Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

23-Jul-2015

English - Or. English

ENV/JM/MONO(2015)26
Unclassified

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO RECONSTRUCTED HUMAN EPIDERMIS (RHE) TEST METHODS FOR SKIN CORROSION TESTING AS DESCRIBED IN TG 431

(Intended for the developers of new or modified similar test methods)
Series on Testing & Assessment
No. 219

JT03380410

Complete document available on OLIS in its original format

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 219

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

AS DESCRIBED IN TG 431[1]

(Intended for the developers of new or modified similar test methods)

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

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2015

[1]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.

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

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

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

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

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document contains the Performance Standards (PS) for the validation of similar or modified RhE methods for skin corrosion testing as described in TG 431. In the past, PS were usually annexed to TGs. However, in view of separating information on the *use* of a test method as contained in the TG from information needed to *validate* test methods as contained in the PS, TGs and PS will now both be stand-alone documents. This approach had been agreed by the Working Group of the National Coordinators of the Test Guidelines Programme (WNT). In case of the current PS for skin *in vitro* corrosion methods according to TG 431, the text was reviewed in regard to harmonising with other relevant documents addressing skin irritation and skin corrosion. The PS were reviewed by the OECD Expert Group on Skin Irritation/Corrosion in November 2014. The PS are intended for the developers of new or modified similar test methods to the validated reference method. The present document was approved by the WNT in April 2015, declassified and published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology on 10 July 2015.

INTRODUCTION

1. This document contains Performance Standards which allow, in accordance with the principles of Guidance Document No. 34 (1), determining the validation status (reliability and relevance) of similar and modified skin corrosion test methods that are structurally and mechanistically similar to the RhE test method in OECD Test Guideline 431 (2).

2. These PS include the following sets of information: (i) Essential Test Method Components that serve to evaluate the structural, mechanistic and procedural similarity of a new similar or modified proposed test method, (ii) a list of 30 Reference Chemicals to be used for validating new or modified test methods and (iii) defined target values of reproducibility and predictive capacity that need to be met by proposed test methods in order to be considered similar to the validated reference methods.

3. The purpose of Performance Standards (PS) is to provide the basis by which new similar 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 (i.e., scientifically valid), in accordance with the principles of Guidance Document No. 34 (1). The PS, based on scientifically valid and accepted test method(s), can be used to evaluate the reliability and relevance of test methods that are based on similar scientific principles and measure or predict the same biological or toxic effect (1). Such methods are referred to as *similar* or *"me-too"* test methods. Moreover, the PS may be used to evaluate *modified* test methods, which may propose potential improvements in comparison to approved earlier versions of a method. In such cases the PS can be used to determine the effect of the proposed changes on the test method's performance and the extent to which such changes may affect the information available for other components of the validation process (e.g. relating to Essential Test Method Components). However, depending on the number and nature of the proposed changes as well as the data and documentation available in relation to these changes, modified test methods may: i) either be found unsuitable for a PS-based validation (e.g. if the changes are so substantial that the method is not any longer deemed sufficiently similar with regard to the PS), in which cases they should be subjected to the same validation process as described for a new test method (1), or ii) suitable for a limited assessment of reliability and relevance using the established PS (1). Similar or modified new test methods (i.e., "me-too" tests) successfully validated according to Performance Standards can be added to TG 431. However, Mutual Acceptance of Data (MAD) will only be guaranteed for those test methods reviewed and adopted by the OECD. Proposed similar or modified test methods validated according to these PS should therefore be submitted to the OECD for adoption and inclusion into TG 431 before being used for regulatory purposes.

4. These PS are based on the ICCVAM PS (3) for evaluating the validity of new or modified RhE test methods. The PS consists of: (i) Essential Test Method Components; (ii) Recommended Reference Chemicals, and; (iii) Defined Reliability and Predictive Capacity Values that the proposed similar or modified test method should meet or exceed. The VRMs used as to develop the present PS are the EpiSkin^{TM} (SM) and EpiDerm^{TM} SCT (EPI-200) test methods as described in TG 431 (2). Definitions are provided in Annex I.

