemical should be carefully washed from the *epidermis* with an aqueous buffer, or 0.9% NaCl. Depending on which of the four validated RhE test methods is used, two or three exposure periods are used per test chemical (for all four valid RhE models: 3 min and 1 hour; for EpiSkin™ an additional exposure time of 4 hours). Depending on the RhE test method used and the exposure period assessed, the incubation temperature during exposure may vary between room temperature and 37ºC.

23. Concurrent negative and positive controls (PC) should be used in each run to demonstrate that viability (with negative controls), barrier function and resulting tissue sensitivity (with the PC) of the tissues are within a defined historical acceptance range. The suggested PC chemicals are glacial acetic acid or 8N KOH depending upon the RhE model used. It should be noted that 8N KOH is a direct MTT reducer that might require adapted controls as described in paragraphs 25 and 26. The suggested negative controls are 0.9% (w/v) NaCl or water.

Cell Viability Measurements

24. The MTT assay, which is a quantitative assay, should be used to measure cell viability under this Test Guideline (28). The tissue sample is placed in MTT solution of appropriate concentration (0.3 or 1 mg/mL) for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (*e.g.* isopropanol, acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm.

25. Test chemicals may interfere with the MTT assay, either by direct reduction of the MTT into blue formazan, and/or by colour interference if the test absorbs, naturally or due to treatment procedures, in the same OD range of formazan (570 ± 30 nm, mainly blue and purple chemicals). Additional controls should be used to detect and correct for a potential interference from these test chemicals (see paragraphs 26 and 27). This is especially important when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*, and is therefore present in the tissues when the MTT viability test is performed. Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the test methods (34) (35) (36) (37). Non-specific MTT reduction (NSMTT) and non-specific colour (NSC) due to these interferences may increase the OD of the tissue extract above the linearity range of the spectrophotometer. It is therefore important for each laboratory to determine the OD linearity range of their spectrophotometer with *e.g.* MTT formazan (CAS # 57360-69-7) commercially available from Sigma (Ref: M2003) before initiating the testing of test chemicals for regulatory purposes. When the OD of the tissue extract falls above the linearity range of the spectrophotometer, it should be diluted in acidified isopropanol or acidic isopropanol and the dilution factor should be taken into account when determining % NSMTT and/or % NSC relative to the negative controls ran concurrently to the test being corrected. Test results for materials inducing %NSMTT and/or %NSC ≥ 50% of negative control should be taken with caution.

26. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT medium. The mixture is incubated in the dark at 37°C, 5% CO_2 for a minimum of 60 min and a maximum of 180 min depending upon the RhE model used (34) (35) (36) (37). MTT medium is used as control. If the MTT mixture containing the test chemical (or suspension for insoluble compounds) turns blue/purple, the test chemical is presumed to directly reduce the MTT, and further functional check on non-viable epidermis should be performed. This additional functional check employs killed tissues that possess no metabolic activity but absorb and bind the test chemical in similar amount as viable tissues. Each MTT reducing chemical is applied on at least two killed tissue replicates per exposure time, which undergo the whole skin corrosion test. True tissue viability is calculated as the difference between the OD obtained with living tissue treated by MTT reducer and the OD obtained with frozen tissue treated by MTT reducer, and subsequently divided by the OD of the Negative Control concurrently tested.

27. To identify colour interference, spectral analysis of a coloured chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed to evaluate if the chemical requires additional controls. If the chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, colorant controls should be performed. Each coloured chemical is applied on at least two viable tissue replicates per exposure time, which undergo the entire skin corrosion test but are incubated with medium instead of MTT solution during the MTT incubation step. An independent NSC control needs to be performed concurrently per exposure time per coloured test chemical (in each run) due to the inherent biological variability of living tissues. True tissue viability is calculated as the difference between the OD obtained with living tissues incubated with MTT solution and the OD obtained with living tissues incubated with medium without MTT, and subsequently divided by the OD of the Negative Control performed in the same run.

Acceptability Criteria

28. For each test method using valid RhE models, tissues treated with the negative control should exhibit OD reflecting the quality of the tissues as described in table 2 and should not be below historically established boundaries. Tissues treated with the PC, *i.e.* glacial acetic acid or 8N KOH, should reflect the ability of the tissues to respond to a corrosive chemical under the conditions of the test method (see Annex 3). The variability between tissue replicates of test and/or control chemicals should fall within the accepted limits for each valid RhE model requirements (see Annex 3) (*e.g.* the difference of viability between the two tissue replicates should not exceed 30%). If either the negative control or PC included in a run fall out of the accepted ranges, the run is considered as not qualified and should be repeated. If the variability of test chemicals falls outside of the defined range, its testing should be repeated.

Interpretation of Results and Prediction Model

29. The OD values obtained for each test chemical should be used to calculate percentage of viability relative to the negative control, which is set at 100%. The cut-off percentage cell viability values distinguishing corrosive from non-corrosive test chemical (or discriminating between different corrosive sub-categories) are defined below in paragraphs 31 and 32 for each of the test methods covered by this Test Guideline and should be used for interpreting the results.

30. A single testing run composed of at least two tissue replicates should be sufficient for a test chemical when the resulting classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements, a second run may be considered, as well as a third one in case of discordant results between the first two runs.

31. The prediction model for the EpiSkin™ skin corrosion test method (11) (34) (24), associated with the UN GHS (1) classification system, is shown in Table 4:

Table 4:EpiSkin™ prediction model

Viability measured after exposure time points (t=3, 60 and 240 minutes)	**Prediction to be considered**
< 35% after 3 min exposure	**Corrosive:** • Optional Sub-category 1A *
≥ 35% after 3 min exposure **AND** < 35% after 60 min exposure **OR** ≥ 35% after 60 min exposure **AND** < 35% after 240 min exposure	**Corrosive:** • A combination of optional Sub-categories 1B and 1C
≥ 35% after 240 min exposure	Non-corrosive

*) According to the data generated in view of assessing the usefulness of the RhE test methods for supporting sub-categorisation, it was shown that around 22% of the Cat1A results of the EpiSkin™ test method may actually constitute Category 1B or 1C substances/mixtures (i.e. over classifications) (see Table 4.0 in Annex 4).

