Organisation for Economic Co-operation and Development

ENV/CBC/WRPR(2023)50

Unclassified

English - Or. English

12 May 2023

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Performance Standards for the Assessment of Proposed Similar or Modified in Vitro Phototoxicity: Reconstructed Human Epidermis (RhE) Test Methods for Testing of Topically Applied Substances, as described in Test Guideline 498

Series on Testing and Assessment
No. 356

This corrected version of the Performance Standards was approved on 28 April 2023 by the Working Party of the National Coordinators of the Test Guidelines. The Chemicals and Biotechnology Committee is invited to endorse the corrected document by 20 June 2023.

JT03518834

Foreword

This document describes the Performance Standards (PS) for the assessment of proposed similar or modified methods to the Reconstructed Human Epidermis (RHE) Phototoxicity test method in TG 498. The PS are intended for the developers of new or modified similar test methods. TG 498 was adopted in 2021, on the basis of a project led by the U.S and Slovakia, who also developed the present Performance Standards, with the collaboration of the OECD expert group on Skin and Eye Irritation, and Phototoxicity. The PS were circulated to the Working Group of the National Coordinators of the Test Guidelines Programme (WNT) for two public commenting rounds, and they were revised accordingly for their submission to WNT, who approved the Performance Standards at its 34th meeting in April 2022. This document is published under the responsibility of the Chemicals and Biotechnology Committee.

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 irritation test methods that are structurally and mechanistically similar to the RhE test method in OECD Test Guideline (TG) 498 [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 12 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 demonstrated 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 [3-5] and already elsewhere accepted test method [6], 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 498. 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 498 before being used for regulatory purposes.

4. These PS have been defined based on the ECVAM sponsored prevalidation study [4, 5, 7] and follow-up ECVAM sponsored feasibility study [8] and related publications using the method [9-15]. 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 VRM used to develop the present PS is the EpiDerm™ test method as described in TG 498 (2). The EpiDerm™ Phototoxicity test method was used as a VRM to define the Essential Test Method Components. Definitions are provided in Annex 1.

5. Similar (me-too) or modified test methods proposed for use under TG 498 [2] should be evaluated to determine their reliability and predictive capacity using Reference Chemicals. Reference Chemicals represent a broad range of the phototoxicity responses *in vivo* (Table 3). Reference chemicals should be tested by the laboratories prior to use for testing other chemicals, in order to ensure that these methods are able to correctly discriminate between substances that are photo-irritating and non-photo-irritating to the skin. The proposed similar or modified test methods should have reliability and predictive capacity (i.e., sensitivity, specificity and accuracy) values which are equal to or better than those derived from the VRM and as described in paragraphs 27 to 31 of these PS (Table 4).

6. The WNT approved the PS at its 34th meeting in April 2022, and corrected the document in 2023. This corrected the Performance Standards document is published under the responsibility of the Chemicals and Biotechnology Committee.

ESSENTIAL TEST METHOD COMPONENTS

7. The Essential Test Method Components consist of essential structural, functional, and procedural elements of the scientifically valid test method (the VRM) that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, quality control measures, and acceptance criteria. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding VRM [1]. The essential test method components to be considered for similar or modified test methods related to TG 498 are described in detail in the following paragraphs.

8. For specific parameters (e.g., Table 1 and Table 2), or modified procedures, adequate values or procedures should be provided for the proposed similar or modified test method. These values or procedures may vary depending on the specific test method and/or its modification.

1. GENERAL CONDITIONS

9. 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 14). 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.

2. FUNCTIONAL CONDITIONS

2.1. Viability

10. The assay used for quantifying tissue viability is the MTT-assay [16]. 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 [17]. 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 of the RhE VRM 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 the negative control are stable in culture (provide similar viability measurements) for the duration of the test exposure period.

Table 1. Acceptability ranges for negative control OD values of the VRM

	Lower acceptance limit	Upper acceptance limit
EpiDerm™ (EPI-200)	≥0.8	≤ 2.8

2.2. Barrier function

11. 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).

2.3. Morphology

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

2.4. Reproducibility

13. The results of the positive and negative controls of the test method should demonstrate reproducibility of the test method over time.