5. Similar (me-too) or modified test methods proposed for use under Test Guideline 431 (2) should be evaluated to determine their reliability and predictive capacity using Reference Chemicals representing the full range of the TG 404 *in vivo* corrosivity scores (Table 5) prior to their use for testing new test chemicals, in order to ensure that these methods are able to identify correctly non-corrosive and corrosive chemicals, and possibly also to discriminate UN GHS Sub-category 1A from a combination of Sub-categories 1B and 1C corrosive chemicals (4) (5). The proposed similar or modified test methods should

have reproducibility, sensitivity, specificity and accuracy values which are equal or better than those derived from the two VRM and as described in paragraphs 29 to 32 of these PS (Tables 6 and 7) (6) (7) (8).

ESSENTIAL TEST METHOD COMPONENTS

6. The Essential Test Method Components consist of essential structural, functional, and procedural elements of scientifically valid test methods (the VMRs) 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 (1) (2). The essential test method components to be considered for similar or modified test methods related to TG 431 are described in detail in the following paragraphs.

7. For specific parameters (*e.g.*, for Tables 1, 2, 3 and 4) 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.

General Conditions

8. Non-transformed human keratinocytes should be used to reconstruct the epithelium. The RhE model is prepared in inserts with a porous synthetic membrane through which nutrients can pass to the cells. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum, stratum granulosum*) should be present under a functional *stratum corneum*. The test chemical is applied topically to the three-dimensional RhE model, which should have a surface in direct contact with air so as to allow for an exposure similar to the *in vivo* situation. 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 benchmark chemicals, *e.g.* sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a benchmark 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 benchmark chemical at a specified, fixed concentration (see paragraph 8). The containment properties of the RhE model should prevent the passage of test chemical 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, and fungi.

Functional Conditions

Viability

9. The assay used for quantifying tissue viability is the MTT-assay (9). The viable cells of the RhE tissue construct can reduce the vital dye MTT into a blue MTT formazan precipitate, which is then extracted from the tissue using isopropanol (or a similar solvent). The Optical Density (OD) of the extraction solvent alone should be sufficiently small, i.e., OD < 0.1. The extracted MTT formazan may be quantified using either a standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure (10). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control. An acceptability range (upper and lower limit) for the negative control OD values should be established by the RhE model developer/supplier. Acceptability ranges for the negative control OD values for the RhE VRMs are given in Table 1. An HPLC/UPLC-Spectrophotometry user should use the negative control OD ranges provided in Table 1 as the acceptance criterion for the negative

control. 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 1: Acceptability ranges for negative control OD values to control batch quality of the VRMs

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

Barrier function

10. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of certain cytotoxic benchmark chemicals (*e.g.* SDS or Triton X-100), as estimated by IC_{50} or ET_{50} (Table 2).

Morphology

11. 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

12. Test results of the positive and negative controls of the test method should demonstrate reproducibility of the test method over time. In case of the use of a test method for sub- categorization, the reproducibility with respect to sub-categorization should also be demonstrated.

Quality control (QC)

13. 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 7), *barrier function* (paragraph 8) and *morphology* (paragraph 9) are the most relevant. An acceptability range (upper and lower limit) for the barrier function as measured by the IC_{50} or ET_{50} (see paragraphs 6 and 8) should be established by the RhE model developer/supplier. The acceptability range of the VRMs are given in Table 2. Adequate ranges should be provided for any new similar or modified test method. These may vary depending on the specific test method. Data demonstrating compliance with all production release criteria should be provided by the RhE model developer/supplier. Only results produced with tissues fulfilling all of these production quality release criteria can be accepted for reliable prediction of corrosive classification.

Table 2: QC batch release criteria of the VRMs

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS) (11)	IC_{50} = 1.0 mg/mL	IC_{50} = 3.0 mg/mL
EpiDerm™ SCT (EPI-200) (1% Triton X-100) (12)	ET_{50} = 4.0 hours	ET_{50} = 8.7 hours

Procedural Conditions

Application of the Test Chemical and Control Substances

14. At least two tissue replicates should be used for each test chemical and each control substance for each exposure time in each run. 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). Whenever possible, solids should be tested as a fine powder.