32. The prediction models for the EpiDerm™ SCT (13) (36), the SkinEthic™ RHE (19) (20) (36), and the epiCS® (18) (37) skin corrosion test methods, associated with the UN GHS (1) classification system, are shown in Table 5:

Table 5:EpiDerm™ SCT, SkinEthic™ RHE and epiCS®

Viability measured after exposure time points (t=3 and 60 minutes)	Prediction to be considered
< 50% after 3 min exposure	**Corrosive:** • Optional Sub-category 1A*
≥ 50% after 3 min exposure **AND** < 15% after 60 min exposure	**Corrosive:** • A combination of optional Sub-categories 1B and 1C
≥ 50% after 3 min exposure **AND** ≥ 15% after 60 min exposure	Non-corrosive

*) According to the data generated in view of assessing the usefulness of the RhE test methods for supporting sub-categorisation, it was shown that around 42% of the Cat1A results of the EpiDerm™ test method, and around 46% of the Cat 1A results of the SkinEthic™ and the epiCS® test method may actually constitute Category 1B or 1C substances/mixtures (i.e. over-classifications) (see Table 4.0 in Annex 4).

DATA AND REPORTING

Data

33. For each test, data from individual tissue replicates (*e.g.* OD values and calculated percentage cell viability for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition, means and ranges of viability and CVs between tissue replicates for each test should be reported. Observed interactions with MTT reagent by direct MTT reducers or coloured test chemicals should be reported for each tested chemical.

Test report

34. The test report should include the following information:

Test Chemical and Control Substance:

- Chemical name(s) such as IUPAC or CAS name and number, if known;
- Purity and composition of the substance or mixture (in percentage(s) by weight);
- Physical-chemical properties relevant to the conduct of the study (*e.g.* physical state, stability, volatility, pH, water solubility, if known);
- Treatment of the test chemical /control substance prior to testing, if applicable (*e.g.* warming, grinding);
- Storage conditions;

RhE model and protocol used and rationale for it (if applicable)

Test Conditions:

- RhE model used (including batch number);
- Calibration information for measuring device (*e.g.* spectrophotometer), band pass used for measuring cell viability, and OD linearity range of measuring device;
- Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to:
  - i) Viability
  - ii) Barrier function
  - iii) Morphology
  - iv) Reproducibility and predictive capacity
  - v) Quality controls (QC) of the model
- Details of the test procedure used. This should include, but is not limited to;
  - i) Washing procedures after exposure period
  - ii) Wavelength and band pass (filter) used to measure OD (cell viability)
- Test doses used, duration of exposure period(s) and temperature(s) of exposure;
- Number of tissue replicates used per test chemical and controls (PC, negative control, and NSMTT and NSC, if applicable), per exposure time;
- Indication of controls used for direct MTT-reducers and/or colouring test chemicals;
- Description of any modifications of the test procedure (including washing procedures);
- Reference to historical data of the model. This should include, but is not limited to;
  - i) Acceptability of the QC data with reference to historical data
  - ii) Acceptability of the positive and negative control values with reference to positive and negative control means and ranges
  - iii) Acceptability of the test results with reference to historical variability between tissue replicates
- Description of decision criteria/prediction model applied based on the RhE model used;

Results:

- Tabulation of data for individual test chemicals and controls, for each exposure period, each run and each replicate measurement, means, ranges and CVs, and the derived classification;
- Results of controls used for direct MTT-reducers and/or colouring test chemicals;
- Description of other effects observed;
- The derived classification with reference to the prediction model/decision criteria used;

Discussion of the results

Conclusions

LITERATURE

(1) UN (2013), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Fifth revised edition, UN New York and Geneva. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html.

(2) OECD (2002), *Test No. 404: Acute Dermal Irritation/Corrosion*, OECD Guideline for the Testing of Chemicals, Section 4, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264070622-en.

(3) OECD (2004), *Test No. 431: In vitro skin model*, OECD Guideline for the Testing of Chemicals, Section 4, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264071148-en.

(4) OECD (2004), *Test No. 430: In vitro skin corrosion*: *Transcutaneous Electrical Resistance (TER)*, OECD Guideline for the Testing of Chemicals, Section 4, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264071124-en.

(5) OECD (2006), *Test No. 435: In vitro membrane barrier test method*, OECD Guideline for the Testing of Chemicals, Section 4, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264067318-en.

(6) OECD (2010), *Test No. 439: In vitro skin irritation: reconstructed human epidermis test method*, OECD Guideline for the Testing of Chemicals, Section 4, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264090958-en.

(7) ICCVAM (2004), Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion, NIH Publication Number 04-4510, Research Triangle Park, NC: National Institute of Environmental Health Sciences, http://iccvam.niehs.nih.gov/docs/dermal_docs/ps/ps044510.pdf .

(8) OECD (2005), "Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 34, OECD Publishing, Paris.

(9) Botham, P.A. et al. (1995), A prevalidation study on *in vitro* skin corrosivity testing: The report and recommendations of ECVAM Workshop 6, *ATLA,* Vol. 23, pp. 219-255.

(10) Barratt, M.D. et al. (1998), The ECVAM international validation study on *in vitro* tests for skin corrosivity. 1. Selection and distribution of the test chemicals, *Toxicology in Vitro,* Vol. 12/4, pp. 471-482.

(11) Fentem, J.H. et al. (1998), The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the Management Team, *Toxicology in Vitro,* 12/4, pp. 483-524.

(12) Liebsch, M. et al. (2000), The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing, *ATLA*, Vol. 28, pp. 371-401.

(13) Balls, M. et al. (1995), Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshops, *ATLA,* Vol. 23, pp. 129-147.

(14) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (1997), Validation and Regulatory Acceptance of Toxicological Test Methods, NIH Publication No. 97-3981, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf.

(15) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2002), ICCVAM evaluation of EpiDerm™ (EPI-200), EPISKIN™ (SM), and the Rat Skin Transcutaneous Electrical Resistance (TER) assay: *In Vitro* test methods for assessing dermal corrosivity potential of chemicals. NIH Publication No. 02-4502. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.
http://iccvam.niehs.nih.gov/methods/epiddocs/epis_brd.pdf

(16) EC-ECVAM (1998), Statement on the scientific validity of the EpiSkin™ test (an *in vitro* test for skin corrosivity), issued by the ECVAM Scientific Advisory Committee (ESAC10), 3 April 1998. Available at: http://ecvam.jrc.ec.europa.eu.

