**EURL ECVAM**

**Test Submission Template (TST)**

**TABLE OF CONTENTS**

[**PURPOSE OF THE TST** 5](#_Toc509237922)

[**NOTE REGARDING TERMS** 6](#_Toc509237923)

[1. GENERAL INFORMATION 9](#_Toc509237924)

[**1.1** **TEST METHOD SUBMITTER** 9](#_Toc509237925)

[**1.2** **ADDITIONAL CONTACT PERSON (OPTIONAL)** 9](#_Toc509237926)

[**1.3** **NAME OF THE TEST METHOD** 10](#_Toc509237927)

[**1.4** **LIST OF ABBREVIATIONS** 10](#_Toc509237928)

[**1.5** **SUBMISSION OF CONFIDENTIAL INFORMATION** 10](#_Toc509237929)

[2. INFORMATION ON VALIDATION MODULES 11](#_Toc509237930)

[**2.1** **MODULE 1: TEST DEFINITION** 11](#_Toc509237931)

[**2.1.1** **Human health, environmental or other biological effects addressed by the test method** 11](#_Toc509237932)

[**2.1.2** **Scientific basis – biological and/or mechanistic relevance** 11](#_Toc509237933)

[**2.1.3** **Intended purpose of the test method** 11](#_Toc509237934)

[**2.1.4** **Improvement in relation to existing test method(s)** 12](#_Toc509237935)

[**2.1.5** **Contribution to the 3Rs** 12](#_Toc509237936)

[**2.1.6** **Protocol(s) of the test method** 12](#_Toc509237937)

[**2.1.7**  **Test items used for developing and optimising the test method (protocol and prediction model)** 14](#_Toc509237938)

[**2.1.8**  **Cost and time estimates per test item** 14](#_Toc509237939)

[**2.1.9**  **Occurrence of non-qualified tests** 14](#_Toc509237940)

[**2.1.10**  **Issues and drawbacks** 14](#_Toc509237941)

[**2.1.11** **Intellectual Property Rights** 14](#_Toc509237942)

[**2.1.12** **Essential information for a validation study based on performance standards (for similar or updated test methods)** 15](#_Toc509237943)

[**2.2**  **MODULE 2: WITHIN-LABORATORY REPRODUCIBILITY** 15](#_Toc509237944)

[**2.2.1** **Test items used for assessing within-laboratory reproducibility** 15](#_Toc509237945)

[**2.2.2** **Assessment of within-laboratory reproducibility** 15](#_Toc509237946)

[**2.2.3** **Identification and discussion of outlying values** 15](#_Toc509237947)

[**2.3** **MODULE 3: TRANSFERABILITY** 16](#_Toc509237948)

[**2.3.1** **Test items used for assessing the transferability** 16](#_Toc509237949)

[**2.3.2** **Training required for transferring the test method** 16](#_Toc509237950)

[**2.3.3** **Obstacles pertaining to transferability that are specific to the test method** 16](#_Toc509237951)

[**2.3.4** **Assessment of the transferability to other laboratories** 16](#_Toc509237952)

[**2.4**  **MODULE 4: BETWEEN-LABORATORY REPRODUCIBILITY** 17](#_Toc509237953)

[**2.4.1** **Test items used for assessing between-laboratory reproducibility** 17](#_Toc509237954)

[**2.4.2** **Assessment of between-laboratory reproducibility** 17](#_Toc509237955)

[**2.4.3** **Identification and discussion of outlying values** 17](#_Toc509237956)

[**2.5** **MODULE 5: PREDICTIVE CAPACITY** 18](#_Toc509237957)

[**2.5.1** **Test items used for assessing predictive capacity** 18](#_Toc509237958)

[**2.5.2** **Assessment of the predictive capacity of the test method** 18](#_Toc509237959)

[**2.5.3** **Identification and discussion of false predictions** 18](#_Toc509237960)

[**2.5.4** **Quality/variability/uncertainty of the reference data used to calculate predictive capacity** 18](#_Toc509237961)

[**2.6** **MODULE 6: APPLICABILITY DOMAIN** 19](#_Toc509237962)

[**2.6.1** **Limitations of the test method identified through testing** 19](#_Toc509237963)

[**2.6.2** **Technical limitations of the test method** 19](#_Toc509237964)

[**2.6.3** **Mechanistic limitations of the test method** 19](#_Toc509237965)

[**2.7** **MODULE 7: PERFORMANCE STANDARDS** 20](#_Toc509237966)

[**2.7.1** **Suggestions of essential test method components for the evaluation of similarity** 20](#_Toc509237967)

[**2.7.2** **Suggestions of possible reference test items** 20](#_Toc509237968)

[**2.7.3** **Suggestions of defined accuracy and reliability values** 20](#_Toc509237969)

[3. ADDITIONAL INFORMATION 21](#_Toc509237970)

[**3.1** **Study organisation and management (Project Plan)** 21](#_Toc509237971)

[**3.2** **History of test method development** 21](#_Toc509237972)

[**3.3** **Collection of existing (historical) information** 21](#_Toc509237973)

[**3.4** **Test items used in the submission** 21](#_Toc509237974)

[**3.5**  **Quality system(s) used in the generation of the data submitted in this TST** 21](#_Toc509237975)

[**3.6** **Study conclusions** 22](#_Toc509237976)

[**3.7** **Recommendations** 22](#_Toc509237977)

[**3.8** **Additional information** 22](#_Toc509237978)

[**3.9** **List of references** 22](#_Toc509237979)

[4. ATTACHMENTS 23](#_Toc509237980)

[5. ABBREVIATIONS 25](#_Toc509237981)

[6. REFERENCES 25](#_Toc509237982)

**PURPOSE OF THE TST**

This Test Submission Template (TST) is the EURL ECVAM standard reporting format for Step 2 of the EURL ECVAM process for submission of test methods for validation and/or peer review. The TST and all accompanying material (Annexes) will be used by EURL ECVAM to evaluate the proposed test method. The TST can be used to submit test methods for validation coordinated by EURL ECVAM or finalised external validation studies for evaluation of the test method's readiness to be peer reviewed by the EURL ECVAM Scientific Advisory Committee (ESAC). It should be noted though that all validation studies, either coordinated by EURL ECVAM or externally, should follow good validation practices which should be laid down in a well-defined project plan. A good project plan should include (but does not need to be limited to) the following points:

1) Study objective and goals.

2) Test method description.

3) Management structure (composition and role of the different actors involved in the study): Validation Management Group (VMG); study coordinator; study sponsor(s); chemicals selection group; entity/person responsible for chemicals acquisition, coding and distribution; participating laboratories (experienced and naïve); entity/person responsible for biostatistics; etc.

4) Chemicals selection criteria.

The use of the same chemicals in different phases of test method development and validation, as described below, should be avoided to the extent possible:

1. Chemicals used for developing and optimising the test method (protocol and prediction model) should not be selected for assessing its within-laboratory reproducibility, between-laboratory reproducibility and predictive capacity.
2. Different chemicals should be used for:

* training laboratories on the test method,
* assessing the transferability of the test method to other laboratories and,
* assessing the test method's within-laboratory reproducibility, between-laboratory reproducibility and predictive capacity.

5) Procedure for chemicals acquisition, coding, distribution, receipt and handling.

6) Identification of study directors, safety officers, quality assurance directors and experimental team in each participating laboratory and description of their roles and responsibilities.

7) Study design (including instructions on the number of replicates measurements per test, the number of tests per test item, the way in which unqualified data are dealt with, the number of re-testing allowed, etc.).

8) Data collection, handling and analysis.

9) Quality assurance/good laboratory practice in the participating laboratories and at the test system producers (quality assurance audits at the production sites may be considered if production is not done under GMP).

10) Timelines of the study.

The types of validation processes include (1) prospective validation, i.e. when some or all information necessary to assess the validity of a test method are not available, and therefore new experimental work is required, (2) retrospective validation, i.e. an assessment carried out based on existing (historical) information, (3) validation based on Performance Standards, i.e.

validation of new or updated test methods found sufficiently similar to previously validated and accepted reference test method(s).

As a general rule, the information submitted in the TST should be:

* as complete as possible, depending also on the type of validation sought
* as detailed as necessary, but
* as concise as possible.

Please note that the TST can be used for submission of single test methods as well as of Defined Approaches (i.e. testing strategies) composed of more than one information source

**NOTE REGARDING TERMS**

Below is a list of terms related to test method validation, as defined in OECD Guidance Document No 34 (OECD, 2005) or defined for the purpose of the TST (identified with an \*).

**\*Benchmark:** A test item that produces a midrange response in the test method, i.e. to assess variability of the test system over time. Please note that if positive controls elicit a too strong response, they cannot be used as benchmark.

**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.

**\*Biological relevance:** Relates to the extent to which the test methods models or reproduces the biological properties of target organ/system or species of interest (e.g. mechanism of action, cell types, cytoarchitecture).

**Endpoint:** The biological or chemical process, response, or effect, assessed by a test.

**False negative:** A substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

**False negative rate:** The proportion of all positive substances falsely identified by a test method as negative. It is one indicator of test method performance.

**False positive:** A substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

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

**\*Negative Control:** The vehicle used and/or a test item known not to elicit a positive response in the test method.

**\*Non-qualified test:** A test that does not meet the acceptance criteria defined in the protocol.

**Performance standards:** 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 (1) essential test method components; (2) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, 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.

**\*Positive Control:** A test item well known to elicit a positive response in the test method.

**\*Predictive capacity:** The ability of a test method to make relevant predictions on defined biological effects (e.g. human health effects).

**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: (1) a definition of the specific purpose(s) for which the test method is to be used; (2) specifications of all possible results that may be obtained, (3) an algorithm that converts each study result into a prediction of the (toxic) effect of interest, and (4) 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.

**\*Qualified test:** A test that meets the acceptance criteria defined in the protocol.

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

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

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

**\*Run:** A run consists of one or more test items tested concurrently, and if applicable, also with a positive and a negative control.

**Sensitivity:** The proportion of all positive/active substances that are correctly classified by the test. 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.

**Specificity:** The proportion of all negative/inactive substances that are correctly classified by the test. 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.

**\*Test:** The use of a test method for testing a single test item within a single experiment (can be composed of one single measurement or several measurements conducted in parallel, i.e. “replicates”).

**\*Test Item:** Any entity to be tested with the test method. These may be single substances, mixtures, biologicals, etc.

**Test method:** A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with “test” and “assay”.

**\*Test System:** A test method is usually composed of three elements: (i) test system, (ii) protocol, (iii) prediction model. The test system is the biological/chemical system that is exposed to the test items to obtain experimental data.

**Transferability:** The ability of a test procedure to be accurately and reliably performed in independent, competent laboratories.

**\*Variability:** Within- and between laboratory variability are also referred to as within- and between laboratory reproducibility. Although reproducibility is generally the preferred term, it refers to the same concept as variability, often used in the literature. The latin prefixed intra- and inter- are often replaced with the English translation within- and between.