15. Concurrent negative and positive controls (PC) should be used in each run to demonstrate that viability (with negative controls), and sensitivity (with the PC) of the tissues are within a defined historical acceptance range. The concurrent negative control also provides the baseline (100% tissue viability) to calculate the relative percent viability of the tissues treated with the test chemical. The positive control suggested for the VRMs 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 15 and 16. The suggested VRMs negative controls are 0.9% (w/v) NaCl or water.

Cell Viability Measurements

16. The MTT assay, which is a quantitative assay, should be used to measure tissue viability (9). It is compatible with use in a three-dimensional tissue construct. The tissue sample is placed in MTT solution of an appropriate concentration (e.g. 0.3 or 1 mg/mL in the VRMs) for 3 hours. The vital dye MTT is reduced into a blue formazan precipitate by the viable cells of the RhE model. 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 quantified by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm, or by an HPLC/UPLC-spectrophotometry procedure (10). The same procedure should be employed for the concurrently tested negative and positive controls.

17. Optical properties of the test chemical or its chemical action on MTT may interfere with the measurement of MTT formazan leading to a false estimate of tissue viability. 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 chemical absorbs, naturally or due to treatment procedures, in the same OD range of formazan (i.e. 570 ± 30 nm, mainly blue and purple chemicals). Pre-checks should be performed before testing to allow identification of potential direct MTT reducers and/or colour interfering chemicals. The corresponding procedures should be standardised and part of the SOP. Additional controls should be used to correct for a potential interference from these test chemicals such as the non-specific MTT reduction (NSMTT) control and the non-specific colour (NSC) control (see paragraphs 16 to 19). 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. For coloured test chemicals or test chemicals that become coloured in contact with water or isopropanol, which are not compatible with the standard absorbance (OD) measurement due to too strong interference with the MTT assay (i.e., strong absorption at 570 ± 30 nm), an HPLC/UPLC-spectrophotometry procedure to measure MTT formazan may be employed (10). A detailed description of how to correct direct MTT reduction and colour interferences by the test chemical should be available in the test method's SOP. A description of the control measures used in the VRMs are summarised in paragraphs 16 to 19 below.

18. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT medium. If the MTT mixture containing the test chemical turns blue/purple, the test chemical is presumed to directly reduce the MTT, and further functional check on non-viable epidermis should be performed, independently of using the standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry

procedure. This additional functional check employs killed tissues (by e.g., exposure to low temperature ("freeze-killed" tissues) or by other means) that possess only residual metabolic activity but absorb and retain the test chemical in a similar way as viable tissues. Each MTT reducing chemical is applied on at least two killed tissue replicates per exposure time, which undergo the entire testing procedure. The true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the MTT reducer **minus** the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT).

19. To identify potential interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol and decide on the need for additional controls, spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, further colorant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required (see paragraph 19). When performing the standard absorbance (OD) measurement, each interfering coloured test chemical should be applied on at least two viable tissue replicates per exposure time, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step to generate a non-specific colour (NSC_{living}) control. The NSC_{living} control needs to be performed concurrently per exposure time to the testing of the coloured test chemical (in each run) due to the inherent biological variability of living tissues. The true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the interfering test chemical and incubated with MTT solution **minus** the percent non-specific colour obtained with living tissues exposed to the interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected ($\%NSC_{living}$).

20. Test chemicals that are identified as producing both direct MTT reduction (see paragraph 16) and colour interference (see paragraph 17) should also require a third set of controls when performing the standard absorbance (OD) measurement, apart from the NSMTT and NSC_{living} controls described in the previous paragraphs. This is usually the case with darkly coloured test chemicals interfering with the MTT assay (e.g., blue, purple, black) because their intrinsic colour impedes the assessment of their capacity to directly reduce MTT as described in paragraph 16. These test chemicals may be retained in both living and killed tissues and therefore the NSMTT control may not only correct for potential direct MTT reduction by the test chemical, but also for colour interference arising from the retention of the test chemical by killed tissues. This could lead to a double correction for colour interference since the NSC_{living} control already corrects for colour interference arising from the retention of the test chemical by living tissues. To avoid a possible double correction for colour interference, a third control for non-specific colour in killed tissues (NSC_{killed}) needs to be performed. In this additional control, the test chemical is applied on at least two killed tissue replicates per exposure time, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. A single NSC_{killed} control is sufficient per test chemical regardless of the number of independent tests/runs performed, but should be performed concurrently to the NSMTT control and, where possible, with the same tissue batch. The true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT minus $\%NSC_{living}$ plus the percent non-specific colour obtained with killed tissues exposed to the interfering test chemical and incubated with medium without MTT, calculated relative to the negative control run concurrently to the test being corrected ($\%NSC_{killed}$).