(17) EC-ECVAM (2000), Statement on the application of the EpiDerm™ human skin model for skin corrosivity testing, issued by the ECVAM Scientific Advisory Committee (ESAC14), 21 March 2000. Available at: http://ecvam.jrc.ec.europa.eu.

(18) Hoffmann, J. et al. (2005), Epidermal-skin-test 1000 (EST-1000)-A new reconstructed epidermis for *in vitro* skin corrosivity testing, *Toxicology in Vitro,* Vol. 19/7, pp. 925-929.

(19) Kandárová, H. et al. (2006), Assessment of the human epidermis model SkinEthic RHE for *in vitro* skin corrosion testing of chemicals according to new OECD TG 431, *Toxicology in Vitro,* Vol. 20/5, pp. 547–559.

(20) Tornier, C., M. Roquet, A.B. Fraissinette (2010), Adaptation of the validated SkinEthic™ Reconstructed Human Epidermis (RHE) skin corrosion test method to 0.5 cm^2 tissue sample, *Toxicology in Vitro,* Vol. 24/5, pp. 1379-1385.

(21) EC-ECVAM (2006), Statement on the application of the SkinEthic™ human skin model for skin corrosivity testing, issued by the ECVAM Scientific Advisory Committee (ESAC25), 17 November 2006. Available at: http://ecvam.jrc.ec.europa.eu.

(22) EC-ECVAM (2009), ESAC statement on the scientific validity of an *in-vitro* test method for skin corrosivity testing: the EST-1000, issued by the ECVAM Scientific Advisory Committee (ESAC30), 12 June 2009. Available at: http://ecvam.jrc.ec.europa.eu.

(23) OECD (2013), "Summary Document on the Statistical Performance of Methods in OECD Test Guideline 431 for Sub-categorisation", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 190, OECD Publications, Paris.

(24) Alépée, N., M.H. Grandidier, J. Cotovio (2014), Sub-categorisation of skin corrosive chemicals by the EpiSkin™ reconstructed human epidermis skin corrosion test method according to UN GHS: Revision of OECD Test Guideline 431, *Toxicology in Vitro,* Vol. 28/2, pp. 131-145.

(25) Worth, A.P. et al. (1998), An Evaluation of the Proposed OECD Testing Strategy for Skin Corrosion, *ATLA*, Vol. 26, pp. 709-720.

(26) Eskes, C. et al. (2012), Regulatory assessment of *in vitro* skin corrosion and irritation data within the European framework: Workshop recommendations, *Regulatory Toxicology and Pharmacology*, Vol. 62/2, pp. 393-403.

(27) Mosmann, T. (1983), Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays, *Journal of Immunological Methods*, Vol. 65/1, pp. 55-63.

(28) Tinois, E. et al. (1994), The Episkin model: Successful reconstruction of human *epidermisin vitro*, in *In vitro Skin Toxicology*, Rougier, A., A.M. Goldberg, H.I. Maibach (eds.), Mary Ann Liebert, Inc., pp. 133-140.

(29) Cannon, C. L. et al. (1994), New epidermal model for dermal irritancy testing, *Toxicology in Vitro*, Vol. 8/4, pp. 889 - 891.

(30) Ponec, M. et al. *(*2000), Lipid and ultrastructural characterization of reconstructed skin models, *International Journal of Pharmaceutics*, Vol. 203/1-2, pp. 211 - 225.

(31) Tinois, E. et al. (1991), *In vitro* and post –transplantation differentiation of human keratinocytes grown on the human type IV collagen film of a bilayered dermal substitute, *Experimental Cell Research*, Vol. 193/2, pp. 310-319.

(32) Parenteau, N.L. et al. (1992), The organotypic culture of human skin keratinocytes and fibroblasts to achieve form and function, *Cytotechnology*, Vol. 9, pp. 163-171.

(33) Wilkins, L.M. et al. (1994), Development of a bilayered living skin construct for clinical applications, *Biotechnology and Bioengineering*, Vol. 43/8, pp. 47-756.

(34) EpiSkin™ SOP, *INVITTOX* Protocol No. 118 (December 2011), EpiSkin™ Skin Corrosivity Test. Available at: http://ecvam.jrc.ec.europa.eu.

(35) EpiDerm™ SOP, Version MK-24-007-0024 (February 2012), Protocol for: *In vitro*EpiDerm™ skin corrosion test (EPI-200-SCT), For use with MatTek Corporation's reconstructed human epidermal model EpiDerm. Available at: http://ecvam.jrc.ec.europa.eu.

(36) SkinEthic™ RHE SOP, *INVITTOX* Protocol (January 2012), SkinEthic™ Skin Corrosivity Test. Available at: http://ecvam.jrc.ec.europa.eu.

(37) epiCS® SOP, Version 4.1 (January 2012), *In Vitro* Skin Corrosion: Human Skin Model Test Epidermal Skin Test 1000 (epiCS®) CellSystems. Available at: http://ecvam.jrc.ec.europa.eu.

ANNEX 1

PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* RECONSTRUCTED HUMAN *EPIDERMIS* (RHE) TEST METHODS FOR SKIN CORROSION[2]

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 demonstrate to have sufficient reliability and relevance for specific testing purposes. The PS, based on valid and accepted test methods, can be used to evaluate the reliability and relevance of other analogous 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 (8). On the other hand, modified test methods, which propose potential improvements to an approved test method, should 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 established PS (9).

2. Similar (me-too) or modified test methods proposed for use under this Test Guideline should be evaluated to determine their reliability and relevance using Reference Substances (Table 1) representing the full range of the TG 404 *in vivo* corrosivity scores, *i.e.*, Corrosive (UN GHS Category 1A and Category 1B and 1C) and non-corrosive chemicals (1). The proposed similar or modified test methods should have reliability and predictive capacity, which are comparable or better than those derived from the two VRM [EpiSkin™ (SM) and EpiDerm™ SCT (EPI-200)] and as described in paragraphs 6 to 10 of this Annex (Tables 2 and 3) (11) (12) (24). The reliability of the new or modified test method, as well as its ability to correctly identify non-corrosive and corrosive chemicals, and possibly also to discriminate UN GHS Category 1A from Category 1B and 1C corrosive chemicals, should be determined prior to its use for testing chemicals.