**Within-laboratory reproducibility:** A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times.

**1. GENERAL INFORMATION**

**1.1 TEST METHOD SUBMITTER**

|  |  |
| --- | --- |
| Organisation/Company | Cosmetics Europe –The Personal Care Association AISBL |
| Department/Faculty/  Institute or Other | Cosmetics Europe Science & Research (S&R) |
| Address | Avenue Herrmann Debroux 40 |
| Postcode | 1160 |
| Town | Brussels |
| Country | Belgium |
| Responsible Contact: | |
| Title | Mr.  Ms.  Dr.  Prof. |
| Surname | Hoffmann |
| First name | Sebastian |
| Function | Consultant for Cosmetics Europe |
| Phone | +49 5251 8700566 |
| Fax | - |
| Email | sebastian.hoffmann@seh-cs.com |

**1.2 ADDITIONAL CONTACT PERSON (OPTIONAL)**

|  |  |
| --- | --- |
| *Title* | Mr.  Ms.  Dr.  Prof. |
| *Surname* | Pfuhler |
| *First name* | Stefan |
| *Function* | Research Fellow, Global Product Stewardship |
| *Phone* | +1 513 3197468 |
| *Fax* | - |
| *Email* | Pfuhler.s@pg.com |
| *If different from 1.1:* | |
| *Organisation/Company* | Procter & Gamble |
| *Department/Faculty/*  *Institute or Other* | Global Product Stewardship, Human Safety |
| *Address* | 8700 Mason-Montgomery Road |
| *Postcode* | OH 45040 |
| *Town* | Mason |
| *Country* | USA |

**1.3 NAME OF THE TEST METHOD**

*Please write the full test method name and any eventual acronym.*

3D reconstructed human skin micronucleus assay (RSMN)

**1.4 LIST OF ABBREVIATIONS**

*Please list the abbreviations used in the submission.*

BLR: between-laboratory reproducibility

BN: binucleated

BSA: bovine serum albumin

2,4-DAT: 2,4-diaminotoluene

CPPE: Cyclopenta[c,d]pyrene

DMBA: 7,12-dimethylbenz[a]thracene

HC: historical control

HET-MN: hen's egg test - micronucleus induction

IARC: International agency for research on cancer

KCl: Potassium chloride

KOW: octanol-water partition coefficient

MMC: mitomycin C

MN: micronucleus

MN-BN: number of MN in number of BN cells scored

MP: misleading positive

MoA: mode of action

MW: molecular weight

OECD: organisation for economic co-operation and development

PC: positive control

RHS: reconstructed human skin

RS Comet assay: 3D Reconstructed Skin Comet assay

RSMN: reconstituted skin micronucleus test

(R)VCC: (relative) viable cell count

SC: solvent control

SCCS: Scientific Committee on Consumer Safety

TG: test guideline

TN: true negative

TP: true positive

WLR: within-laboratory reproducibility

**1.5 SUBMISSION OF CONFIDENTIAL INFORMATION**

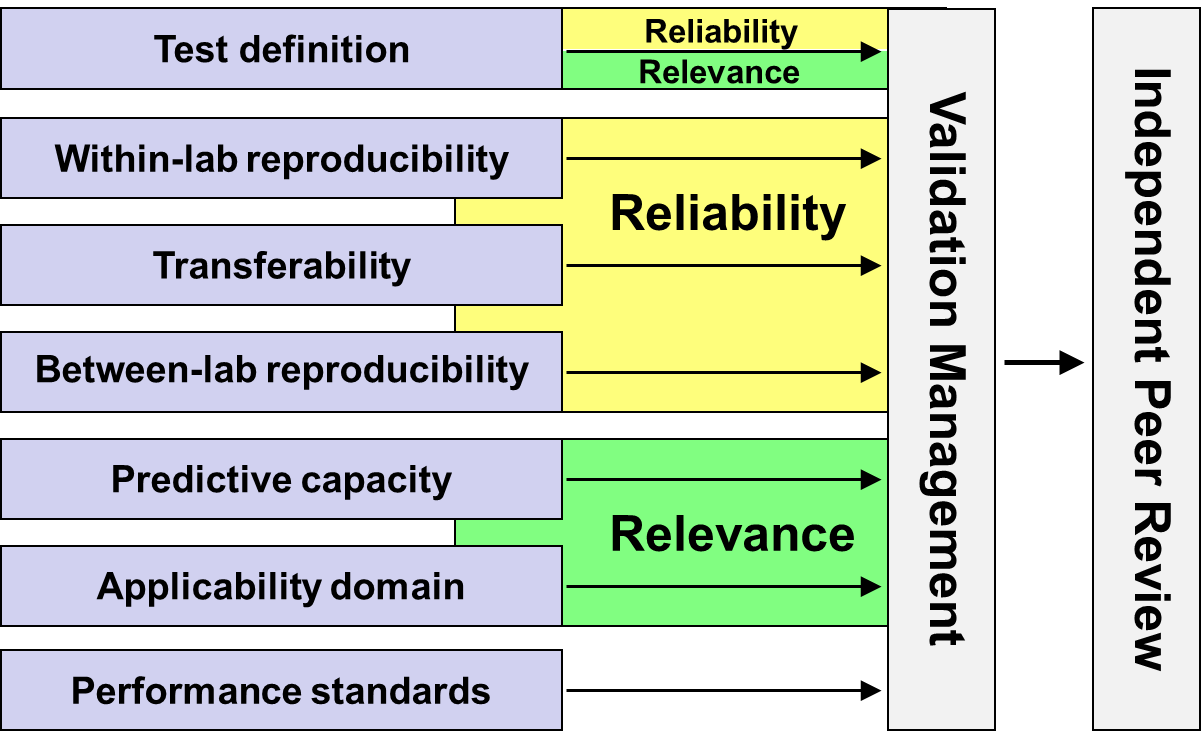
*It should be noted that during the evaluation phase of test method submissions, EURL ECVAM might share information contained in the TST with its Advisory Structure, e.g. the EURL ECVAM Scientific Advisory Committee (ESAC), and to the extent possible, with European regulatory authorities. Whoever gets access to the TST will be bound to treat all information submitted to EURL ECVAM in a confidential manner. Nevertheless, EURL ECVAM procedures involve that Background Review Documents on the evaluation of the test method and its outcome are compiled and made publicly available which might imply disclosing information/data submitted with the TST together with those generated during validation.*

*Therefore, if you consider some of the information provided in the TST as confidential (e.g. confidential business information), please clearly identify below those paragraphs where confidential information has been entered. Briefly describe (not more than 100 words per paragraph) why this information is considered confidential. Please specify if such confidential information may be made publicly available.*

This submission does not include any confidential information.

**2. INFORMATION ON VALIDATION MODULES**

*The EURL ECVAM modular approach to validation breaks down the information required for peer review of a test method and for establishing its scientific validity into seven independent modules (Hartung et al., 2004).*

**

*According to this modular approach, the information requirements can be fulfilled by using data obtained from a prospective study, by a retrospective evaluation of already existing data/information, or by a combination of both. The TST is structured according to these modules and the information necessary to complete each module is identified below.*

**2.1 MODULE 1: TEST DEFINITION**

*This module defines the scientific purpose of the test method and includes, among other, a description of its biological and/or mechanistic relevance in view of current scientific knowledge, its contribution to the 3Rs, intellectual property rights associated with the test method, and its protocol which should include specification of the test system, readout(s) and endpoint(s) measured, derivation and expression of results, acceptance criteria, interpretation of results via a prediction model and adequate controls.*

**2.1.1 Human health, environmental or other biological effects addressed by the test method**

*Please describe which human health, environmental or other biological effects the test method addresses/informs.*

The RSMN addresses the effect of toxicity to the genome, which is an integral part of genotoxicity and carcinogenicity hazard and risk assessment in regulations of industrial chemicals, cosmetic ingredients, biocides, pharmaceutical, plant protection products and medical devices around the globe. It transfers the approach of the well-established in vivo and in vitro micronucleus tests (see OECD TG 474 (OECD, 2016c) and 487 (OECD, 2016a)), which address the potential of a test item to cause genotoxicity in the form of chromosomal damage (clastogenicity and aneugenicity), to a three-dimensional human skin model.

The micronucleus assay is widely used for a variety of purposes (Sommer et al., 2020), including regulatory applications.

The RSMN uses three-dimensional reconstructed human skin (RHS) models to provide a relevant test method for test items that are primarily associated with dermal exposure.

**2.1.2 Scientific basis – biological and/or mechanistic relevance**

*Please describe the relationship between the test method and the effect of interest. This may include a) the mechanistic relevance (e.g. the mechanism of action) and/or b) the biological relevance (e.g. how well the test method models the target organ) of the test system/method.*

Mechanistic relevance

As any micronucleus assay, the RSMN can detect two types of genotoxic damage potentially induced by test items, i.e. clastogenicity (chromosomal breaks and translocations) and aneugenicity (abnormal number of chromosomes). As these types of genotoxicity are predicted with high sensitivity (Hayashi, 2016), the micronucleus test is generally considered of high mechanistic relevance and to have high predictivity as evidenced by decades of use of the in vivo micronucleus test, e.g. as OECD TG 474 (OECD, 2016c), and by the acceptance of the in vitro mammalian cell micronucleus test after formal retrospective validation, ultimately leading to OECD TG 487 (Corvi et al., 2008; OECD, 2016a).

The RSMN has been developed and optimised to identify genotoxicity hazard in the context of dermal exposure, especially in a tiered strategy to follow-up positive results obtained with the in vitro micronucleus or chromosomal aberration test. It is not intended to identify the genotoxicity mechanism. However, it can be adapted to distinguish clastogens and aneugens (Roy et al., 2016).

Biological relevance

The micronucleus assay can be applied to any cell type or tissue, provided that cells are dividing in the time frame of the assay. The RSMN uses 3D RHS models. Specifically, the validation of the RSMN assay was conducted with EpiDerm™ (MatTek, Ashland, USA). The EpiDerm™ is cultured from normal human epidermal keratinocytes derived from neo-natal foreskin tissue on specially prepared tissue culture inserts. As EpiDerm™ resembles the structure, morphology, and xenobiotic metabolism of the human epidermis (Hu et al., 2010; Götz et al., 2012; Hewitt et al., 2013; Yuki et al., 2013), it is of high biological relevance when assessing the genotoxic potential of test items via the dermal exposure route.

**2.1.3 Intended purpose of the test method**

*Please describe the intended purpose (i.e. practical use) of the test method (e.g. regulatory and/or non-regulatory applications) and whether it is intended to be used alone or in combination with other methods (e.g. within an Integrated Approach to Testing and Assessment (IATA) and/or a Defined Approach (DA)).*

The RSMN is intended to be used within regulatory genotoxicity hazard identification testing strategies to follow up positive results from the classical in vitro test battery, which is used as a first step. It is proposed to be used instead of in vivo genotoxicity test methods (e.g, OECD TG 474 (OECD, 2016c) and TG 475 (OECD, 2016b)) in the context of test items that are causing clastogenic/aneugenic effects in vitro and are primarily associated with dermal exposure, for which other exposure routes are negligible or can be addressed by other means. Together with the RS Comet assay, all genotoxicity endpoints that usually need to be addressed for regulatory purposes (gene mutation, clastogenicity and aneugenicity) can be covered.

**2.1.4 Improvement in relation to existing test method(s)**

*Please describe whether the test method represents an improvement compared to an existing method [e.g. better/more information (e.g. higher accuracy or addressing a mechanism of action), higher throughput, lower cost, etc.].*

In parallel to or in combination with the RS Comet assay, the RSMN offers for the first time an in vitro testing solution to follow up positive results from the standard in vitro genotoxicity testing battery. In addition, the RSMN potentially provides information of higher human relevance for test items with a dominantly dermal exposure as compared to dermal, or oral, in vivo genotoxicity test with rodents or nude mice which suffer from intrinsic and technical limitations. In such cases, uncertainty in hazard and risk assessment can be reduced, e.g. as the use of RHS can eliminate interspecies differences.