21. NSC_{living} or NSC_{killed} controls are never required when using HPLC/UPLC-spectrophotometry, independently of the chemical being tested. NSMTT controls should nevertheless be used if the test chemical is suspected to directly reduce MTT or has a colour (intrinsic or when mixed with water) that impedes the assessment of the capacity to directly reduce MTT as described in paragraph 16. When using HPLC/UPLC-spectrophotometry to measure MTT formazan, the percent tissue viability is calculated as

percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. For test chemicals able to directly reduce MTT, true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT. Finally, it should be noted that in very rare cases, direct MTT-reducers or MTT-reducers that are also colour interfering and are retained in the tissues after treatment may not be assessable by the VRMs if they lead to ODs (using standard OD measurement) or peak areas (using UPLC/HPLC-spectrophotometry) of the tested tissue extracts that fall outside of the linearity range of the spectrophotometer.

Acceptability Criteria

22. For each run, tissues treated with the negative control should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes and should not be outside of the historically established boundaries (see paragraph 7 and table 1). Similarly, tissues treated with the positive control, should show a mean tissue viability (relative to the negative control) within an historically established range, thus reflecting the ability of the tissues to respond to a corrosive chemical under the conditions of the test method. The variability between tissue replicates of test chemicals and/or control substances should fall within the accepted limits also established from historical values (*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 outside of the accepted ranges, the run is considered non-qualified and should be repeated. If the variability between tissue replicates of test chemicals falls outside of the accepted range, the test chemical should be re-tested. Paragraph 33 provides more details on re-testing in case of non-qualified runs during validation studies. Importantly, an increased frequency of non-qualified runs may indicate problems with either the test system (e.g. the intrinsic RhE tissue quality) or with the handling (e.g. shipment, SOP execution). Therefore, occurrence of non-qualified runs in validation studies should be carefully monitored and all non-qualified runs need to be reported.

Interpretation of Results and Prediction Model

23. 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%. In case HPLC/UPLC-spectrophotometry is used, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. The cut-off value of percentage cell viability distinguishing corrosive from non-corrosive test chemical (and/or discriminating between different corrosive sub-categories), and the statistical procedure(s) used to evaluate the results should be clearly defined, documented, and proven to be appropriate. The cut-offs defined for the VRMs are defined below in paragraphs 23 and 24.

24. 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.

25. The prediction model for the VRM EpiSkin™ skin corrosion test method (6) (8) (11), associated with the UN GHS (4) classification system, is shown in Table 3:

Table 3: Prediction model of the VRM EpiSkin™

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

26. The prediction models for the VRM EpiDerm™ SCT (7) (12) (13) test method associated with the UN GHS (4) classification system, are shown in Table 4:

Table 4: Prediction model of the VRM EpiDerm™ SCT

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

MINIMUM LIST OF REFERENCE CHEMICALS

27. Reference Chemicals are used to determine whether the reliability and predictive capacity of a proposed similar or modified test method, proven to be structurally and functionally sufficiently similar to the VRM, or representing a minor modification of the VRM, are equal or better than those derived from the VRMs (6) (7) (8). The 30 recommended Reference Chemicals listed in Table 5 include chemicals representing different chemical classes (*i.e.* chemical categories based on functional groups), and are representative of the full range of TG 404 *in vivo* skin corrosion scores. The chemicals included in this list comprise representatives of the following UN GHS (Sub-)categories: 10 Sub-category 1A chemicals, 10 chemicals of sub-categories 1B and 1C (the *in vivo* data do not permit distinction between the two categories) as well as 10 non-corrosive chemicals. The Reference Chemicals were selected from the test chemicals used in the validation studies of the VRMs (6) (7) (8) (14) using the selection criteria as described in Table 5 (foot-note 1), with due regard to e.g., chemical functionality and physical state.