3. These PS are based on the US-ICCVAM PS (7) for evaluating the validity of new or modified RhE test methods. The PS consists of (8): (i) essential test method components; (ii) recommended Reference Substances, and; (iii) defined reliability and accuracy values that the proposed test method should meet or exceed.

[2]Proposed new or modified test methods following the PS of this Test Guideline should be submitted to the OECD for adoption and inclusion into the Test Guideline before being used for regulatory purposes.

ESSENTIAL TEST METHOD COMPONENTS

4. These consist of essential structural, functional, and procedural elements of a validated test method 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 VRMs (6). The essential test method components are described in detail in paragraphs 16 to 32 of the Test Guideline:

The general conditions (paragraph 16)

The functional conditions, which include:

- Viability (paragraph 17)
- Barrier function (paragraph 18)
- Morphology (paragraph 19)
- Reproducibility (paragraph 20)
- Quality control (paragraph 21)

The procedural conditions (paragraphs 22 to 32)

For specific parameters (*e.g.*, for Tables 2, 3, 4, 5, and 6), adequate values should be provided for any new similar or modified test method, these specific values may vary depending on the specific test method.

MINIMUM LIST OF REFERENCE SUBSTANCES

5. Reference Substances 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 VRMs, or representing a minor modification of one of the VRMs, are comparable or better than those of the VRMs (11) (12) (24). The 30 recommended Reference Substances listed in Table 1 include substances representing different chemical classes (*i.e.* chemical categories based on functional groups), and are representative of the full range of TG 404 *in vivo* scores. The substances included in this list comprise 10 UN GHS Category 1A, 10 UN GHS Category 1B and 1C (the *in vivo* data do not permit distinction between the two categories) and 10 non-corrosive substances. The substances listed in Table 1 are selected from the substances used in the validation study of the VRMs, with regard to chemical functionality and physical state (10) (11) (12) (24). These Reference Substances represent the minimum number of chemicals that should be used to evaluate the reliability and relevance of a proposed similar or modified test method able to discriminate between Category 1A, Category 1B and 1C and non-corrosive substances and mixtures (1A vs. 1B and 1C vs. NC), in accordance with the UN GHS (1) . For similar or modified test methods able to discriminate corrosive from non-corrosive substances and mixtures but not able to sub-categorize corrosive chemicals (C vs. NC), only 20 of the 30 substances listed in Table 1 (the ones not in *italics*) need to be evaluated (5 UN GHS Category 1A, 5 UN GHS Category 1B and 1C and 10 non-corrosive substances). The use of these Reference Substances for the development/optimization of new similar test methods should be avoided to the extent possible. In situations where a listed substance is

unavailable, other substances for which adequate *in vivo* reference data are available could be used, primarily from the substances used in the validation study of the VRMs. If desired, additional substances representing other chemical classes and for which adequate *in vivo* reference data are available may be added to the minimum list of Reference Substances to further evaluate the accuracy of the proposed test method.

Table 1: Minimum list of Reference Substances for determination of Reliability and Predictive Capacity for similar or modified *in vitro*RhE-based skin corrosion test methods. The 20 chemicals NOT in *italics* should be tested with similar or modified test methods proposed to discriminate Corrosive from Non-Corrosive chemicals (without sub-categorization). Additional reference substances should be tested with similar or modified test methods proposed to identify Cat. 1A, a combination of Category 1B and 1C (referred to as 1B/1C below) and non-corrosive chemicals. These additional reference substances are indicated *in italics*.

Chemical[1]	CASRN	Chemical Class[2]	Physical State	EpiSkin™ [4]	EpiDerm™ [4]	SkinEthic™ [4]	epiCS® [4]
Non-corrosive chemicals based on *in vivo* results[3]							
Phenethyl bromide*	103-63-9	Electrophile	L	(3) NC	(3) NC	(3) NC	(2) NC
4-Amino-1,2,4-triazole	584-13-4	Organic base	S	(3) NC	(3) NC	(3) NC	(2) NC
4-(methylthio)-benzaldehyde*	3446-89-7	Electrophile	L	(3) NC	(3) NC	(3) NC	(2) NC
Lauric acid	143-07-7	Organic acid	S	(3) NC	(3) NC	(3) NC	(2) NC
1,9-Decadiene	1647-16-1	Neutral organic	L	(3) NC	(3) NC	(3) NC	(2) NC
2,4-Dimethylaniline	95-68-1	Organic base	L	(2) NC (1) 1B/1C	(1) NC (2) 1B/1C	(2) 1B/1C (1) 1A	(1) NC (1) 1B/1C
3,3-Dithiopropionic acid	1119-62-6	Organic acid	S	(3) NC	(3) NC	(3) NC	(2) NC
Methyl palmitate	112-39-0	Neutral organic	S	(3) NC	(3) NC	(3) NC	(2) NC
2-Hydroxyiso-butyric acid	594-61-6	Organic acid	S	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C
Sodium undecylenate (33%)	3398-33-2	Soap / Surfactant	L	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C
UN GHS Cat. 1B-and-1C based on *in vivo* results[3]							
Glyoxylic acid monohydrate	563-96-2	Organic acid	S	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C
Lactic acid	598-82-3	Organic acid	L	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C
Sodium bisulphate monohydrate	10034-88-5	Inorganic salt	S	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C (1) NC	(2) 1B/1C
Ethanolamine*	141-43-5	Organic base	Viscous	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C
60/40 Octanoic/decanoic acid	*68937-75-7*	*Organic acid*	*L*	*(3) 1B/1C*	*(3) 1B/1C*	*(3) 1B/1C*	*(2) 1B/1C*
Hydrochloric acid (14.4%)	7647-01-0	Inorganic acid	L	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C
Fluoroboric acid	*16872-11-0*	*Inorganic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Propionic acid	*79-09-4*	*Organic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
*2-tert-Butylphenol**	*88-18-6*	*Phenol*	*L*	*(3) 1B/1C*	*(3) 1A*	*(3) 1A*	*(2) 1A*