**2.1.5 Contribution to the 3Rs**

*Please describe the impact of the test method in terms of the 3Rs principle (Replacement, Reduction, Refinement of animal testing). In this context, please also indicate whether the test method implies the use of animal derived serum (e.g. foetal bovine serum) and/or of animal derived antibodies (e.g. monoclonal or polyclonal antibodies). Has the use of chemically defined media (i.e. serum-free alternatives) and/or non-animal derived recombinant antibodies (i.e. produced with phage display technology) or other new generation animal-free affinity reagents, such as aptamers, affimers, DARPINs, etc. been considered?*

Under regulations that allow animal testing, the RSMN can be used instead of in vivo follow-up testing (triggered by positive results in the standard in vitro genotoxicity test battery) for chemicals primarily associated with dermal exposure and for which other exposure routes are negligible or can be addressed by other means. The main motivation for developing and validating the RSMN, however, was to provide a viable follow-up solution of high human relevance under regulations that prohibit in vivo testing, such as the EU cosmetics regulation.

The validation of the RSMN was performed with the EpiDerm™ skin model, which, during the cultivation phase of the tissue, is cultured with a specifically defined medium that contains bovine serum albumin (BSA) as the only animal-derived component. BSA is essential for the cultivation of differentiated skin models to maintain their unique physiological and structural features. So far, no adequate solution to replace BSA without quality loss could be identified. To obtain a single-cell suspension the skin models are trypsinised. To inactivate the trypsinisation, small amounts of trypsin neutralizing solution (Dulbecco’s Modified Eagle’s Medium) that contains 10% foetal bovine serum is used. Note that due to the intrinsic metabolic capacity of the skin tissues (Aardema et al., 2013; Kidd et al., 2021) which reflects human skin metabolism (Hewitt et al, 2013), no rat liver S9 mix needs to be added to identify pro-mutagens as is the case with the assays of the classical in vitro test battery.

**2.1.6 Protocol(s) of the test method**

*Please append as Attachment 1 the protocol(s) as used to generate the data submitted in this TST, preferentially in the DB-ALM protocol*[[1]](#footnote-1) *format. You can also submit the protocol(s) in your own format, as long as all essential information identified in the DB-ALM template are included. If different versions of the protocol have been used (e.g. (i) changed after the training and/or transfer phases in a prospective study, or (ii) in case existing data, gathered from literature and generated with different protocols, are included in this submission), please append or summarise all protocol versions used and describe the main differences/changes between them. If existing data, gathered from literature and generated with different protocols, are included in this submission, please link each identified protocol to the appropriate numbered information/data set(s) listed in section 3.3 (Attachment 13) and please indicate and justify the preferred or recommended protocol(s) for future use.*

The protocol as used in the last stages of the validation (Phases 2c and 2d, see Figure 4 in Chapter 3.1) and as summarized by Pfuhler et al. (2021), is provided as Attachment 1.1. It addresses all essential information of the DB-ALM template. This protocol evolved from the original protocol (Dahl et al., 2011), which was used in Phase 1. It was iteratively improved based on observations made during the early stages of the validation study (Phases 1, 2a and 2b, see Figure 4 in Chapter 3.1). The following protocol amendments took place:

• After Phase 1:

- While some laboratories already measured viable cell count in Phase 1, it was formally   
 added to the protocol as a second cytotoxicity readout before the start of Phase 2.

- Precipitation of test item onto the tissue surface, which can interfere with scoring or  
 impact tissue health, was added to the list of mandatory observations and it was  
 specified that only the lowest precipitating concentration was to be included in the  
 statistical analysis.

• After Phase 2b:

As an interim analysis of data from Phase 2a indicated that the sensitivity of the RSMN was lower than expected and the data from Phase 2b confirmed the higher sensitivity of the 72h exposure, the 72-experiment was added to the protocol.

The protocol recommended for future use foresees to start with a 72h-experiment, instead of a 48h-experiment. This potentially reduces the workload per substance, while the predictive capacity is maintained.

*The protocol(s) should contain a list and description of the materials needed, a description of what is measured and how (specification of parameters/endpoints measured), as well as a (preliminary) prediction model if applicable. In addition, a brief description of the following key elements of the protocol(s) should be provided in the following paragraphs:*

***a) Brief description of the test system***

*Briefly specify the test system used (e.g. tissue model, specific cells grown to confluence, etc.).*

The RSMN uses three-dimensional RHS models. In particular, Mattek’s EpiDermTM EPI-200-MNA skin tissues were used for the validation. These tissues are cultured from normal human epidermal keratinocytes derived from neo-natal foreskin tissue on specially prepared tissue culture inserts. They have been shown to resemble the structure, morphology and xenobiotic metabolism of the human epidermis (Hu et al., 2010; Hewitt et al., 2013) and are therefore of high biological relevance when assessing the genotoxic potential of test items via the dermal exposure route.

***b) Quality criteria applied to the test system***

*Please specify the quality criteria applied to the test system, if any (e.g. appropriate stratification and barrier function of a reconstructed human epidermis in each lot/batch).*

The tissue producer Mattek applies several criteria to qualify tissues:

- all tissues are visually inspected for physical imperfections.

- the time of exposure required to reduce the tissue viability (ET-50) using the MTT viability assay is determined for each tissue lot (acceptance range: 4.77 – 8.72 hours)

- all media used throughout the production process is checked for sterility

- all cells are screened and are negative for HIV, hepatitis B and hepatitis C using PCR

Only tissues meeting all criteria are released (with a quality certificate).

In addition, the RSMN protocol specifies to inspect EpiDerm kits for the air inclusion between the agarose gel and the insert after arrival of the shipment at the respective laboratory.

Note that the same tissue qualification procedure applies as for the tissues used for skin corrosion and skin irritation test methods according to OECD TG 431 and 439.

***c) Positive control, negative control, benchmarks***

*Please indicate all concurrent controls used and specify if they are used as acceptance criteria for the run (a run consists of one or more test items tested concurrently with a positive and a negative control). Please include also the acceptance range for the control responses and, where available, any historical data used to establish the acceptance range.*

**Positive Control(s)**

*The positive control(s) should consist of a test item(s) well known to elicit a positive response in the test method.*

Mitomycin C (MMC), which also is recommended in the OECD TG 487 for this purpose, is used as a positive control (PC) at a concentration of 3 µg/ml. This concentration induces a clear but not excessive increase in BN cells with micronuclei and no or low cytotoxicity, as demonstrated by the historical control data presented in Roy et al. (2016). The same PC response was consistently observed in the validation study across all laboratories and experiments.

**Negative Control(s)**

*The negative control(s) can consist of treatment with the vehicle used and/or a test item known not to elicit a positive response in the test method.*

Both a negative control of untreated tissues and a solvent control (SC) are used if no sufficient historical data demonstrating that the solvent has no impact on the test system are available. If sufficient historical SC data demonstrating no impact are available, only a SC is needed. For most test substances acetone has been used successfully. An alternative approved solvent is 70% ethanol (in water). Initially, Roy et al. (2016) have demonstrated a stable negative response across several solvents. Pfuhler et al. (2021) published solvent control data for the four laboratories of the validation study, showing how the 95%-quantile developed over time.

**Benchmarks (if applicable)**

*Benchmarks consist of test item(s) that produce a midrange response in the test method.*

not applicable

***d) Readout(s) and endpoint(s)***

*Please specify the readout(s) used (e.g. optical density), how they are measured and how the raw data are processed and analysed to generate the endpoint results (e.g. cell viability, EC50).*

The RSMN measures three readouts related to two endpoints (chromosomal damage and two cytotoxicity readouts) after repeated topical exposure of the skin tissues with the test item for 48h or 72h, which are called here 48h-experiment and 72h-experiment.

To measure chromosomal damage, keratinocytes are isolated from the skin tissues and binuleated (BN) cells with micronuclei (MN) are scored (see Dahl et al., 2011), after acridine-orange staining using a fluorescence microscope.

The number of MN counted, and the number of BN cells scored (at least 500, but typically 1000 per tissue) are recorded per slide, i.e. tissue. Basically, and as in OECD TG 487 (OECD, 2016a) and 474 (OECD, 2016c), a statistically significant and/or biological relevant increase in the number of MN in number of BN cells scored (MN-BN) in comparison to the SC indicate genotoxicity.

The ratio of the number of cells with at least one MN\*100 divided by the total number of BN cells scored is the % MN per tissue, which is used for graphical representation.

Cytotoxicity was determined by measuring the viable cell count (VCC) and the frequency of binucleated (BN) cells. The VCC from each tissue is determined by trypan blue dye exclusion or by differential staining with an AO/DAPI solution using a NucleoCounter NC250 instrument (ChemoMetec A/S, Allerod, Denmark). The BN frequency is evaluated using fluorescence microscopy of the fixed cells after AO staining. The % relative viable cell count (RVCC) of a tissue is determined by dividing the viable cell count\*100 by the average viable cell count of the SC tissues.

The % binucleation of a tissue is the number of binucleated cells\*100 divided by the sum of mononucleated, binucleated and multinucleated cells. To obtain the % relative binucleation (of a tissue), the % binucleation\*100 is divided by the average % binucleation of the SC.

As in OECD TG 487, the threshold for excessive cytotoxicity is set at 60% (55 ± 5%) reduction compared to the SC in either cytotoxicity read-out. If cytotoxicity is observed with both measurements, the more sensitive parameter is used. Test item doses exceeding those cut-off values for toxicity are not assessed for DNA damage.

***e) Acceptance criteria applied to the results***

*Please specify the acceptance criteria for the experimental data.*

An experiment is valid if it includes a valid SC, a valid PC and at least three valid test item doses, each with at least two valid tissues.

• A tissue is considered valid, if

- at least 500 total cells were analysed for proliferation (1N, 2N, ≥3N); any tissue with   
 < 500 cells is considered 100% toxic and averaged into the dose group toxicity.

- the yield of viable cells was higher than 50000

- SC tissues showed a binucleation rate of at least 25%.

Non-valid tissues are eliminated.

• The SC is considered valid, if it has at least two valid tissues and the % MN is not excessively high, i.e. similar to the negative control or not exceeding the historical SC range. If not valid, the entire experiment is not valid.

• The PC is considered valid, if it has at least two valid tissues and induces a statistically significant increase in the % MN compared to the VC (one sided fisher’s exact with a significance level of 0.05). If not valid, the entire experiment is not valid.

• A test item dose group is considered valid, if the threshold set for cytotoxicity (i.e. a >60% reduction in either % BN or % VCC compared to the SC) is not exceeded. If excessive cytotoxicity is observed, the dose group is excluded.

In the case of precipitation of the test item on top of the tissues at the end of treatment, only the lowest precipitating concentration is considered valid, as test material on the tissue surface can interfere with scoring and impact tissue health.

An experiment with only two valid doses can be considered valid if

(1) the first two doses were positive and not excessively cytotoxic, i.e. below the cytotoxicity threshold of 60% were not exceeded, or if

(2) genotoxic effects were absent in all doses and the third dose showed excessive cytotoxicity.

***f)* *Prediction Model/Data Interpretation Procedure***

*If appropriate, please describe the (preliminary) procedure to translate the test method(s) results into predictions of human health, environmental, and/or other biological effects.*

Only data of valid experiments is submitted to the data interpretation procedure, which considers the results of statistical tests, the historical SC range, dose-response and biological relevance. Fisher’s Exact Test (one-sided) is used to determine the statistical significance (p < 0.05, without accounting for multiple comparison) of differences between the SC and each of the doses for the number of MN in number of BN cells scored (MN-BN). For equivocal results, the dose-response is assessed using the one-sided Cochran–Armitage trend test, with p < 0.05 indicating significance.