2.5. Quality control (QC)

14. 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, amongst which those for *viability* (paragraph 10), *barrier function* (paragraph 11) and *morphology* (paragraph 12) are the most relevant. An acceptability range (upper and lower limit) for the IC_{50} or the ET_{50} (see paragraph 11) should be established by the RhE model developer/supplier. The acceptability range of the VRM is given in Table 1. 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 criteria can be accepted for reliable prediction of irritation classification.

Table 2. QC batch release criteria of the VRM

	Lower acceptance limit	Upper acceptance limit
EpiDerm™ (EPI-200) (1% Triton X-100) (14)	ET_{50} = 4.0 hr	ET_{50} = 8.7 hr

2.6. Radiation sensitivity of RhE tissues:

15. An irradiance dose of approximately 6 J/cm^2 (as measured in the UVA range) was determined to be non-cytotoxic in the VRM and sufficiently potent to excite chemicals to elicit phototoxic reactions. The viability of the irradiated tissues should be ≥ 80% relative to the tissues that were not irradiated [3-5, 7, 12, 13]. As an example, to achieve 6 J/cm^2 within a time period of 60 minutes, irradiance was adjusted to 1.7 mW/cm^2 of UVA/visible light (see Annex 2 of the OECD TG 498, Figure 2). Alternate exposure times and/or irradiance values may be used to achieve 6 J/cm^2 using the formula:

$$\text{t (min)} = \frac{\text{irradiation dose (J/cm}^2\text{)} \times 1000}{\text{irradiance (mW/cm}^2\text{)} \times 60} \qquad (1\,\text{J} = 1\,\text{Wsec})$$

The RhE tissue model is tolerant to UVB irradiation [14] and inclusion of UVB irradiation may be appropriate in some cases (e.g., when absorption for the test chemical of interest is exclusively in the UVB wavelength region). The presence of the UVB portion of the spectra should be monitored and reported.

Similarly, if a different dose and/or a different light source is used, the irradiation should be calibrated so that a dose regimen can be selected that is not deleterious to the cells but sufficient to excite standard phototoxins. This might be necessary in case of models with more complex structure (e.g., including melanocytes or other cells and layers) or with thicker *stratum corneum*.

3. PROCEDURAL CONDITION

3.1. Application of the Test Chemical and Control Substances

16. At least two tissue replicates should be used for each test chemical concentration and each control substance 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 (e.g., a minimum of 25 µL/0.6 cm^2 for the VRM). The chemical exposure, irradiation period and temperature need to be optimized for each RhE model and are related to the different intrinsic properties of the RhE model (e.g., for VRM 21 ± 3 hour exposure to chemicals and 60 min ± 5 min irradiation with a total of 6 J/cm^2 UVA). Furthermore, the viability measurement should not be performed immediately after exposure to the test chemical, but after a sufficiently long post-treatment incubation period. This period allows both for recovery from weak cytotoxic and phototoxic effects and for appearance of clear cytotoxic effects. A 21± 3 hours post-treatment incubation period was found optimal for the VRM [3, 5].

17. Concurrent negative control (NC) and positive control (PC) should be used in each run to demonstrate that viability (using the NC), and sensitivity (using 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 in the presence and absence of UVA/visible light irradiation. The PC suggested for the VRM is 0.02 % aqueous solution of Chlorpromazine hydrochloride. The suggested VRM NCs are phosphate buffered saline (DPBS) or Hanks' Balanced Salt Solution (HBSS) without phenol red.

3.2. Cell Viability Measurements

18. The MTT assay, which is a quantitative assay, should be used to measure tissue viability. It is compatible with use in a three-dimensional tissue construct. The tissue sample is placed into wells containing MTT solution of an appropriate concentration (1 mg/mL in the VRM) 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 using an HPLC/UPLC-spectrophotometry procedure [17]. The same procedure should be employed for the concurrently tested negative and positive controls.

19. 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 20 to 23). 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. 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 VRM is summarised in paragraphs 19 to 22 below.

20. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT solution, for example by adding the maximum volume and concentration of test chemical into the volume of MTT solution used in the test method. If the MTT mixture containing the test chemical turns blue/purple, the test chemical is presumed to directly reduce MTT and a further functional check on non-viable RhE tissues 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 test chemical is applied on at least two killed tissue replicates 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).

21. 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. When performing the standard absorbance (OD) measurement, each interfering coloured test chemical should be applied on at least two viable tissue replicates, 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 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 ($\%NSC_{living}$).

22. Test chemicals that are identified as producing both direct MTT reduction (see paragraph 20) *and* colour interference (see paragraph 21) will 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 19. 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, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. 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 ($\%NSC_{killed}$).

23. 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 19). 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 VRM 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.

3.3. Acceptability Criteria

24. 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 10 and Table 1). Similarly, tissues treated with

the PC should show mean tissue viability (relative to the negative control) within a historically established range, thus reflecting the ability of the tissues to respond to a phototoxic 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 in the relative viability values between the two replicate tissues treated with the vehicle (i.e., negative) or positive controls should not exceed 20 %). If either NC or PC included in a run falls 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. 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.

3.4. Interpretation of Results and Prediction Model

25. The OD values obtained with each test chemical should be used to calculate the percentage of viability relative to the negative control, which is set to 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 phototoxic from non- phototoxic test chemicals and the statistical procedure(s) used to evaluate the results and identify irritant chemicals should be clearly defined, documented, and proven to be appropriate (see SOPs of adopted test methods for information). The cut-off values of the VRM for the prediction of phototoxicity are given below [3, 5]:

a. A chemical is predicted to be **phototoxic** (or to have phototoxicity potential) if the relative viability values for one or more test concentrations treated in the presence of irradiation result in a decrease in viability exceeding 30% when compared to the relative viability values for the same concentrations treated in the absence of irradiation.

b. A chemical is predicted to be **non-phototoxic** (or to not have phototoxicity potential) if none of the relative viability values for the test concentrations treated in the presence of irradiation result in a decrease in viability exceeding 30% when compared to the relative viability values for the same concentrations treated in the absence of irradiation.

c. If none of the test concentrations result in a phototoxic prediction and at least one of the concentrations falls within 5% of the cutoff value, and/or non-concordant results from replicate tissues are obtained, a second run should be considered, as well as a third one in case of discordant predictions between the first two runs. In this case, it is recommended to consider a concentration range that is closer to the concentration in which the potentially phototoxic outcome was observed.

4. MINIMUM LIST OF REFERENCE CHEMICALS

26. 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 VRM [3, 5]. The 12 recommended Reference Chemicals listed in Table 3include chemicals representing different chemical classes (i.e., chemical categories based on functional groups), and are

representative of a broad range of photo-irritation potency (from non-phototoxic to strongly phototoxic materials). The Reference Chemicals were selected using the selection criteria as described in Table 3 on the basis of data from the VRM and relate to chemicals used for the pre-validation study [3, 5] and the follow-up studies [8-10, 12-14, 18-20]. Due regard has been given to chemical functionality and physical state when composing this list.

27. The 12 Reference Chemicals listed in Table 3 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. 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 3 and adequate *in vivo* reference data are available, e.g., preferentially from the test chemicals used during optimisation, from the pre-validation study or follow-up feasibility studies of the VRM [3-5] [8-10, 12-14, 18-20].

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 (preferably human)* reference data are available may be tested in addition to the minimum list of Reference Chemicals.

Table 3. Minimum List of 12 Reference Chemicals for Determination of Reproducibility and Predictive Capacity of similar or modified RhE for *in vitro* phototoxicity testing of topically applied substances.