*Cyclohexyl amine**	*108-91-8*	*Organic base*	*L*	*(3) 1B/1C*	*(3) 1A*	*(3) 1A*	*(2) 1A*
UN GHS Cat. 1A based on *in vivo* results[3]							
Acrylic acid	*79-10-7*	*Organic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Bromoacetic acid	79-08-3	Organic acid	S	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Boron trifluoride dehydrate	13319-75-0	Inorganic acid	L	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Phenol	108-95-2	Phenol	S	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Phosphorus tribromide	*7789-60-8*	*Inorganic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Silver nitrate	*7761-88-8*	*Inorganic salt*	*S*	*(1) 1A (2) 1B/1C*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Formic acid	*64-18-6*	*Organic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Dichloroacetyl chloride	79-36-7	Electrophile	L	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Sulphuric acid (98%)	*7664-93-9*	*Inorganic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
N,N-Dimethyl dipropylene triamine*	10563-29-8	Organic base	L	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; UN GHS = United Nations Globally Harmonized System (1); NC = Not Corrosive

[1]These substances, sorted first by corrosives versus non-corrosives, then by corrosive sub-category, were selected from the substances used in the ECVAM validation studies of EpiSkin™ and EpiDerm™ SCT (10) (11) (12) and from post-validation studies based on data generated by EpiSkin™ (24), EpiDerm™, SkinEthic™ and epiCS® developers. Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (10) (12). The selection includes, to the extent possible, substances that: (i) are representative of the range of corrosivity responses (*e.g.* non-corrosives; weak to strong corrosives) that the VRMs are capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation studies; (iii) reflect the performance characteristics of the VRM; (iv) have chemical structures that are well-defined; (v) induce reproducible results in the VRM; (vi) induce definitive results in the *in vivo* reference test method; (vii) are commercially available; and (viii) are not associated with prohibitive disposal costs. Chemicals marked with an * are potential direct MTT reducers.

[2]Chemical class assigned by Barratt *et al.* (1998) (10).

[3]The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS 1A, 1B and 1C.

[4]The *in vitro* predictions reported in this table were obtained with the various test methods during post-validation testing performed by the test method developers. These predictions were corrected for direct MTT reduction using killed control tissues.

DEFINED RELIABILITY AND ACCURACY VALUES

6. For purposes of establishing the reliability and relevance of proposed similar or modified RhE test methods to be used by several independent laboratories, all 30 Reference Substances listed in Table 1 should be tested in at least three laboratories (24 Reference Substances for methods not able to sub-categorize corrosive chemicals). It is however essential that all PS-based validation studies are independently assessed by internationally recognized validation bodies, in agreement with international guidelines (8). In each laboratory, all relevant Reference Substances should be tested for each exposure time in three independent runs performed with different tissue batches and at sufficiently spaced time points. Each run should consist of at least two concurrently tested tissue replicates per exposure time for each test chemical, negative control, PC and adapted controls for direct MTT reduction and/or colour interference.

7. The calculation of the reliability and predictive capacity of the proposed test method should be done according to the rules described below to ensure that a predefined and consistent approach is used:

1. Within-laboratory reproducibility (WLR) should be calculated based on concordance of classifications using only qualified tests obtained with Reference Substances for which at least two qualified tests are available. In addition, it should be reported the number and identity of the Reference Substances which per laboratory have none or only one qualified test (omitted from WLR calculations), as well as how many and which Reference Substances per laboratory have two or three qualified tests (used for WLR calculations).

2. For the calculation of between-laboratory reproducibility (BLR) the final classification for each Reference Chemical in each participating laboratory should be obtained by using the arithmetic mean value of viability over the different qualified tests performed. BLR should be calculated based on concordance of classifications using only qualified tests from Reference Substances for which at least one qualified test per laboratory is available. It should be reported how many and which Reference Substances do not have at least one qualified test per laboratory (omitted from BLR calculations), as well as how many and which Reference Substances have 3, 4, 5, 6, 7, 8 or 9 qualified tests that can be used to calculate BLR (with at least one qualified test per laboratory).

3. The calculation of predictive capacity (e.g. sensitivity, specificity and accuracy for C vs. NC) should be done using all qualified tests obtained for each Reference Chemical in each laboratory. The calculations should be based on the individual predictions of each qualified test for each Reference Chemical in each laboratory and not on the arithmetic mean values of viability over the different qualified tests performed.

In this context, a qualified test consists of a test that meets the criteria for an acceptable test, as defined in the corresponding SOP, and is within a qualified run. Otherwise, the test is considered as non-qualified. A qualified run consists of a run that meets the test acceptance criteria for the NC and PC, as defined in the corresponding SOP. Otherwise, the run is considered as non-qualified.

Within-laboratory reproducibility

8. An assessment of within-laboratory reproducibility for similar or modified test method proposed

to discriminate corrosive from non-corrosive chemicals (but not to sub-categorize corrosive chemicals) should show a concordance of classifications (corrosive or non-corrosive) obtained in different, independent tests of the 24 relevant Reference Substances within one single laboratory equal or higher (≥) than 90% (actual for EpiSkin™: 100%, 100% and 96% in each laboratory, respectively). An assessment of within-laboratory reproducibility for similar or modified test method proposed to discriminate between Category 1A, Category 1B and 1C and non-corrosive chemicals should show a concordance of classifications (Category 1A, Category 1B and 1C or non-corrosive) obtained in different, independent tests of the 30 Reference Substances within one single laboratory equal or higher (≥) than 80% (actual for EpiSkin™: 96%, 96% and 88% in each laboratory, respectively).

Between-laboratory reproducibility

9. For similar or modified test methods proposed to discriminate corrosive from non-corrosive chemicals (but not to sub-categorize corrosive chemicals), the concordance of classifications (corrosive or non-corrosive) between a minimum of three laboratories, obtained for the 24 relevant Reference Substances, should be equal or higher (≥) than 80% (actual for EpiSkin™: 88%). For similar or modified test methods proposed to discriminate between Category 1A, Category 1B and 1C and non-corrosive chemicals, the concordance of classifications (Category 1A, Category 1B and 1C or non-corrosive) between a minimum of three laboratories, obtained for the 30 Reference Substances, should be equal or higher (≥) than 70% (actual for EpiSkin™: 80%).