A test item is considered

• positive,

- if there were two or more concentrations that produced statistically significant  
 increases in the % MN-BN outside the upper 95% control limit of the historical SC  
 data, as established by the 95%-quantile (in a 48h or a 72h-experiment), or

- if only the highest concentration produced a statistically significant increase in % MN-  
 BN and it was reproducible in an independent experiment (in 48h or 72h-experiments),  
 or

- if a mid-concentration induced a significant increase greater than the upper limit of the  
 historical control (HC) range, the trend test was significant, and it was reproducible in   
 an independent experiment (in 48h or 72h-experiments).

• negative, if none of the above criteria were met in two independent experiments, preferably including at least one 72-h experiment (Note that testing according to the recommended future protocol will foresee to start with a 72h-experiment.)

• equivocal, otherwise

***g)* *Expertise required for performing the test method protocol***

*Please describe the level of expertise and demonstrated proficiency required by the study personnel for performing the test method protocol.*

The conduct of a RSMN requires standard technical laboratory skills as well as expertise in interpreting data and results of genotoxicity test methods. In particular, experience with 3D tissue handling and with defining the dose range for in vitro genotoxicity tests is important.

In addition, expertise in integrating statistical results with an assessment of the biological relevance of effects is required to come to an overall conclusion per test item. To gain confidence in proper conduct of the RSMN, it is recommended to perform a series of experiments, e.g. using several concentrations of MMC. Such experiments can also be used to build a historical data base (HC) for SC (and PC), which is required to determine the validity of individual experiments, and which inform the biological relevance of effects.

***h) Critical material/equipment of the test method protocol***

*Please identify critical material/equipment that must be used in the protocol or that, if replaced, must be checked for its equivalence (e.g. by testing a set of proficiency chemicals).*

Some reagents with specific properties (for example, Cytochalasin B & Mitomycin C) are essential components of the procedure. Preferred vendors/product numbers are identified in the protocol. If vendor need to be changed for any reason, it is recommended to ensure that equivalent results are obtained with materials from the old and the new vendor.

***i) Critical steps of the test method protocol***

*Please identify critical steps in the protocol(s) that may largely influence the results of a test, potentially impacting on the reproducibility and/or predictive capacity of the test method, if performed improperly.*

Several protocol steps are critical that are also flagged in the protocol:

• When exposing the skin tissues to a test item or control, it is important to homogeneously distribute the solution/solvent over the whole skin tissue surface.

• Tissues should be inspected for precipitation at the beginning and end of each treatment to ensure proper dose selection and tissue health.

• Potassium chloride (KCl) treatment duration should be strictly adhered to. Over or under swelling the cells will impede scoring.

• Proper resuspension of cell pellets during fixation steps is critical to prevent clumping of cells on microscope slides, which could impede scoring.

• It is recommended to perform a second fixation step to reduce the potential to form salt crystals on the slide

• Prior to evaluation of BN & MN, slides should be coded by personnel not involved in the analysis.

• Scoring of binucleated cells and micronuclei should follow the guidelines

by Dahl et al. (2011).

**2.1.7 Test items used for developing and optimising the test method (protocol and prediction model)**

*Please indicate the test items used to develop and optimise the test method. Please append the full list of these test items in the form of a table as Attachment 2 (in EXCEL format), including their names, CAS numbers, commercial source, purity, physical form and any other relevant physical/chemical properties (e.g. chemical classes/organic functional groups, MW, LogP), and acquired data. If some of these test items were also used for assessing the test method's within-laboratory reproducibility, between-laboratory reproducibility and/or predictive capacity (please see section "Purpose of the TST" above, in particular point 4 – chemicals selection criteria), please indicate it under section 3.4 (see also sections 2.2.1, 2.4.1 and 2.5.1).*

The development of the RSMN was initiated by Curren et al. (2006), who tested MMC and vinblastine sulfate. Mun et al. (2009) refined the RSMN and tested seven genotoxins and five non-genotoxins (see attachments 2 and 14). Hu et al. (2009) explored the within- and between-laboratory reproducibility of the RSMN testing MMC, vinblastine sulfate, methyl methanesulfonate, 4-nitrophenol, trichloroethylene, 2-ethyl-1,3-hexanediol, and 1,2-epoxydodecane. Subsequently the protocol was further standardised (Dahl et al., 2011).

The test items used for development and optimisation of the RSMN are identified, including structures, CAS and EC numbers, in Attachment 2, which also contains information on the commercial supplier as well as their purity, physical form, chemical class/organic functional groups, molecular weight, octanol-water partition coefficient, water solubility, vapour pressure as well as links to websites, from which the physical-chemical properties were retrieved.

**2.1.8 Cost and time estimates per test item**

*Please give an estimate of the testing cost per test item, considering that the laboratory is equipped with all necessary standard equipment and not considering labour cost. Please indicate as well an estimation of the time required to complete data acquisition for a run and specify how many chemicals can be included in a typical run.*

The cost for consumables required for one experiment, incl. the MatTek EPI-200-MNA and the medium, amount to approximately 3400€. Note that the number of experiments required for a test item ranges between one and three (double check requirements for 48hr vs. 72hr in final validation protocol).

Considering the arrival and cultivation of tissues as the start (day 0), four consecutive working days are needed to conduct an experiment. Note that day 0 is typically a Tuesday and tissue harvest and slide preparation is performed on Friday. For a 72h-experiment, the first treatment will occur on the day tissues arrive. Slides are allowed to dry over the weekend and are stained/scored the following week.

Evaluation of a test article typically includes a dose range finding test as well as one or two definitive studies. However, experienced laboratories can test two test items in in parallel (e.g. dose range finder for one & definitive study for the other).

Table 1: Schematic timeline to conduct one RSMN experiment

Day 0 Arrival of tissues; Cultivation of tissues; 1st treatment (for 72h-experiment)

Day 1 1st or 2nd treatment

Day 2 2nd or 3rd treatment

Day 3 Tissue harvest; Preparation of Slides

Day 4+ Slide Staining; Microscopic Evaluation & Data Analysis (approx. 1 week)

**2.1.9 Occurrence of non-qualified tests**

*On the basis of your experience/historical data please provide an estimate (e.g. in percentage) of the frequency in occurrence of non-qualified tests (i.e. tests which do not meet the acceptance criteria - see section 2.1.6 e).*

Of the total of 210 validation study experiments, 207, i.e. 98.6%, fulfilled all validity criteria. Three experiments, i.e. 1.4%, were considered invalid. These are highlighted in Attachment 9 (worksheet ‘Summary by experiment).

**2.1.10 Issues and drawbacks**

*If applicable, please specify any known issues and drawbacks such as:*

*■ The test method requires expensive equipment and/or is based on an expensive test system*

*■ Equipment / test system needed is no longer commercially available*

*■ Etc.*

not applicable

**2.1.11 Intellectual Property Rights**

*Please state whether any component of the test method (e.g. protocol, test system, equipment) is patented, copyright protected, trade-marked, registered or treated as confidential business information (CBI). Please specify who is holding the Intellectual Property Rights.*

EpidermTM is a registered trademark and commercially available for more than ten years. It is also used, for example, in skin corrosion and skin irritation test methods (OECD TG 431 and 439).

**2.1.12 Essential information for a validation study based on performance standards (for similar or updated test methods)**

*For similar or updated test methods validated according to existing performance standards (developed on the basis of Validated Reference Method(s)), please indicate the performance standards used and how the similar or updated test method complies with the essential test method components as specified in the performance standard (see module 7). Please append the performance standards used as Attachment 3.*

not applicable

**2.2 MODULE 2: WITHIN-LABORATORY REPRODUCIBILITY**

*This module addresses the reproducibility of results within a single laboratory over time, using a defined protocol and the same laboratory set-up.*

**2.2.1 Test items used for assessing within-laboratory reproducibility**

*Please append the full list of the test items used for assessing within-laboratory reproducibility in the form of a table as Attachment 4 (in EXCEL format), including their names, CAS numbers, commercial source, purity, physical form and any other relevant physical/chemical properties (e.g. chemical classes/organic functional groups, MW, LogP), and acquired data. Please describe the criteria applied for selecting these test items. Please specify the extent to which these test items represent the applicability domain of the test method and the range of observed effects (e.g. non-toxic to highly toxic effect). If some of these test items were also used for developing and optimising the test method, training laboratories or assessing the transferability of the test method to other laboratories, please indicate it under section 3.4 and justify here below their selection for assessing the test method's within laboratory reproducibility (see also sections 2.1.7 and 2.3.1).*

The test items informing the within-laboratory reproducibility (WLR) are identified in Attachment 4, which also contains information on the commercial supplier as well as their purity, physical form, chemical class/organic functional groups, molecular weight, octanol-water partition coefficient, water solubility and vapour pressure.

Structures, EC-numbers and links to websites from which the physical-chemical properties were retrieved are included in the first worksheet of Attachment 9.

No specific criteria to select test items specifically for WLR were applied. WLR was addressed by considering the test items, for which at least two valid 48h-experiments and/or at least two valid 72h-experiments were available. As this was the case for 38 of the total 43 test items, they cover almost the entire applicability domain of the RSMN as explored so far.

No limitations in testing these test items were observed. Evaporation of test items from the tissue potentially leading to false negative results has not been an issue in any of the validation phases.

**2.2.2 Assessment of within-laboratory reproducibility**

*Please provide an assessment of the within-laboratory reproducibility of the test method (i.e. the agreement among results obtained from testing the same test items in independent experiments using the same protocol in one laboratory), if possible using bootstrap/resampling techniques and providing confidence intervals. Please specify possible sources of variability. Please append the data in the form of tables and/or figures within Attachment 4 (in EXCEL format - see section 2.2.1). If the assessment is performed with both existing (historical) data (see section 3.3) and newly generated data, please specify it in the appended tables.*

Across all laboratories, a total of 38 different test items was considered in the WLR assessment of Pfuhler et al. (2021), comprising 69 instances with at least two valid 48h-experiments (60) and at least two valid 72h-experiments (9). Note that the terminology ‘instance’ is used, as a test item could be considered for WLR twice (for 48h and for 72h experiments) per laboratory (e.g. 5-fluorouracil and n-butyl chloride of laboratory A). The number of instances for which WLR could be applied in the individual laboratories ranged from seven to 24. The WLR of the individual (four) laboratories was 74%, 88%, 86% and 93%, resulting in a WLR across laboratories of 84%. The results are summarised in Attachment 4. This analysis included one formally invalid experiment, i.e. the first 72-h experiment of laboratory B with cadmium chloride. Excluding this experiment results in a slightly reduced WLR for laboratory B of 87% and only marginally decreases the WLR across laboratories.

Since the validation study was not specifically designed to address WLR and that in the majority of cases two experiments per test item were available, the use of resampling approaches to obtain WLR distributions would not be informative and has therefore not been conducted. However, probability calculus, which is equally suitable approach, was applied to obtain the distributions of two reproducible experiments (in terms of the results being reproducibly positive/negative) for all instances per laboratory (see Figure 1).



Figure 1: Distributions of the within-laboratory reproducibility obtained by probability calculus for the individual laboratories.

Potential experimental sources of variability that can result in reduced WLR are borderline results, high or low (but acceptable) background levels, dose selection, variability of cytotoxicity outcomes across experiments using similar concentrations, precipitation, and any combination of two or more of these.

**2.2.3 Identification and discussion of outlying values**

*Please identify and discuss any outlying values.*

Cadmium chloride was not reproducible in 72h-experiments of laboratory C as the dose range, in which cadmium chloride was positive (and not cytotoxic) was hit in the second experiment only. Cadmium chloride was also not reproducible in laboratory A (72h-experiments), potentially due to differences in cytotoxicity between experiments. In general, cadmium chloride appears to be a test item that is difficult to test in the RSMN, as for example indicated by the substantially different cytotoxicity across laboratories.