28. The 30 Reference Chemicals listed in Table 5 represent the minimum number of chemicals that should be used to evaluate the reliability and predictive capacity of a proposed similar or modified test method able to discriminate between Subcategory 1A, a combination of Sub-categories 1B and 1C as well as non-corrosive substances and mixtures in accordance with the UN GHS (4) (1A vs. 1B-and-1C vs. NC). For similar or modified test methods able to discriminate corrosive from non-corrosive substances and mixtures but not able to support sub-categorisation of corrosive chemicals (C vs. NC), only 20 of the 30 chemicals listed in Table 5 (the ones not in *italics*) need to be evaluated: 5 Sub-category 1A chemicals, 5 chemicals of the combined Sub-categories 1B and 1C as well as 10 non-corrosive chemicals. The exclusive 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 Reference Chemical is unavailable, or cannot be used for other justified reasons, another chemical could be used provided it fulfils the selection criteria as described in Table 5 (foot-note 1) and adequate *in vivo* reference data are available, e.g. preferentially from the test chemicals used during the validation studies of the VRMs (6) (7) (8) (14). To gain further information on the predictive capacity of the proposed test method, additional chemicals representing other chemical classes and for which adequate *in vivo* reference data are available may be tested in addition to the minimum list of Reference Chemicals.

Table 5: Minimum list of Reference Chemicals for determination of Reproducibility and Predictive Capacity of 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 chemicals should be tested with similar or modified test methods proposed to identify Sub-category 1A, a combination of Category 1B and 1C (referred to as 1B/1C below) and non-corrosive test chemicals. These additional reference chemicals 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
Combination of UN GHS Sub-categories 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*

Chemical[1]	**CASRN**	**Chemical Class**[2]	**Physical State**	**EpiSkin™** [4]	**EpiDerm™** [4]	**SkinEthic™** [4]	**epiCS®** [4]
*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 Sub-category 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 (4); NC = Not Corrosive

[1]The reference chemicals, sorted first by corrosives versus non-corrosives, then by corrosive sub-category, were selected from the test chemicals used in the ECVAM validation studies of EpiSkin™ and EpiDerm™ SCT (6) (7) (14) and from post-validation studies based on data generated by EpiSkin™ (8), EpiDerm™, SkinEthic™ and epiCS® developers. Unless otherwise indicated, these chemicals were tested at the purity level obtained when purchased from a commercial source (6) (7). The selection includes, to the extent possible, chemicals 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.* (14).

[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. The numbers in brackets indicate, for each chemical, the number of the corresponding type of *in vitro* predictions for the test method considered. These predictions were corrected for direct MTT reduction using killed control tissues.

DEFINED RELIABILITY AND PREDICTIVE CAPACITY VALUES

29. For purposes of establishing the reliability (i.e., within- and between laboratory reproducibility) and predictive capacity (i.e., sensitivity, specificity and accuracy) of proposed similar or modified RhE test methods to be used by several independent laboratories, all 30 (or 24 for methods not able to sub-categorize corrosive chemicals) Reference Chemicals listed in Table 5 should be tested in at least three laboratories. In each laboratory, all relevant Reference Chemicals 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, positive control and adapted controls for direct MTT reduction and/or colour interference.

30. 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:

1. Within-laboratory reproducibility (WLR) should be calculated based on concordance of classifications using only qualified tests obtained with Reference Chemicals for which at least two qualified tests are available. In addition, it should be reported the number and identity of the Reference Chemicals which per laboratory have none or only one qualified test (not considered for WLR calculations), as well as how many and which Reference Chemicals 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 Chemicals for which at least one qualified test per laboratory is available. It should be reported how many and which Reference Chemicals do not have at least one qualified test per laboratory (not considered for BLR calculations), as well as how many and which Reference Chemicals 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 corrosive vs. non-corrosive) as well as, in case of subcategorisation, over- and under-prediction rates, 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 (15).

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 negative control and positive control, as defined in the corresponding SOP. Otherwise, the run is considered as non-qualified.