Predictive capacity

10. The predictive capacity of the proposed similar or modified test method should be comparable or better to that of the VRMs. For similar or modified test method proposed to discriminate corrosive from non-corrosive chemicals (C vs. NC) but not to sub-categorize corrosive chemicals, the sensitivity and specificity obtained with the 20 relevant Reference Substances (Table 1) should be equal or higher (≥) than 95% and 70%, respectively, and the accuracy should be equal or higher (≥) than 82.5% (Table 2). For similar or modified test method proposed to discriminate between Category 1A, Category 1B-and-1C and non-corrosive chemicals (Cat. 1A vs. Cat. 1B-and-1C vs. NC), the minimum predictive capacity values that should be obtained with the 30 Reference Substances (Table 1) are indicated in Table 3 for RhE test methods similar to EpiSkin™ and for RhE test methods similar to EpiDerm™.

Table 2: Required sensitivity, specificity and accuracy for similar or modified RhE test methods to be considered valid to discriminate corrosive from non-corrosive chemicals (C vs. NC) but not able to sub-categorize corrosive chemicals. Values are based on the results of the two VRMs (EpiSkin™ and EpiDerm™) for the 20 Reference Substances not in italics from Table 1.

Sensitivity	Specificity	Accuracy
≥ 95% (actual for EpiSkin™: 100%; actual for EpiDerm™: 100%)	≥ 70% (actual for EpiSkin™: 76.7%; actual for EpiDerm™: 73.3%)	≥ 82.5% (actual for EpiSkin™: 88.3%; actual for EpiDerm™: 86.7%)

Table 3: Required predictive capacity for similar or modified RhE test method to be considered valid to discriminate between Cat. 1A, a combination of Category 1B and 1C (referred to as 1B-and-1C below) and non-corrosive chemicals (Cat. 1A *vs.* Cat. 1B-and-1C *vs.* NC)*. Values are based on the results of the two VRMs (EpiSkin™ and EpiDerm™) for the 30 Reference Substances (see table 1).

VRM	EpiSkin™	EpiDerm™
Sensitivity (C vs NC)	≥ 95% (actual for EpiSkin™: 100.0%)	≥ 95% (actual for EpiDerm™: 100.0%)
Correctly classified 1A	≥ 80% (actual for EpiSkin™: 83.3%)	≥ 90% (actual for EpiDerm™: 90.0%)
1A Underclassified 1B-and-1C	≤ 20% (actual for EpiSkin™: 16.7%,)	≤ 10% (actual for EpiDerm™: 10.0%,)
1A Underclassified NC	0% (actual for EpiSkin™: 0.0%)	0% (actual for EpiDerm™: 0.0%)
Correctly classified 1B-and-1C	≥ 80% (actual for EpiSkin™: 80.0%)	≥ 55% (actual for EpiDerm™: 60.0%)
1B-and-1C Overclassified 1A	≤ 20% (actual for EpiSkin™: 20.0%)	≤ 45% (actual for EpiDerm™: 40.0%)
1B and 1C Underclassified NC	≤ 5% (actual for EpiSkin™: 0.0%)	≤ 5% (actual for EpiDerm™: 0.0%)
Specificity	≥ 70% (actual for EpiSkin™: 76.7%)	≥ 70% (actual for EpiDerm™: 73.3%)
NC Overclassified 1A	≤ 5% (actual for EpiSkin™: 0.0%)	≤ 5% (actual for EpiDerm™: 0.0%)
NC Overclassified 1B-and-1C	≤ 30% (actual for EpiSkin™: 23.3%)	≤ 30% (actual for EpiDerm™: 26.7%)
Accuracy (C vs. NC)	≥ 87% (actual for EpiSkin™: 92.2%)	≥ 87% (actual for EpiDerm™: 91.1%)
Accuracy (1A vs. 1B-and-1C vs. NC)	≥ 78% (actual for EpiSkin™: 80.0%)	≥ 72% (actual for EpiDerm™: 74.4%)

* Depending on the results obtained with a similar or modified RhE test method for the 30 Reference Substances, it may be considered similar to EpiSkin™ or similar to EpiDerm™ for the purpose of this Test Guideline. The EpiSkin™ and EpiDerm™ test methods are able to sub-categorize (i.e. 1A versus 1B-and-1C versus NC) but differences are observed (SkinEthic™ and epiCS® are considered similar to EpiDerm™). For RhE test methods that demonstrate similarity to EpiSkin™ , results can be directly used based on the outcoming predictions. For RhE test methods that demonstrate similarity to EpiDerm™, chemicals that are classified as Category 1B-and-1C can be considered as Category 1B-and-1C, whereas chemicals for which cell viability at 3 minutes is below 50% should be considered as Category 1, since the Category 1A predictions of these three test methods contain a high rate of over-predictions of chemicals of Categories 1B and 1C (see also paragraph 7 of the Test Guideline). The regulatory framework in member countries will decide how this Test Guideline will be used, e.g. acknowledging the significant probability of overclassification, a Category 1A classification may still be accepted or further testing may be conducted to confirm the result.

Study Acceptance Criteria

11. It is possible that one or several tests pertaining to one or more test chemical and control substance does/do not meet the test acceptance criteria or is/are not acceptable for other reasons (non-

qualified tests). To complement missing data, a maximum of two additional tests for each test chemical is admissible per laboratory ("retesting"). More precisely, since in case of retesting also PC and NC have to be concurrently tested, a maximum number of two additional runs may be conducted for each test chemical in each laboratory. Importantly, each laboratory should not produce more than three qualified tests per test chemical. Excess production of data and subsequent data selection are regarded as not appropriate.

12. It is conceivable that even after retesting, three qualified tests are not obtained for every Reference Chemical in every participating laboratory, leading to an incomplete data matrix. In such cases the following three criteria should all be met in order to consider the datasets acceptable for purposes of PS-based validation studies:

1. All relevant Reference Substances (24 for Category 1 vs. Non Corrosive; 30 for Cat. 1A vs. Cat. 1B and 1C vs. Non Corrosive) should have at least one complete test sequence in one laboratory.
2. Each of the at least three participating laboratories should have a minimum of 85% complete test sequences (for 24 Reference Substances: 3 incomplete test sequences are allowed per laboratory; for 30 Reference Substances: 4 incomplete test sequences are allowed per laboratory).
3. At least 90% of all test sequences from at least three laboratories need to be complete (for 24 Reference Substances tested in 3 laboratories: a total of 7 incomplete test sequences are allowed; for 30 Reference Substances tested in 3 laboratories: a total of 9 incomplete test sequences are allowed).