	Substance[1]	CAS	In vivo	Vehicle[2]	VRM# Cat. based on *in vitro*	References
	Phototoxic Substances					
1	Chlorpromazine HCl	69-09-0	PT	Water	PT	[3-5, 12, 13, 15, 18, 20]
2	Anthracene	120-12-7	PT	EtOH or Acetone: Olive oil (4:1)	PT	[5, 12-15]
3	Bergamot oil (non-purified)[3]	8007-75-8	PT	Oil	PT	[3-5] [10, 12-15]
4	Acridine hydrochloride	17784-47-3	PT	Oil	PT	[3, 4] [13-15]
5	8-Methoxypsoralen	298-81-7	PT	Oil	PT	[3-5, 15]
6	Neutral red	553-24-2	PT	Water	PT	[3, 4, 13]
	Non-Phototoxic Substances					
7	Sodium dodecyl sulphate	151-21-3	NPT	Water	NPT up to highest conc. tested (1%)	[3, 4, 15]
8	Octyl salicylate	151-21-3	NPT	Oil	NPT up to highest conc. tested (10%)	[3, 4, 15]
9	4-Aminobenzoic acid (PABA)	150-13-0	NPT	Oil or EtOH	NPT up to highest conc. tested (10%)	[3-5, 12]
10	Penicillin G, Natrium salt	69-57-8	NPT	Water	NPT up to highest conc. tested (1%)	[3, 4, 15]
11	Octyl methoxycinammate	5466-77-3	NPT	Oil	NPT up to highest conc. tested (10%)	[3-5] [14, 15]
12	6-Methyl coumarin[4]	92-48-8	NPT	Oil	NPT up to highest conc. tested (0,1%)	[3, 4]

* VRM = validated reference method (EpiDerm, see paragraph 2 for explanations)
PT = Phototoxic; NPT = Non-Phototoxic; NPT/PA = Non-Phototoxic/Photoallergen
[1] The Reference Chemicals selection was based on the following criteria; (i) the chemicals are commercially available; (ii) they are representative of the broad range of phototoxicity responses (from non-phototoxic to strongly-phototoxic); (iii) they have a well-defined chemical structure; (iv) they are representative of the chemical functionality used in the validation process; (v) they are not associated with an extremely toxic profile (e.g., carcinogenic or toxic to the reproductive system); and (vi) they are not associated with prohibitive disposal costs.
[2] Vehicles are suggested, based upon the pre-validation and follow-up study references (EtOH - Ethanol/ Oil - Sesame seed oil).
[3] The phototoxic response of Bergamot oil is caused by impurities. Although non-purified Bergamot oil does not meet Reference Chemical selection criterion (iii) above, it is readily available from commercial chemical vendors. Non-purified Bergamot oils do have significant absorption in the UVA and UVB portions of the spectra (10) and result in phototoxic responses. It is recommended to assess the UVB/UVA spectra and/or conduct an HPLC analysis for impurities prior to use as a reference chemical.
[4] *6-Methylcoumarin* (6-MC), a fragrance material, has been reported to be photoallergenic both in man and in guinea pigs. 6-MC is however not classified as acutely phototoxic chemical. In the 3T3 NRU PT assay, 6-MC is usually classified as positive. In the 3D skin model test with exaggerated irradiation conditions or exposure (high concentrations or dose/area ratio) it may turn borderline or false positive.

5. DEFINED RELIABILITY AND PREDICTIVE CAPACITY VALUES

28. For purposes of establishing the reliability and predictive capacity (i.e., sensitivity, specificity and accuracy) of proposed similar or modified RhE test methods to be used by several laboratories, all 12 Reference Chemicals listed in Table 3 should be tested in at least three laboratories. In each laboratory, all 12 Reference Chemicals should be tested in three independent runs performed with different tissue batches . Each run should consist of at least two concurrently tested tissue replicates for each test chemical, negative control, positive control and adapted controls for direct MTT reduction and/or colour interference.

29. 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:

a. Only the data of runs from complete run sequences qualify for the calculation of the test method within-, and between-laboratory variability and predictive capacity (accuracy).

b. The final classification for each Reference Chemical in each participating laboratory should be obtained by using the mean value of viability over the different runs of a complete run sequence.

c. Only the data obtained for chemicals that have complete run sequences in all participating laboratories qualify for the calculation of the test method between-laboratory variability.

d. The predictive capacity should be determined using a weighted calculation in which 1) the final outcome of each individual qualified test obtained for each Reference Chemical (from all laboratories participating in the validation study) is captured as an independent prediction in the calculation and 2) correction factors are applied so that all Reference Chemicals have an equal weight in the calculations, even in cases where it was not possible to obtain the same number of qualified tests for all Reference Chemicals during the validation study. In summary, the prediction for each Reference Chemical obtained at each laboratory participating in the study should be divided by the total number of available predictions to determine the number of correct, over-, and under-predictions for each Reference Chemical

and these should be used to calculate sensitivity, specificity, and accuracy in a manner that all chemicals exert an equal weight in the calculations.