Within-laboratory reproducibility

31. 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 in every laboratory, a concordance of predictions (corrosive or non-corrosive) obtained in different, independent tests of the 24 relevant Reference Chemicals equal or higher (≥) than 90% (actual for EpiSkin^{TM}: 100%, 100% and 96% in each laboratory, respectively).

32. An assessment of within-laboratory reproducibility for similar or modified test method proposed to discriminate between Sub-category 1A, a combination of Sub-categories 1B and 1C as well as non-corrosive chemicals should show in every laboratory, a concordance of predictions obtained in different, independent tests of the 30 Reference Chemicals equal or higher (≥) than 80% (actual for EpiSkin™: 96%, 96% and 88% in each laboratory, respectively).

Between-laboratory reproducibility

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

Predictive capacity

34. The predictive capacity of the proposed similar or modified RhE test method should be equal or better than the target values derived from the VRMs. For similar or modified test methods proposed to discriminate corrosive from non-corrosive chemicals but unable to support sub-categorisation of corrosive chemicals, the sensitivity, specificity and accuracy obtained with the 20 relevant Reference Chemicals (Table 5) should be equal or higher (≥) than 95%, 70% and 82.5% respectively (Table 6). For similar or modified test methods proposed to discriminate between Sub-category 1A, a combination of Sub-categories 1B and 1C as well as non-corrosive chemicals, the minimum predictive capacity values that should be obtained with the 30 Reference Chemicals (Table 5) are indicated in Table 7. A distinction is made between RhE-based test methods similar to EpiSkin™ on the one hand and similar to EpiDerm™ on the other hand due to their differences in Sub-categorization predictive capacities.

Table 6: 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 support sub-categorisation of corrosive chemicals.

Sensitivity	Specificity	Accuracy
≥ 95% (actual for EpiSkin™: 100%; actual for EpiDerm™: 100%)[1]	≥ 70% (actual for EpiSkin™: 76.7%; actual for EpiDerm™: 73.3%)[1]	≥ 82.5% (actual for EpiSkin™: 88.3%; actual for EpiDerm™: 86.7%)[1]

[1] **Values are based on the results of the two VRMs (EpiSkin™ and EpiDerm™) for the 20 Reference Chemicals not in italics from Table 5.**

Table 7: Required predictive capacity for similar or modified RhE test method to be considered valid to discriminate between Sub-category 1A, a combination of Sub-categories 1B and 1C (referred to as 1B-and-1C below) and non-corrosive chemicals *.

VRM	EpiSkin™ [1]	EpiDerm™ [1]
Sensitivity (for predictions 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 (i.e., correct NC predictions)	≥ 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%)

[1] Actual values are based on the results of the two VRMs (EpiSkin™ and EpiDerm™) for the 30 Reference Chemicals (see table 5).

* Depending on the results obtained with a similar or modified RhE test method for the 30 Reference Chemicals, 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 Sub-category 1B-and-1C can be considered as Sub-category 1B-and-1C, whereas chemicals for which cell viability at 3 minutes is below 50% should be considered as Category 1, since the Sub-category 1A predictions of these three test methods contain a high rate of over-predictions of chemicals of Sub-categories 1B-and-1C (see also paragraph 7 of the Test Guideline 431 (2)). The regulatory framework in member countries will decide how this Test Guideline will be used, e.g. acknowledging the significant probability of overclassification, a Sub-category 1A classification may still be accepted or further testing may be conducted to confirm the result.

Study Acceptance Criteria

35. It is possible that one or several tests pertaining to one or more Reference Chemical does/do not

meet the test acceptance criteria (non-qualified tests) or is/are not acceptable for other reasons such as technical reasons or because they were obtained in a non-qualified run due to failure of the concurrent positive and/or negative control. To complement missing data, a maximum of two additional tests for each Reference Chemical is admissible per laboratory ("re-testing"). More precisely, since in case of re-testing also the positive and negative control substances have to be concurrently tested, a maximum number of two additional runs may be conducted for each Reference Chemical in each laboratory. Non-qualified tests should be documented and reported. Importantly, each laboratory should not produce more than three qualified tests per Reference Chemical. Excess production of data and subsequent data selection are regarded as inappropriate. All tested tissues should be reported. The extent of unacceptable tests/runs should be documented and the basis for the likely cause of each should be provided.