In this context, a test sequence consists of the total number of independent tests performed for a single Reference Chemical in a single laboratory, including any re-testing (a total of 3 to 5 tests). A complete test sequence consists of a test sequence containing three qualified tests. A test sequence containing less than 3 qualified tests is considered as incomplete.

ANNEX 2

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 (9).

Cell viability: Parameter measuring total activity of a cell population *e.g.* as ability of cellular mitochondrial dehydrogenases to reduce the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue), which depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Chemical: means a substance or a mixture

Complete test sequence: A test sequence containing three qualified tests. A test sequence containing less than 3 qualified tests is considered as incomplete (see also definition of "test sequence" below).

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 (9).

ET_{50}: Can be estimated bydetermination of the exposure time required to reduce cell viability by 50% upon application of the marker chemical at a specified, fixed concentration, see also IC_{50}.

IC_{50}: Can be estimated by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, see also ET_{50}.

Infinite dose: Amount of test chemical applied to the *epidermis* exceeding the amount required to completely and uniformly cover the *epidermis* surface.

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. Interchangeably used with similar test method (9).

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

MTT: 3-[4,5-dimethylthiazol-2-yl]-2, 5-diphenyltetrazolium bromide)

NC: Non corrosive

OD: Optical Density

PC: Positive Control

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 Substances 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 Substances (9).

Qualified run: A run that meets the test acceptance criteria for the NC and PC, as defined in the corresponding SOP. Otherwise, the run is considered as non-qualified.

Qualified test: A test that meets the criteria for an acceptable test, as defined in the corresponding SOP, and is within a qualified run. Otherwise, the test is considered as non-qualified.

Reference Substances: 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 Substances may be required for the different stages of the validation process, and for different test methods and test uses (9).

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 (9).

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 (9).

Run: A run consists of one or more test chemicals tested concurrently with a negative control and with a PC.

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 (9).

Skin corrosion *in vivo:* The production of irreversible damage of the skin; namely, visible necrosis through the *epidermis* and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at

14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

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 (9).

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.

Test: A single test substance concurrently tested in a minimum of two tissue replicates as defined in the corresponding SOP.

Test sequence: The total number of independent tests performed for a single test substance in a single laboratory, including any re-testing. A test sequence may include both qualified and non-qualified tests.

Test chemical: means what is being tested

Tiered testing strategy: Testing which uses test methods in a sequential manner; the test methods selected in each succeeding level are determined by the results in the previous level of testing (9).

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 (1).

ANNEX 3

MAIN TEST METHOD COMPONENTS OF THE RhE TEST METHODS VALIDATED FOR SKIN CORROSION TESTING

Test Method Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Model surface	0.38 cm^2	0.63 cm^2	0.5 cm^2	0.6 cm^2
Number of tissue replicates	At least 2 per exposure time	2-3 per exposure time	At least 2 per exposure time	At least 2 per exposure time
Treatment doses and application	Liquids and viscous: 50 µL ± 3 µL (131.6 µL/cm2) Solids: 20 ± 2 mg (52.6 mg/cm2) + 100 µL ± 5µL NaCl solution (9 g/L) Waxy/sticky: 50 ± 2 mg (131.6 mg/cm2) with a nylon mesh	Liquids: 50 µL (79.4 µL/cm^2) with or without a nylon mesh *Pre-test compatibility of test chemical with nylon mesh* Semisolids: 50 µL (79.4 µL/cm^2) Solids: 25 µL H_2O (or more if necessary) + 25 mg (39.7 mg/cm^2) Waxes: flat "disc like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15 µL H_2O.	Liquids and viscous: 40 µL ± 3µl (80 µL/cm^2) using nylon mesh *Pre-test compatibility of test chemical with nylon mesh* Solids: 20 µL ± 2µl H_2O + 20 ± 3 mg (40 mg/cm^2) Waxy/sticky: 20 ± 3 mg (40 mg/cm^2) using nylon mesh	Liquids: 50 µL (83.3 µL/cm^2) using nylon mesh *Pre-test compatibility of test chemical with nylon mesh* Semisolids: 50 µL (83.3 µL/cm^2) Solids: 25 mg (41.7 mg/cm^2) + 25 µL H_2O (or more if necessary) Waxy: flat "cookie like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15 µL H_2O
Pre-check for direct MTT reduction	50 µL (liquid) or 20 mg (solid) + 2 mL MTT 0.3 mg/mL solution for 180 ± 5 min at 37°C, 5% CO_2, 95% RH → if solution turns blue/purple, water-killed adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 1 mL MTT 1 mg/mL solution for 60 min at 37°C, 5% CO_2, 95% RH → if solution turns blue/purple, freeze-killed adapted controls should be performed	40 µL (liquid) or 20 mg (solid) + 1 mL MTT 1 mg/mL solution for 180± 15 min at 37°C, 5% CO_2, 95% RH → if solution turns blue/purple, freeze-killed adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 1 mL MTT 1 mg/mL solution for 60 min at 37°C, 5% CO_2, 95% RH → if solution turns blue/purple, freeze-killed adapted controls should be performed