In this context, a **run sequence** consists of three independent runs from one laboratory for one test chemical. A **complete run sequence** is a run sequence from one laboratory for one test chemical where a total of three qualified runs have been achieved. Once three qualified runs have been conducted, no further testing of the test chemical will be conducted. If for a test chemical any runs are not qualified, no more than five attempts will be allowed.

5.1. Within-laboratory reproducibility

30. An assessment of within-laboratory reproducibility should show in one single laboratory, a concordance of predictions (Phototoxic and non-Phototoxic) obtained in different, independent test runs of the 12 Reference Chemicals equal or higher (≥) than 90%.

5.2. Between-laboratory reproducibility

31. An assessment of between-laboratory reproducibility is not essential if the proposed test method is to be used in a single laboratory only. For methods to be transferred between laboratories, the concordance of predictions obtained in different, independent test runs of the 12 Reference Chemicals between a minimum of three laboratories should be equal or higher (≥) than 80%.

5.3. Predictive capacity

32. The predictive capacity (sensitivity, specificity and accuracy) of the proposed similar or modified test method should be similar to or better than the target values derived from the VRM (Sensitivity 87%, Specificity 93%) [3, 4, 7]. Accordingly, the sensitivity and specificity with the 12 Reference materials should each be at least 80% (Table 4). There is no further restriction with regard to the specificity of the proposed *in vitro* test method, i.e., any participating laboratory may misclassify any *in vivo* non-Phototoxic chemical as long as the final specificity of the test method is within the acceptable range. Test method accuracy should be at least 80%.

Table 4. Required sensitivity, specificity and accuracy values for similar or modified RhE test method to be considered valid to discriminate phototoxic chemicals from non-phototoxic chemicals

Sensitivity	Specificity	Accuracy
≥ 80%	≥ 80%	≥ 80%

Note: each misclassified substance from the set of 12 contributes to the decrease of prediction by 17 %

5.4. Study Acceptance Criteria

33. It is possible that one or several tests pertaining to one or more Reference Chemicals 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 runs are admissible ("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.

ANNEX I

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).

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 run sequence: A run sequence containing three qualified tests. A run sequence containing less than 3 qualified tests is considered as incomplete.

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 (3).

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 combination of, or solution composed of two or more substances in which they do not react (3).

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

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

NSC_{living}: 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 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.

Phototoxicity *in vivo*: Phototoxicity (photoirritation) is defined as an acute toxic response elicited by topically or systemically administered photoreactive chemicals after the exposure of the body to environmental light. Within the context of skin exposures to phototoxic chemicals, phototoxic responses are elicited after the first acute exposure of skin to photoactive chemicals and subsequent exposure to light.

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 vehicle which may be separated without affecting the stability of the substance or changing its composition.

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

Test chemical: means what is being tested.

Validated Reference Method(s) (VRM(s)): one (or more) test method(s) officially endorsed as scientifically 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).