In laboratory A, also ethyl methane sulfonate, n-butyl chloride, phenantrene, resorcinol and tolbutamide were not reproducible. There are no obvious reasons that could explain this observation. Note that laboratory A was the only laboratory that conducted experiments with ethyl methane sulfonate and resorcinol that informed the WLR.

In laboratory B, d-limonene and tolbutamide were not reproducible due to borderline results that were considered positive. There is no apparent explanation why diclofenac was not reproducible in laboratory B.

In laboratory C, only cadmium chloride was not reproducible (see above) and in laboratory D only 2,6-diaminotoluene (for no apparent reason).

**2.3 MODULE 3: TRANSFERABILITY**

*The transferability measures the ability of a test method protocol to be accurately and reliably performed in independent competent laboratories. It therefore provides an estimate of how much training is needed to be able to perform the test in a naïve laboratory and produce reproducible results. A naïve laboratory refers to a laboratory that is inexperienced in performing the test method. The transferability gives an indication on the robustness of a test method (e.g. its reliable performance under different conditions).*

**2.3.1 Test items used for assessing the transferability**

*Please append the full list of the test items used for assessing transferability in the form of a table as Attachment 5 (in EXCEL format), including their names, CAS numbers, commercial source, purity, physical form and any other relevant physical/chemical properties (e.g. chemical classes/organic functional groups, MW, LogP), and acquired data. Please describe the criteria applied for selecting these test items. Please specify the extent to which the test items represent the applicability domain of the test method and the range of observed effects (e.g. non-toxic to highly toxic effect). If some of these test items were also used for training laboratories or for assessing the test method's within-laboratory reproducibility, between-laboratory reproducibility and/or predictive capacity (please see section "Purpose of the TST" above, in particular point 4 – chemicals selection criteria), please indicate it under section 3.4 and justify here below their selection for assessing the test method's transferability (see also sections 2.2.1, 2.4.1 and 2.5.1).*

The transferability of the RSMN was successfully demonstrated by Hu et al. (2009). A RSMN protocol similar to the protocol used during the validation study was transferred from the developing laboratory Institute for In Vitro Sciences, Inc. (IIVS) to two American laboratories (Procter and Gamble and MatTek). Fife test items, three positives and two negatives, were correctly predicted by all laboratories. Two additional negative test items were tested by two laboratories only and were also correctly predicted.

The transferability of the RSMN was also assessed in Phase 1 of the validation study. The protocol used, as described by Dahl et al. (2011), was basically the same as used in Phase 2 of the validation study. However, RVCC was not yet used as a cytotoxicity read-out and 72h-experiments were not yet used to follow-up negative 48h-experiments. The RSMN was successfully transferred to two naïve European laboratories (L’Oréal and Henkel). Successful transfer was demonstrated using over-sea shipped tissues testing the two positive test items vinblastine sulfate and MMC. The results have been reported by Aardema et al. (2010).

The test items used for the transferability assessment are identified, including structures, CAS and EC numbers, in Attachment 5, which also contains information on the commercial supplier as well as their purity, physical form, chemical class/organic functional groups, molecular weight, octanol-water partition coefficient, water solubility, vapour pressure as well as links to websites, from which the physical-chemical properties were retrieved.

**2.3.2 Training required for transferring the test method**

*Please provide an estimation of the amount of training that is necessary to establish the test method in a naïve laboratory (i.e. a laboratory which is familiar with the techniques involved but not with the test method). If available, please append the training protocol as Attachment 6.*

The conduct of a RSMN requires standard technical laboratory skills as well as expertise in interpreting data and results of genotoxicity test methods. In addition, experience in handling 3D tissue and in defining the dose range for in vitro genotoxicity tests is important.

Each laboratory should establish experimental competency in the micronucleus assay in accordance with OECD TG 487 by demonstrating the ability to obtain single cell suspensions of sufficient quality from EpiDerm tissues. The quality of the preparations and tissues investigated will be evaluated firstly assessing the percentage of BN for untreated or solvent treated tissues, which should be at least 25% of total cells.

Further experiments, e.g. using several doses of the positive control MMC, should be conducted to establish proficiency in the method, including scoring, as well as to establish laboratory-specific historical control ranges. Respective ranges of SC of the laboratories that participated in the validation study are included in Pfuhler et al. (2021). It is recommended that laboratories use these to assess the appropriateness of their historical control data.

In case slides of one experiment are scored by more than one technician, scoring should be harmonized. It is recommended to qualify the respective personnel by scoring the same set of coded slides. A comparison of the results allows identification of potential differences or issues and to harmonise the scoring by further training. It is recommended to repeat this procedure frequently, or when new personnel are involved.

Successful training should be demonstrated by testing at least one positive or, preferably, one positive and one negative test item not previously tested, following the entire protocol, i.e. from the determination of solubility to the final call. It is recommended to select the test item(s) from the list of those tested in the validation exercise.

We consider that the recommendations made for a transfer of the assay are covered by the SOP and that a specific training protocol is not required.

**2.3.3 Obstacles pertaining to transferability that are specific to the test method**

*Please provide a summary of expected obstacles or difficulties that may impact on the transferability of the test method, e.g. level of complexity of some procedures in the protocol(s), cost, required equipment, etc.*

The critical steps for a test naïve laboratory are same as listed in section 2.1.6 i). There are no other obstacles.

**2.3.4 Assessment of the transferability to other laboratories**

*Please provide an assessment of the transferability of the submitted test method to other laboratories. Please explain the criteria applied to assess success of transfer. If available, please append the transfer report as Attachment 7. If the assessment is performed with both existing (historical) data (see section 3.3) and newly generated data, please specify it in the appended report and/or the table(s) included in Attachment 5 (see section 2.3.1).*

The transfer of the RSMN was assessed by Hu et al (2009) with seven test items tested in two US laboratories (P&G and MatTek) and by Aardema et al. (2010) in Phase 1 of the validation study with two items tested by two naïve European laboratories (L’Oréal and Henkel). Transfer success was demonstrated by repeatedly obtaining the correct results for the test items.

While MatTek and Henkel did not participate in the validation study, another laboratory that had already successfully used the RSMN, so that no formal transfer was required, participated in the validation study.

Based on the experience generated during the validation study, it is recommended that a naïve laboratory should conduct at least two experiments each with the three test items cyclohexanone (true negative), mitomycin C (positive control) and n-ethyl-n-nitrosourea (true positive). To conclude transfer success, the experiments have to meet the validity criteria and have to show clear positive results for mitomycin C and n-ethyl-n-nitrosourea, and negative results for cyclohexanone. Note that additional experiments are required to establish a reliable laboratory-specific historical control range.

**2.4 MODULE 4: BETWEEN-LABORATORY REPRODUCIBILITY**

*This module addresses the reproducibility of results in different qualified laboratories, using the same protocol and testing the same test items. The between-laboratory reproducibility is usually assessed in three well-trained laboratories. However, the number of laboratories as well as the number and type of test items should be decided according to the purpose of the validation study.*

**2.4.1 Test items used for assessing between-laboratory reproducibility**

*Please append the full list of the test items used for assessing between-laboratory reproducibility in the form of a table as Attachment 8 (in EXCEL format), including their names, CAS numbers, commercial source, purity, physical form and any other relevant physical/chemical properties (e.g. chemical classes/organic functional groups, MW, LogP), and acquired data. Please describe the criteria applied for selecting these test items. Please specify the extent to which these test items represent the applicability domain of the test method and the range of observed effects (e.g. non-toxic to highly toxic effect). If some of these test items were also used for developing and optimising the test method, training laboratories or assessing the transferability of the test method to other laboratories, please indicate it under section 3.4 and justify here below their selection for assessing the test method's between laboratory reproducibility (see also sections 2.1.7 and 2.3.1).*

The test items informing the between-laboratory reproducibility (BLR) are identified in Attachment 8, which also contains information on the commercial supplier as well as their purity, physical form, chemical class/organic functional groups, molecular weight, octanol-water partition coefficient, water solubility and vapour pressure. Structures, EC-numbers and links to websites, from which the physical-chemical properties were retrieved, are included in the first worksheet of Attachment 9.

BLR was addressed in both phases of the validation study. In phase 1, three chemicals, which were selected by an independent chemical selection expert team (Aardema et al., 2010), were tested in three laboratories. In phase 2, a more flexible approach was taken that resulted in seven test items being tested in two laboratories, nine test item being tested in three laboratories and three test items being tested in four laboratories. The test items of phase 2 were selected by external experts (R. Corvi and D. Kirkland) from a chemical master list prepared for Cosmetics Europe by a larger group of external experts.

In total, 22 test items informed the BLR. In this master list data sets were collated that had previously been investigated in in vivo genotoxicity and/or carcinogenicity studies with dermal exposure. The chemicals were primarily selected to obtain a balanced selection regarding the test items reference result (11 true positives (TP), 11 with expected negative outcome - 4 misleading positives (MP) and 7 true negatives (TN) were selected) and to address various modes of actions.

More details, e.g. the in vitro and in vivo reference results, are included in Supplemental Table S1 of Pfuhler et al. (2021) and Annex A10.

The 22 test items comprised five liquids and 17 solids, which all had a low to very low to low vapour pressure (< 7 hPa). Their molecular weight ranged from 93 to 589 and their octanol-water partition coefficient (logKOW) from -3.1 to 4.5. The substances were mainly organic representing various chemical classes and many functional groups. In addition, two metal salts were included.

Three of these 22 test items were also used during the development of the RSMN (MMC by Curren et al. (2006); n-ethyl-n-nitrosourea and 4-nitrophenol by Mun et al. (2009)). In addition, three test items were also used for the assessment of the transferability (MMC by Hu et al. (2006) and Aardema et al. (2010); methyl methanesulfonate and 4-nitrophenol by Hu et al. (2009)).

**2.4.2 Assessment of between-laboratory reproducibility**

*Please provide an assessment of the between-laboratory reproducibility of test method (i.e. the agreement among results obtained from testing the same test items using the same protocol in different laboratories), if possible using bootstrap/resampling techniques and providing confidence intervals. Usually at least three laboratories are requested to properly evaluate between-laboratory reproducibility. Please specify possible sources of variability. Please append the data in the form of tables and/or figures within Attachment 8 (in EXCEL format - see section 2.4.1). If the assessment is performed with both existing (historical) data (see section 3.3) and newly generated data, please specify it in the appended tables.*

Twentytwo (22) test items were considered in the BLR assessment, as reported by Pfuhler et al. (2021). Sixteen (16) test items had concordant and four test items (cadmium chloride, 2,4-diaminotoluene, 5-fluorouracil, phenantrene) discordant calls in two to four laboratories. The remaining two test items (resorcinol, tolbutamide) were negative in one or two laboratories and equivocal in one laboratory. The results are summarised in Attachment 8.

Considering resorcinol and tolbutamide due to the remaining uncertainty that led to equivocal results to be between-laboratory reproducible with a likelihood of 50%, the overall BLR, when weighing all test items equally, was 17/22 = 77.3%.

Probability calculus was applied to obtain the distributions of the 22 test items obtaining reproducible results in three laboratories (see Figure 2).



Figure 2: Distributions of the between-laboratory reproducibility obtained by probability calculus.

Potential experimental sources of variability that can result in reduced BLR primarily are experiments with borderline results and differences in dose selection as well as cytotoxicity between laboratories.