Test Method Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Pre-check for colour interference	10 µL (liquid) or 10 mg (solid) + 90 µL H_2O mixed for 15 min at RT → if solution becomes coloured, living adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 300 µL H_2O for 60 min at 37°C, 5% CO_2, 95% RH → if solution becomes coloured, living adapted controls should be performed	40 µL (liquid) or 20mg (solid) + 300 µL H_2O mixed for 60 min at RT → if test chemical is coloured, living adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 300 µL H_2O for 60 min at 37°C, 5% CO_2, 95% RH → if solution becomes coloured, living adapted controls should be performed
Exposure time and temperature	3 min, 60 min (± 5 min) and 240 min (± 10 min) In ventilated cabinet Room Temperature (RT, 18-28°C)	3 min at RT, and 60 min at 37°C, 5% CO_2, 95% RH	3 min at RT, and 60 min at 37°C, 5% CO_2, 95% RH	3 min at RT, and 60 min at 37°C, 5% CO_2, 95% RH
Rinsing	25 mL 1x PBS (2 mL/throwing)	20 times with a constant soft stream of 1x PBS	20 times with a constant soft stream of 1x PBS	20 times with a constant soft stream of 1x PBS
Negative control	50 µL NaCl solution (9 g/L) Tested with every exposure time	50 µL H_2O Tested with every exposure time	40 µL H_2O Tested with every exposure time	50 µL H_2O Tested with every exposure time
Positive control	50 µL Glacial acetic acid Tested only for 4 hours	50 µL 8N KOH Tested with every exposure time	40 µL 8N KOH Tested only for 1 hour	50 µL 8N KOH Tested with every exposure time
MTT solution	2 mL 0.3 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL

Test Method Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
MTT incubation time and temperature	180 min (± 15 min) at 37°C, 5% CO_2, 95% RH	180 min at 37°C, 5% CO_2, 95% RH	180 min (± 15 min) at 37°C, 5% CO_2, 95% RH	180 min at 37°C, 5% CO_2, 95% RH
Extraction solvent	500 µL acidified isopropanol (0.04 N HCl in isopropanol) (isolated tissue fully immersed)	2 mL isopropanol (extraction from top and bottom of insert)	1.5 mL isopropanol (extraction from top and bottom of insert)	2 mL isopropanol (extraction from top and bottom of insert)
Extraction time and temperature	Overnight at RT, protected from light	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT
OD reading	570 nm (545 - 595 nm) without reference filter	570 nm (or 540 nm) without reference filter	570 nm (540 - 600 nm) without reference filter	540 - 570 nm without reference filter
Tissue Quality Control	18 hours treatment with SDS 1.0 mg/mL ≤ IC_{50} ≤ 3.0 mg/mL	Treatment with 1% Triton X-100 4.08 hours ≤ ET_{50} ≤ 8.7 hours	Treatment with 1% Triton X-100 4.0 hours ≤ ET_{50} ≤ 10.0 hours	Treatment with 1% Triton X-100 2.0 hours ≤ ET_{50} ≤ 7.0 hours

Test Method Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Acceptability Criteria	1. Mean OD of the tissue replicates treated with the negative control (NaCl) should be ≥ 0.6 and ≤ 1.5 for every exposure time 2. Mean viability of the tissue replicates exposed for 4 hours with the positive control (glacial acetic acid), expressed as % of the negative control, should be ≤ 20% 3. In the range 20-100% viability and for ODs≥ 0.3, difference of viability between the two tissue replicates should not exceed 30%.	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 2.8 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be < 15% 3. In the range 20 - 100% viability, the Coefficient of Variation (CV) between tissue replicates should be ≤ 30%	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 3.0 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour (and 4 hours, if applicable) with the positive control (8N KOH), expressed as % of the negative control, should be < 15% 3. In the range 20-100% viability, and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not exceed 30%	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 2.8 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be < 20% 3. In the range 20-100% viability, and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not exceed 30%

ANNEX 4

PERFORMANCE OF TEST METHODS FOR SUB-CATEGORISATION

This annex provides a table where performances of the four test methods were calculated based on a set of 80 chemicals tested by the four test developers. Calculations were performed by the OECD Secretariat, reviewed and agreed by an expert subgroup (23).

EpiSkin™, EpiDerm™ , SkinEthic™ and epiCS® test methods are able to sub-categorize (i.e. 1A versus 1B-and-1C versus NC) but differences are observed between EpiSkin™ and the three other test methods, EpiDerm™, SkinEthic™ and epiCS®, for sub-categorization. Results from EpiSkin™ can be directly used based on the outcoming results, whereas results from EpiDerm™, SkinEthic™ and epiCS®, should take into account high over-classification rates from those three test methods for 1B-and-1C category (see table 4.0 in Annex 4). Therefore, for EpiDerm™, SkinEthic™ and epiCS®, chemicals that are classified as 1B-and-1C can be considered as 1B-and-1C, and chemicals for which cell viability at 3 minutes is below 50% should be considered as 1, that is to say that either under the prediction principle they could be claimed as 1A or they should undergo further testing to be possibly confirmed as 1B-and-1C. The regulatory framework in member countries will decide how this Test Guideline will be used.

Table 4.0: Performances, Overclassification rates, Underclassification rates, and Accuracy (Predictive capacity) of the four test methods based on a set of 80 chemicals all tested over 2 or 3 runs in each test method.

STATISTICS ON ENTIRE SET OF CHEMICALS (n= 80 chemicals tested over 2 or 3 runs, i.e. 159* or 240 classifications) *one chemical was tested once because of no availability				
	EpiSkin™	**EpiDerm™**	**SkinEthic™**	**epiCS®**
Overclassifications:				
Cat. 1BC chemicals that are overclassified 1A	21.50%	41.94%	46.24%	45.90%
Cat. NC chemicals that are overclassified 1B/1C	20.72%	23.42%	24.32%	28.38%
Cat. NC chemicals that are overclassified 1A	0.00%	2.70%	2.70%	0.00%
Cat. NC chemicals that are overclassified Corr.	20.72%	26.13%	27.03%	28.38%
Global overclassification rate (all categories)	17.92%	28.33%	30.42%	30.82%
Underclassifications:				
Cat. 1A Underclassified 1B/1C	16.67%	8.33%	13.89%	8.33%
Cat. 1A Underclassified NC	0.00%	0.00%	0.00%	0.00%
Cat. 1BC Underclassified NC	2.15%	0.00%	7.53%	6.56%
Global Underclassification rate (all categories)	3.33%	2.47%	5.00%	3.77%
Correct Classifications:				
1A Correctly classified	83.33%	91.67%	86.11%	91.67%
1B/1C Correctly classified	76.34%	58.06%	46.24%	47.54%
NC Correctly classified	79.28%	73.87%	72.97%	71.62%
Accuracy (Predictive capacity)	78.75%	70.42%	64.58%	65.41%

NC=non corrosive.