36. It is conceivable that even after re-testing, 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 Chemicals (24 for Category 1 vs. Non Corrosive; 30 for Sub-cat. 1A vs. Sub-cat. 1B-and-1C vs. Non Corrosive) should have at least one complete test sequence in one laboratory.
2. Each of at least three participating laboratories should have a minimum of 85% complete test sequences (for 24 Reference Chemicals: 3 incomplete test sequences are allowed per laboratory; for 30 Reference Chemicals: 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 Chemicals tested in 3 laboratories: a total of 7 incomplete test sequences are allowed; for 30 Reference Chemicals 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 test sequence may include both qualified and non-qualified 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.

LITERATURE

(1) 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].

(2) OECD (2015), OECD Guideline for the Testing of Chemicals No. 431. Reconstructed human epidermis (RhE) method for skin corrosion testing. OECD, Paris.

(3) 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. Available at: [http://iccvam.niehs.nih.gov/docs/dermal_docs/ps/ps044510.pdf]

(4) 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]

(5) OECD (2015), OECD Guideline for Testing of Chemicals. No. 404: Acute Dermal Irritation, Corrosion. OECD, Paris.

(6) Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhutter, H.-G., and Liebsch, M. (1998), The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the Management Team. *Toxicol.in Vitro* 12, 483-524.

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

(8) Alépée, N., Grandidier, M.H., and Cotovio, J. (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. *Toxicol.In Vitro* 28, 131-145.

(9) Mosmann, T. (1983), Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* 65, 55-63.

(10) Alépée, N., Barroso, J., De Smedt, A., De Wever, B., Hibatallah, J., Klaric, M., Mewes, K.R., Millet, M., Pfannenbecker, U., Tailhardat, M., Templier, M., and McNamee, P. Use of HPLC/UPLC-spectrophotometry for detection of MTT formazan in in vitro Reconstructed human Tissue (RhT)-based test methods employing the MTT assay to expand their applicability to strongly coloured test chemicals. Manuscript in preparation.

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

(12) 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].

(13) EC-ECVAM (2000), Statement on the application of the EpiDerm™ human skin model for skin corrosivity testing, issued by the ECVAM Scientific Advisory Committee (ESAC 14), 21 March 2000. Available at: [http://ecvam.jrc.ec.europa.eu].

(14) Barratt, M.D., Brantom, P.G., Fentem, J.H., Gerner, I., Walker, A.P., and Worth, A.P. (1998), The ECVAM international validation study on *in vitro* tests for skin corrosivity. 1. Selection and distribution of the test chemicals. *Toxicol.in Vitro* 12, 471-482.

(15) OECD (2013), OECD Serires on Testing and Assessment No. 190. Summary Document on the Statistical Performance of Methods in OECD Test Guideline 431 for Sub-Categorisation

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 (1).

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 (1).

C: Corrosive.

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 (1).

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

GHS (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 (4).

HPLC: High Performance Liquid Chromatography.

IC_{50}: Can be estimated by determination of the concentration at which a benchmark 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 (1). The term is interchangeably used with similar test method.

Mixture: means a mixture or solution composed of two or more substances in which they do not react (4).

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide.

NC: Non corrosive.

NSC_{killed}: Non-Specific Colour in killed tissues.

NSC: Non-Specific Colour in living tissues.

NSMTT: Non-Specific MTT reduction.
OD: Optical Density

PC: Positive Control, a replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

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 (1).

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 (1).

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.

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 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 (1).

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 (1).

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 (1).

Reproducibility: The agreement among results obtained from testing the same substance using the same test protocol (1).

Run: A run consists of one or more test chemicals tested concurrently, by one laboratory, with a negative control and with a positive control.

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 (1).

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 (5).

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 (1).

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).

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.

Validated Reference Method(s) (VRM(s)): one (or more) test method(s) officially endorsed as scientific valid that was(were) used to develop the related official Test Guidelines and Performance Standards (PS). The VRM is considered the reference test method 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 (1).