**2.4.3 Identification and discussion of outlying values**

*Please identify and discuss any outlying values.*

Cadmium chloride, which was also not within-laboratory reproducibility in two laboratories, was not reproducible between laboratories, potentially due to differences in cytotoxicity between experiments and laboratories. In general, cadmium chloride appears to be a test item that is difficult to test in the RSMN.

The other three not between-laboratory reproducible test items were positive in laboratory A and negative in the other laboratories. 2,4-diaminotoluene and phenanthrene were clearly positive in laboratory A in 48h-experiments and 5-fluorouracil in 72h-experiments. This may indicate a slightly increased sensitivity of the RSMN in laboratory A as compared to the other laboratories, which is also supported by the observation that laboratory A obtained equivocal and the other laboratories negative results for resorcinol and tolbutamide.

**2.5 MODULE 5: PREDICTIVE CAPACITY**

*The predictive capacity determines the ability of a test method to predict the in vivo result and/or a human health, environmental, and/or other biological effect of concern. This is usually done by relating the predictions to the result obtained with a reference method. The calculated predictive capacity is influenced by the number (sample size) and type of test items and the quality of the reference data.*

**2.5.1 Test items used for assessing predictive capacity**

*Please append the full list of the test items used for assessing predictive capacity in the form of a table as Attachment 9 (in EXCEL format), including their names, CAS numbers, commercial source, purity, physical form and any other relevant physical/chemical properties (e.g. chemical classes/organic functional groups, MW, LogP), and acquired data. Please describe the criteria applied for selecting these test items. Please specify the extent to which these test items represent the applicability domain of the test method and the range of observed effects (e.g. non-toxic to highly toxic effect). If some of these test items were also used for developing and optimising the test method, training laboratories or assessing the transferability of the test method to other laboratories, please indicate it under section 3.4 and justify here below their selection for assessing the test method's predictive capacity (see also sections 2.1.7 and 2.3.1).*

The full list of test items, including their names, CAS numbers, commercial source, purity, physical form and any other relevant physical/chemical properties (e.g. chemical classes/organic functional groups, MW, LogP), is presented in the first worksheet of Attachment 9.

As described in Pfuhler et al. (2021), Phase 2 test items were selected by external experts [Raffaella Corvi (EC, Joint Research Centre) and David Kirkland (Kirkland Consulting)] from a master list prepared for Cosmetics Europe by another group of external experts. These chemicals had previously been tested in in vivo genotoxicity and/or carcinogenicity studies, ideally with dermal exposure, and were assigned to three categories: true negative (TN) and true positive (TP) chemicals, with concordant in vitro and in vivo data, as well as misleading positives (MP) for which positive in vitro findings were reported, but not confirmed in in vivo studies.

The chemicals were primarily selected to obtain a balanced selection regarding the test items reference result (21 TP, 12 MP and 10 TN were selected) and to include various modes of actions.

More details, e.g. the modes of actions of the selected true positives, are provided in the worksheet ‘Test item reference result’ of Attachment 10.

The 43 chemicals comprised eight liquids and 35 solids, which all had a low to very low vapour pressure. Their molecular weight ranged from 93 to 854 and their octanol-water partition coefficient (logKOW) from -3.1 to 5.8. The chemicals were, with two exceptions (cadmium chloride and potassium bromate), organics representing various chemical classes and featuring numerous functional groups. Pro-mutagens (e.g., DMBA, 2-AAF) have been included to explore the potential of the RSMN to detect substances that need to be metabolically activated to show their mutagenic potential.

**2.5.2 Assessment of the predictive capacity of the test method**

*Please provide all available information on the predictive capacity of the test method, if possible using bootstrap/resampling techniques and providing confidence intervals. Please describe the accuracy values (i.e., overall accuracy, sensitivity, specificity, positive and negative predictive values, false positive and negative proportion) and associated confidence intervals of the proposed test method with respect to its ability to measure or predict the effect of interest. The accuracy values of the proposed test should be compared to that obtained for the appropriate reference test method (if available) and/or other data or information, especially from the species of interest (if available). Please append the data in the form of tables and/or figures within Attachment 9 (in EXCEL format - see section 2.5.1). If the assessment is performed with both existing (historical) data (see section 3.3) and newly generated data, please specify it in the appended tables.*

The predictive capacity of the RSMN was calculated in two ways

1) across all laboratories, based on predictive capacity (as sum of laboratories and as mean) of the individual laboratories as in done by Pfuhler et al., 2021 (see Table 4)

2) across all test items, based on the result per test item across laboratories, where applicable, whereas equivocal calls were assigned a value of 0.5. Repeatedly tested test items with discordant results were assigned a value reflecting the proportion of correct predictions.

Due to the small sample sizes, especially for laboratories C and D for approach 1), 95% confidence intervals were calculated using the exact Clopper-Pearson as implemented in the R package ‘PropCIs’ for approach 2) only. Note that the more suitable mid-p approach is not reasonably applicable to non-integer numbers and was thus not applied.

In addition, distributions for the predictive parameters estimates were obtained by probability calculus (Figure 3) - leading to the same results as resampling – for approach 2).

The results for calculation approach 1) are presented in Table 3.

Table 3: Predictive capacity (in %) per individual laboratory and summarising the parameters across laboratory by the arithmetic mean (unweighted). Numbers is brackets are the respective ratios (see also Table 4 of Pfuhler et al. (2021), which was amended here with the column ‘overall (mean)’).

Lab A Lab B Lab C Lab D overall overall  
 (mean)

Sensitivity 93.3 (14/15) 61.5 (8/13) 75.0 (6/8) 50 (2/4) 75.0 (30/40) 70.0

Specificity 71.4 (10/14) 85.7 (12/14) 100 (6/6) 90 (9/10) 84.1 (37/44) 86.8

Accuracy 82.8 (24/29) 74.1 (20/27) 85.7 (12/14) 78.6 (11/14) 79.8 (67/84) 80.2

For approach 2, the following results were obtained:

• specificity: 83.3% (18.33/22) with a 95%-confidence interval of 61.5% - 95.6%

• sensitivity: 76.2% (16/21) with a 95%-confidence interval of 52.8% - 91.8%

• accuracy: 79.8% (34.33/43) with a 95%-confidence interval of 64.8% - 90.5%



Figure 3: Distributions of the predictive parameters sensitivity, specificity, and accuracy based on 21, 22 and 43 test items.

**2.5.3 Identification and discussion of false predictions**

*Please identify and discuss any false predictions (e.g. false positives and false negatives).*

In the validation study a total of 13 test item were not correctly predicted in at least one laboratory. These comprised seven test items incorrectly classified as negative:

• 2-acetylaminofluorene (2-AAF) was tested negatively in three laboratories in 48-experiements and in two in 72h-experiemnts. As 2-AAF is a pro-mutagen that forms genotoxic metabolites. The negative results in the RSMN indicate that 2-AAF may not be sufficiently metabolised in the skin to form micronuclei.

• Cadmium chloride was tested in three laboratories. It was positive, but not within-laboratory reproducible in two laboratories and reproducibly negative in the third laboratory (B). Cadmium chloride appears to a test item that is difficult to test in the RSMN, as primarily indicated by the lack of within-laboratory reproducibility and the substantially different cytotoxicity across laboratories.

• Cyclopenta[c,d]pyrene (CPPE) was tested in two laboratories, being correctly positive in laboratory B in a 72-experiment and being incorrectly negative in laboratory C, where it was tested twice in 48-experiemnts. A likely reason for the incorrect negative result is that CPPE was not tested in a 72h-experiment.

• Cytosine arabinoside was tested in 48h and 72-experiments in laboratory B. There are no apparent reasons that could explain the incorrect negative result.

• 2,4-diaminotoluene (2,4-DAT) was positive in laboratory A in 48h-experiments, negative in laboratory D in 48h-experiments (not tested in 72h-experiments) and negative in 48h- and 72h-experiments in laboratory B. There are no apparent reasons that could explain the incorrect negative results.

• 7,12-dimethylbenz[a]anthracene (DMBA) tested negative by laboratory A in a 48h- and a 72h-experiment. DMBA is a pro-mutagen that requires activation by cytochrome P450. Basal expression of phase I enzymes has been shown to be low in both native human skin and RS tissues but can be upregulated within 24–72 h. DMBA was shown to be positive in a previous RSMN study that evaluated DMBA and other pro- mutagens, although the % VCC, which was a much more sensitive indicator of the cytotoxic effects of DMBA, in the present study was not used as a measure of cytotoxicity in the earlier study and allowed higher doses of DMBA to be tested.

• 5-fluorouracil was correctly positive in three 72h-experiments in laboratory A, but negative in laboratory B (48h- and 72h-experiments). There are no apparent reasons that could explain the incorrect negative result of laboratory B.

In addition, incorrect positive results were obtained for four test items and equivocal results for two test items:

• Curcumin was found to be positive by laboratory B in two 48h-experiments. As at the beginning of Phase 2 RVCC, which was the more sensitive cytotoxicity endpoint (several tissues below 40%), was not yet used as an acceptance criterion, these results could potentially be explained by excessive cytotoxicity of the doses tested.

• Diclofenac was tested by three laboratories with positive results after 48h and 72h. Pfuhler et al. (2021) discuss this substance in detail, inter alia observing that Diclofenac is a non-steroidal anti-inflammatory drug (NSAID) and also the active ingredient in dermal drugs for treating pre-cancerous skin lesions. The mechanism of action is not completely understood but likely involves the induction of apoptosis and the normalisation of metabolism and immune cell infiltration and function in actinic keratosis lesions via cyclooxygenase (COX-1/-2)-dependent and independent pathways. To note, Lab A reported a strong increase of apoptotic cells alongside the increase of micronuclei, a well-known cause of false-positive results in MN scoring.

• Eugenol was incorrectly positive in laboratory A after 48h and 72h. There are no apparent reasons that could explain the incorrect positive result.

• Phenantrene was tested by four laboratories, with correct negative results in laboratories C and D (48h-experiments only) and in laboratory B (48h- and 72h-experiments). It showed variable results in five 48-h experiments in laboratory A (three positive and two negative), with an overall call as positive. There are no apparent reasons that could explain the incorrect positive result of laboratory A.

• Resorcinol was correct negative in laboratory B (48h and 72h). Laboratory A conducted three 48h-experiments, two being correct negative and one being positive. Considering the higher solvent control response of the positive experiment, the overall call of laboratory A was equivocal.

• Tolbutamide was tested in three laboratories. Based on two 48h-experiments each, it was negative in laboratories B and C. Laboratory A conducted four 48h-experiments, two being negative and two being positive. The first positive experiment showed an unclear dose-response curve with precipitation of the two highest doses. In the second positive experiment only the highest dose, which precipitated, was positive, marginally exceeding the historical control range. As these responses could not be reproduced in two subsequent experiments, tolbutamide was considered equivocal in laboratory A.

These incorrect predictions need, however, be also viewed in the context of the proposed application of the RSMN, i.e. as one of three higher tier test methods (RS Comet assay and HET-MN) to follow-up potentially incorrect positive results from in the standard in vitro test battery.

**2.5.4 Quality/variability/uncertainty of the reference data used to calculate predictive capacity**

*Please provide information on the origin (i.e. source), quality and relevance of the reference data used to calculate the predictive capacity. Please append all available reference data as Attachment 10 (in EXCEL format). If possible, please comment on the uncertainty associated with the reference data used to conclude on the properties of the test items (e.g. in vivo classification), considering its relevance for the species of interest (e.g. humans when predicting human health effects) and the consistency between multiple reference data for the same test item (weight-of-evidence assessment). If multiple reference data of the same type (e.g. obtained with the same reference method) exist for each test item, please report them and comment on the variability of the data and its potential influence on the calculation of predictive capacity, and justify the final choice of reference classification used for each test item having multiple reference data available.*

All relevant information on the reference data used to determine the reference results, i.e. TP. MP and TN, are summarized in Attachment 10 ‘A10 reference data 210629.xls’. The quality and reliability of the reference data was the primary selection criterion. It needs to be acknowledged, however, that potentially additional relevant evidence has been generated since the selection of the test items. It is assumed that that evidence will not lead to changes in reference results.