LITERATURE

1. OECD, 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. 2005 https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2005)14&doclanguage=en.
2. OECD, OECD Test Guideline for the Testing of Chemicals No.498. In vitro Phototoxicity test for topically applied substances: reconstructed human epidermis test method. . 2021: OECD, Paris.
3. Liebsch, M., et al., Prevalidation ofthe EpiDerm Phototoxicity test. 1999, CPL Press, Newbury, UK: In Alternatives to Animal Testing II: Proceedings of the second international scientific conference organised by the European Cosmetic Industry, Brussels, Belgium (ed. D. Clark, S. Lisansky & R. Macmillan). p. 160-165.
4. Liebsch, M., Prevalidation of the "EpiDerm™ Phototoxicity Test" FINAL REPORT (Phases I, II,III). 1999b: https://tsar.jrc.ec.europa.eu/.
5. Liebsch, M., et al., Development of a new in vitro test for dermal phototoxicity using a model of reconstituted human epidermis (EpiDerm™). ALTEX - Alternatives to animal experimentation, 1997. **14**(4).
6. ICH, ICH S10 Photosafety Evaluation of Pharmaceuticals. Guidance for Industry. . January 2015.: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/s10-photosafety-evaluation-pharmaceuticals.
7. Liebsch, M.T., D. Barrabas, C. Spielmann, H. Gerberick, GF. Cruse, L. Diembeck, W. and U.S. Pfannenbecker, J. Holzhütter, HG. Brantom, P. Aspin, P. Southee, J., Prevalidation of the EpiDerm phototoxicity test.,in Alternatives to animal testing II: proceedings of the second international scientific conference organised by the European Cosmetic Industry, D. Clark, S. Lisansky, and R. Macmillan, Editors. 1999, Newbury: CPL Press: Brussels, Belgium. p. 160–6.
8. Liebsch, M., Interim Report. A feasibility study on whether the prevalidated human 3-D epidermis model in vitro phototoxicity test (EpiDerm-PT), could successfully be used for toxic potency testing. Contract No. 19868-2002-09 F1ED ISP DE. 2006: https://tsar.jrc.ec.europa.eu/.
9. Jirová, D., et al., Phototoxicity of bituminous tars-correspondence between results of 3T3 NRU PT, 3D skin model and experimental human data. Toxicol In Vitro, 2005. **19**(7): p. 931-4.
10. Kejlová, K., et al., Phototoxicity of bergamot oil assessed by in vitro techniques in combination with human patch tests. Toxicol In Vitro, 2007. **21**(7): p. 1298-303.
11. Kejlová, K., et al., Toxicological testing of a photoactive phthalocyanine-based antimicrobial substance. Regul Toxicol Pharmacol, 2020. **115**: p. 104685.
12. Líšková, A., et al., Evaluation of phototoxic and cytotoxic potential of TiO_2 nanosheets in a 3D reconstructed human skin model. ALTEX, 2020;37(3):441-450.
13. Kandarova, H. and M. Liebsch, The EpiDerm™ Phototoxicity Test (EpiDerm™ H3D-PT). in Alternatives for Dermal Toxicity Testing. Editors: C. Eskes, E. van Vliet, and H. Maibach 2017. p. 483-503.
14. Jones, P., et al., Phototoxicity test ing using 3D Reconstructed Human skin models.in Alternatives to Animal Testing II: Proceedings of the second international scientific conference organised by the European Cosmetic Industry, Brussels, Belgium,Newbury, UK: CPL Press. 1999 p. 138-141.
15. Jones, P.A., et al., In Vitro Testing for Phototoxic Potential Using the EpiDerm 3-D reconstructed Human Skin Model. Toxicology Methods.2001. **11**: p. 1-19.
16. Mosmann, T., Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. J. Immunol. Methods 1983. **65**: p. 55- 63.
17. Alépée, N., et al., Use of HPLC/UPLC-spectrophotometry for detection of formazan in in vitro

Reconstructed human Tissue (RhT)-based test methods employing the MTT-reduction assay to expand their applicability to strongly coloured test chemicals. Toxicol In Vitro, 2015. **29**(4): p. 741-61.

18. Kandarova, H., Evaluation and Validation of Reconstructed Human Skin Models as Alternatives to Animal Tests in Regulatory Toxicology, in Department of Biology, Chemistry and Pharmacy. 2006, F reie Universität Berlin, Berlin, 2006, pp. 167.
19. Kejlová, K., et al., Phototoxicity of essential oils intended for cosmetic use. Toxicology in Vitro. 2010. **24**(8): p. 2084-2089.
20. Kim, S.Y., et al., Evaluation of phototoxicity of tattoo pigments using the 3 T3 neutral red uptake phototoxicity test and a 3D human reconstructed skin model. Toxicol In Vitro, 2020. **65**: p. 104813.