TP test items reference results have been established based on evidence for carcinogenicity from rodent studies and their genotoxic mode of action, with 5-fluorouracil being the only exception as it was considered as a non-carcinogenic in vivo genotoxicant. Most TP have been assigned to the IARC (International Agency for Research on Cancer) groups 1 (carcinogenic to humans), 2A (probable human carcinogen) or 2B (possibly carcinogenic to humans). Due to the consistency of mechanistic and experimental evidence, the uncertainty associated with the reference result of the TP is low.

The majority of TN test items were negative for carcinogenicity (not dermal route), and mainly negative in in vivo and in vitro studies, so that they are considered to have low uncertainty. Also the uncertainty associated with additional three TN substances that have no in vivo data, mainly negative in vitro data, and negative carcinogenicity data (not dermal route), is considered low. In addition, one non-genotoxic carcinogen (d-limonene) with a well-elucidated, human irrelevant mode of action (MoA) was included, which was carcinogenic in rodent studies (not dermal route) and negative in in vivo and in vitro genotoxicity studies. Due to the MoA, the uncertainty in the reference result is also considered low.

MP test items include six test items that were positive in some in vitro genotoxicity studies, predominantly negative in in vivo genotoxicity studies and negative in rodent carcinogenicity studies (one (4-nitrophenol) tested via the dermal route). Considering the in vitro studies as misleading positive due to their low specificity, uncertainty associated with these reference results is low. In addition, the MP group contains four non-carcinogens with predominantly positive in vitro and in vivo results. The uncertainty associated with the reference results for these four test items is slightly higher than compared to the other MP test items, as it is for the non-genotoxic carcinogen phenol, which was negative in a dermal mouse carcinogenicity study, but was consistently positive in vitro and no in vivo genotoxicity studies.

**2.6 MODULE 6: APPLICABILITY DOMAIN**

*The Applicability Domain of an in vitro test method is defined by its limitations and by considerations of the physico-chemical or other properties of the chemicals for which a method is applicable for use as determined from empirical testing.*

**2.6.1 Limitations of the test method identified through testing**

*Please describe where possible, the chemical categories (e.g. based on organic functional groups and/or physicochemical properties) for which the test method does not make reliable (i.e. generates variable predictions) and/or relevant predictions (i.e. generates false predictions).*

The data generated during the validation for 43 test items does not indicate any chemical category for which the reliability or relevance of predictions is reduced. Some false predictions could be explained (see chapter 2.5.3). For several test items no reasons that could explain the incorrect result were apparent. The number of test items with incorrect results is, however, too low for a meaningful analysis of the applicability domain in terms of chemical structure and physicochemical properties.

Noting that the eligibility of test items was primarily limited by the availability of reliable reference data, the applicability domain of the RSMN may need to be evaluated further. Since the assay is in the meantime used commercially it is also expected that published data will contribute to an extension of the applicability domain.

**2.6.2 Technical limitations of the test method**

*If applicable, please identify on the basis of experimental evidence the known technical limitations of the test method (e.g. test method not applicable to the testing of poorly soluble materials).*

The only known technical limitation is related to test item solubility. If a test item is not sufficiently soluble according to the procedure described in the protocol in one of the two established solvents (acetone and 70% ethanol), a different solvent may be used. However, the suitability of the new solvent, i.e. that it does not induce cytotoxicity and does not modify the characteristics of test item, has to be established first in order to allow for a sound interpretation of results. Therefore, the RSMN is limited to substances that are soluble in a suitable solvent.

Note that while a solvent may appear suitable at the beginning of an experiment, the test item may have precipitated on the surface of the skin model as visually observed by the end of an experiment. It is therefore recommended to use the lowest precipitating concentration as the highest dose tested in the main experiment.

**2.6.3 Mechanistic limitations of the test method**

*If applicable, please define the mechanistic limitations of the test method related to known modes of action (e.g. skin sensitisation test method only applicable to chemicals binding to cysteine residues).*

For the intended purpose, no mechanistic limitations of the RSMN are known. With 43 substances of different chemical classes tested so far, no mechanistic limitations of the RS Comet assay have been identified. However, its applicability will need to be evaluated further, e.g. for mixtures.

**2.7 MODULE 7: PERFORMANCE STANDARDS**

*Performance Standards of (a) validated test method(s) are usually defined upon completion of a validation study and are only required at OECD level when there are proprietary issues involved. In the absence of proprietary elements in the test method, this section can therefore be considered as optional.*

*The three elements of Performance Standards are: (i) essential test method components, (ii) a minimum list of reference test items and (iii) defined accuracy and reliability values (see OECD Guidance Document No. 34; OECD, 2005). They can be used to evaluate the reliability and accuracy of proposed similar or updated test methods (e.g. modifications of already validated test methods) with respect to the validated reference method(s).*

*If available as a separate document, please append the Performance Standards as Attachment 11.*

**2.7.1 Suggestions of essential test method components for the evaluation of similarity**

*Please indicate the essential structural, functional and procedural components of a validated test method that should be retained in a proposed test method to be judged as similar.*

Suggestions will be provided at a later stage.

**2.7.2 Suggestions of possible reference test items**

*Please suggest test items which should be used to assess the accuracy and reliability of a similar test method. These test items should be a representative subset of those used to demonstrate the reliability and accuracy of the validated test method.*

Suggestions will be provided at a later stage.

**2.7.3 Suggestions of defined accuracy and reliability values**

*Please suggest the target accuracy and reliability values that should be achieved by the similar test method when evaluated using the reference test items.*

Suggestions will be provided at a later stage.

**3. ADDITIONAL INFORMATION**

**3.1 Study organisation and management (Project Plan)**

*Where available, please append as Attachment 12 the project plan developed prior to the initiation of the study, describing the study objective and goals, the study organisation and the study management. If a project plan is not available, please summarise here below how the study was designed (e.g. number of participating laboratories, number of replicates within a single experiment, number of independent repetitions, etc.) and managed, and specify the objective and goals of the study (e.g. evaluation of the transferability of the test method, evaluation of within-laboratory reproducibility and/or between-laboratory reproducibility, and/or assessment of predictive capacity).*

For the organisation and management, a steering committee was formed by the sponsor Cosmetics Europe (at the time still called COLIPA) that included representatives of the participating laboratories as well as several experts, most of which had formed the chemical selection committee. The original composition, as described by Aardema et al. (2010), is summarised in Table 4. Over time, some institutions and roles changed, especially as during the later stages less steering was required. Note that the initial chemical selection was supported by Thomas Slaga (UTHSCSA-Pharmacology), Johannes Doehmer (GenPharmTox BioTech AG) and Günter Speit (University Ulm).

Table 4: Steering committee members and roles (modified from Aardema et al., 2010)

Name (alphabetical order) Institution (role)

Marilyn Aardema initially Procter & Gamble (chair); later Bioreliance (laboratory  
 representative)

Raffaella Corvi EURL-ECVAM (independent expert)

Rodger Curren IIVS (member and laboratory representative)

Cyrille Krul TNO (member)

Gladys Ouédraogo L’Oréal (member and laboratory representative)

Stefan Pfuhler Procter & Gamble (member, later chair)

Kerstin Reisinger Henkel (co-chair)

David Kirkland Covance Laboratories (chemical selection)

Jan van Benthem RIVM (independent expert)

The study was designed by the steering committee with the aim to validate - guided by international validation requirements (OECD, 2005) - the RSMN for use within regulatory genotoxicity hazard identification testing strategies to follow up positive results from the classical in vitro test battery. The modular validation approach by Hartung et al. (2004) was followed, focusing on the modules of test definition, transferability, with- and between-laboratory reproducibility and predictive capacity. The test method protocol was defined at the beginning (called Phase 0) of the validation activities (Dahl et al., 2011). Transferability, which has been demonstrated earlier by Hu et al. (2009), was confirmed (Aardema et al., 2010). The validation was conducted in two main phases. In Phase 1, within-laboratory reproducibility was assessed using three test items. Predictivity was assessed in Phase 2a with the experimental duration of 48h. Subsequently, the protocol was amended by adding a 72h-experiement to increase the sensitivity of the RSMN. Therefore, nine test items that were already tested in Phase 2a were re-tested in Phase 2b in 72h-experiments. As the Steering Committee concluded that a more thorough assessment of the performance of the extended 72-h protocol was needed, a Phase 2c study was conducted, in which a total of 12 coded chemicals (six TP and six MP/TN) from the previous Phases 2a and 2b were retested in two labs each. In Phase 2c, chemicals were first tested using the 48-h protocol, and any chemical that showed negative or equivocal results was retested using the 72-h protocol. In the last phase (2d), additional five coded chemicals that were chosen for testing by the external experts to close a data gap with chemicals that were tested in the RS Comet assay only (but were positive in both the Ames and a clastogenicity assay) were tested in one laboratory. Figure 4 (of the file 'TM2020-04\_TST with Tabs and Figs RSMN 210910'provides an overview of the various phased of the RSMN validation study.

While three laboratories (A, B and D) participated in Phase 1, laboratory C joined the validation study in Phase 2a. Laboratory D participated in Phases 1 and 2a only. Phase 2d testing was conducted by laboratory A only.

The test items of Phases 1 and 2a were coded by Covance Laboratories and BioTeSys GmbH coded the test items for the other phases, where applicable. All data were analysed by an independent statistician (Sebastian Hoffmann, seh consulting + service), who was contracted by EURL-ECVAM for Phases 1 and 2a, and for Phases 2b, 2c and 2d by Cosmetics Europe.



Figure 4. RSMN validation timeline (from Pfuhler et al., 2021)

**3.2 History of test method development**

*Please provide any information on the process of developing the test method that might be of relevance for its validation.*

The RSMN was developed Curren et al. (2006), who had experience working with the EpiDerm™ skin model for skin irritation testing, in collaboration with Procter & Gamble. They established the 48h-protocol and % binucleation as a cytotoxicity readout. The RSMN was further developed by Mun et al. (2009), who explored the applicability by testing seven genotoxins and five non-genotoxins. Hu et al. (2009) reported successful transfer to other laboratories and showed reproducibility of results for EpiDermTM tissues from different donors.

**3.3 Collection of existing (historical) information**

*If some or all data use in this submission come from existing (historical) information/data (retrospective assessment) collected from different sources (such as validation studies, in-house data, unpublished company data, publications, communications, etc.), please summarise and number the information/data sets collected in a Table using the following format and append it as Attachment 13 (in EXCEL format).*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Continuous number of information/data set* | *Identification of information/data set retrieved* | *Proprietary data included (YES/NO)* | *Source* | *Indicate under which module(s) the information/data were considered* | *Justification for inclusion in this submission on the basis of the selection criteria used (e.g. data quality, test system used, protocol used, etc.)* |
| Not applicable | | | | | |

**3.4 Test items used in the submission**

*Please append as Attachment 14 a list of all test items used in the submission (including name and CAS number), indicating in which of the following phases they were used: (i) developing and optimising the test method (protocol and prediction model), (ii) training, (iii) transferability, (iv) within-laboratory reproducibility, (v) between laboratory reproducibility, (vi) predictive capacity, and/or (vii) performance standards.*

**3.5 Quality system(s) used in the generation of the data submitted in this TST**

*Please state the quality system(s) in place in the laboratories involved (e.g. GLP, ISO). For GLP-like conditions, please specify the extent and area of compliance. Please indicate the extent of coverage of good in vitro method practices (GIVIMP) (OECD, 2018).*

To 3.4. (due to technical issues): Attachment 14 list all test items and indicate for which purposes they have been used.

The acquisition of the data for the assessment of the reproducibility and predictive capacity did not follow any formal quality assurance system. However, the following safeguards recommended by Balls et al. (1995) were applied:

• qualified personnel, and appropriate facilities, equipment and materials were available

• records of the qualifications, training and experience, and a job description for each professional and technical individual, were maintained

• for each study, an individual with appropriate qualifications, training and experience was appointed to be responsible for its overall conduct and for any report issued

• instruments used for the generation of experimental data were inspected regularly, cleaned, maintained and calibrated according to manufacturers' instructions. Records of these processes were kept.

• reagents were labelled, as appropriate, to indicate their source, identity, concentration and stability. The labelling included the preparation and expiry dates, and specific storage conditions.

• all data generated during a study were recorded by the individual(s) responsible, entries were attributable and dated

In addition, test items were provided coded from a central repository at Covance Laboratories and, at a later stage, BioTeSys GmbH (Germany). Data were submitted to an independent statistician for analysis. The initial data analysis was conducted with codes still in place.

**3.6 Study conclusions**

*Please provide your conclusions regarding the outcome of the study [e.g. is the test method easily transferable to other laboratories? Is the test method generating reproducible results within a single laboratory and between laboratories? Is the test method sufficiently accurate compared to the reference data (e.g. in vivo data)/target accuracy and reliability values (for performance standards-based validation studies)?], in the light of the study objectives and in consideration of the extent to which such conclusions are supported by the study results.*

The validation study demonstrated that the RSMN

• works well with the MatTek Epi-200TM Skin model, with reproducibly low background DNA damage and sufficient metabolic capacity to activate pro-mutagens

• can be transferred to other laboratories

• is well reproducible within and between laboratories

• has sufficiently high predictive capacity for the intended use as a higher tier to follow-up potentially false positive results from the standard in vitro test battery, especially if used in combination with the RS Comet assay – as intended, when the expected route of exposure is dermal (Pfuhler et al, 2020)

In conclusion, the RSMN is considered sufficiently validated for its intended use by the test developers, as supported by its acceptance into the OECD’s test guideline development program.

**3.7 Recommendations**

*Considering the study objectives and outcome, please provide recommendations, where applicable, regarding a) possible improvements of the test method (e.g. in relation to the SOP, prediction model etc.), b) future activities to be undertaken to better characterise the performance of the test method in view of its envisaged use (e.g. better characterisation of the predictive capacity by testing additional coded chemicals), c) the possible current use of the test method (e.g. screening method, method as part of a testing strategy, replacement method), d) any other recommendation on future activities.*

Based on the thorough characterisation of the performance through a formal validation study, the RSMN is considered fit-for-purpose, as demonstrated by:

- its use in two ingredient safety assessments of the Scientific Committee on Consumer Safety (SCCS 2019a; SCCS 2019b)

- its inclusion into ‘The SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation’ since the addendum to the 8th revision in 2014 (SCCS, 2021), as well as a recommendation by an international external expert group (IWGT recommendation, Pfuhler et al, 2020).

Increased use for genotoxic hazard assessment, both in research and regulatory contexts, would inform further protocol optimisation, e.g., a more efficient cytotoxicity testing procedure. More importantly, a broader database would allow to better define the applicability domain and its limitation, as well as its performance as a higher tier test method within genotoxicity assessment strategies. Such data are being generated as the RSMN is already offered and used commercially.

**3.8 Additional information**

*Please add any additional information that was not covered in the TST.*

All relevant information on the EpiDermTM-based RSMN has been included in this TST. Note that the RSMN is currently also being validated using the 3D EpiSkin® model.

**3.9 List of references**

*Please provide a list of scientific or other appropriate references, briefly describing their relevance with regard to the submitted test method. Please include all publications that provide background information on the test method’s biological and mechanistic relevance. If available, please include publications that provide direct information of the performance of the proposed test method. In cases of submissions of putative similar or updated test methods, please include publications relating to the validated reference test method(s) on which – if available – the performance standards are based. Please append publications as electronic files (pdf or scans of paper copies) as Attachment 15.*

All references cited in this TST are listed here, grouped according to their relevance. Full texts are provided as Attachment 15 (Please respect copyrights where applicable!).

**Validation guidance**

• Balls M. Blaauboer, BJ, Fentem J, et al (1995). Practical Aspects of the Validation of Toxicity Test Procedures: The Report and Recommendations of ECVAM Workshop 5. Altern Lab Anim. 23(1):129-147.

• Hartung T, Bremer S, Casati S, et al. (2004). A modular approach to the ECVAM principles on test validity. Altern Lab Anim. 32(5):467-72.

• OECD (2005). Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. Environmental Health and Safety Monograph Series on Testing and Assessment No. 34. OECD Publishing, Paris.

Development and validation of the RSMN

• Curren RD, Mun GC, Gibson DP, Aardema MJ. (2006). Development of a method for assessing micronucleus induction in a 3D human skin model (EpiDerm). Mutat Res. 607(2):192-204

• Mun GC, Aardema MJ, Hu T, et al. (2009). Further development of the EpiDerm 3D reconstructed human skin micronucleus (RSMN) assay. Mutat Res. 673(2):92-9.

• Hu T, Kaluzhny Y, Mun GC, et al. (2009). Intralaboratory and interlaboratory evaluation of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. Mutat Res. 673(2):100-8.

• Aardema MJ, Barnett BC, Khambatta Z, et al. (2010). International prevalidation studies of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. Mutat Res. 701(2):123-31.

• Dahl EL, Curren R, Barnett BC, et al. (2011). The reconstructed skin micronucleus assay (RSMN) in EpiDerm™: detailed protocol and harmonized scoring atlas. Mutat Res. 720:42-52.

• Pfuhler S, Downs TR, Hewitt NJ, et al. (2021). Validation of the 3D reconstructed human skin micronucleus (RSMN) assay: an animal-free alternative for following-up positive results from standard in vitro genotoxicity assays. Mutagenesis. 36(1):1-17.

**Investigations of the properties of the skin model (EpiDermTM) and the RSMN**

• Hu T, Khambatta ZS, Hayden PJ, et al. (2010). Xenobiotic metabolism gene expression in the EpiDerm in vitro 3D human epidermis model compared to human skin. Toxicol In Vitro. 24(5):1450-63.

• Götz C, Hewitt NJ, Jermann E, et al. (2012). Effects of the genotoxic compounds, benzo[a]pyrene and cyclophosphamide on phase 1 and 2 activities in EpiDerm™ models. Xenobiotica. 42(6):526-37

• Aardema MJ, Barnett BB, Mun GC, et al. (2013). Evaluation of chemicals requiring metabolic activation in the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. Mutat Res. 750(1-2):40-9.

• Yuki K, Ikeda N, Nishiyama N, Kasamatsu T. (2013). The reconstructed skin micronucleus assay in EpiDerm™: reduction of false-positive results - a mechanistic study with epigallocatechin gallate. Mutat Res. 757(2):148-57

• Hewitt NJ, Edwards RJ, Fritsche E, et al. (2013). Use of human in vitro skin models for accurate and ethical risk assessment: metabolic considerations. Toxicol Sci. 133(2):209-17.

• Roy S, Kulkarni R, Hewitt NJ, Aardema MJ (2016). The EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: Historical control data and proof of principle studies for mechanistic assay adaptations. Mutat Res Genet Toxicol Environ Mutagen. 805:25-37.

• Kidd D, Phillips S, Chirom T, et al (2021). The 3D reconstructed skin micronucleus assay: considerations for optimal protocol design. Mutagenesis. 36(1):37-49.

**Strategic use of the RSMN**

• Pfuhler S, Kirst A, Aardema M, et al. (2010). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. Regul Toxicol Pharmacol. 57(2-3):315-24 (not cited here)

• Pfuhler S, van Benthem J, Curren R, et al (2020). Use of in vitro 3D tissue models in genotoxicity testing: Strategic fit, validation status and way forward. Report of the working group from the 7th International Workshop on Genotoxicity Testing (IWGT). Mutat Res. 850-851:503135.

• SCCS. 2021. SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation 11th revision, SCCS/1628/21.

**Use of the RSMN in regulatory contexts**

• SCCS (2019a). Opinion on hair dye 1,2,4-trihydroxybenzene (1,2,4-THB). SCCS/1598/18.

• SCCS (2019b). Opinion on Methoxypropylamino Cyclohexenylidene Ethoxyethylcyanoacetate (S87). SCCS/1605/19.

**Fundamental micronucleus literature**

• Corvi, R., Albertini, S., Hartung, T., et al. (2008). ECVAM retrospective validation of in vitro micronucleus test (MNT). Mutagenesis. 23; 271-283.

• Hayashi M. (2016). The micronucleus test-most widely used in vivo genotoxicity test. Genes Environ. 38:18.

• OECD (2016a). Test No. 487: In Vitro Mammalian Cell Micronucleus Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris.

• OECD (2016b). Test No. 475: Mammalian Bone Marrow Chromosomal Aberration Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris.

• OECD (2016c). Test No. 474: Mammalian Erythrocyte Micronucleus Test, OECD Publishing, Paris.

• Sommer S, Buraczewska I, Kruszewski M. (2020). Micronucleus Assay: The State of Art, and Future Directions. Int J Mol Sci.;21(4):1534.

**4. ATTACHMENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Attachment Number** | **Description** | **Tick if attached** | **File name** |
| Attachment 1 | Protocol(s) of the test method |  | A1 RSMN protocol 210913 |
| Attachment 2 | Test items used to develop and optimise the test method |  | A2 RSMN test iten developments and optimisation 210902 |
| Attachment 3 | If applicable, Performance Standards used to validate the similar or updated test method |  |  |
| Attachment 4 | Test items used to assess within-laboratory reproducibility, acquired data and their assessment |  | A4 RSMN WLR 210902 |
| Attachment 5 | Test items used to assess transferability |  | A5 RSMN test items transferability 210902 |
| Attachment 6 | Training protocol (if available) |  |  |
| Attachment 7 | Transfer report (if available) |  |  |
| Attachment 8 | Test items used to assess between-laboratory reproducibility, acquired data and their assessment |  | A8 RSMN BLR 210902 |
| Attachment 9 | Test items used to assess predictive capacity, acquired data and their assessment |  | A9 RSMN detailed validation results 210902 |
| Attachment 10 | Reference data used to calculate the predictive capacity including their origin, quality and relevance |  | A10 RSMN reference data 210910 |
| Attachment 11 | Performance Standards developed on the basis of the submitted method (for validation of similar or updated test methods) |  |  |
| Attachment 12 | Project Plan of the validation study |  |  |
| Attachment 13 | Existing (historical) information/data collected for retrospective assessment |  |  |
| Attachment 14 | List of all test items included in the submission and the different phases of method development/validation in which they were used |  | A14 RSMN Overview table of substances 210902 |
| Attachment 15 | Relevant references |  | A15 RSMN References 1-3 |

***NOTE:*** *Please label appended files by indicating the relevant Attachment Number at the beginning of the file name. If more than one file needs to be attached for each description, please use ZIP compression.*

**5. ABBREVIATIONS**

* CAS Chemical Abstracts Service
* DARPins Designed Ankyrin Repeat Proteins
* DB-ALM DataBase on ALternative Methods
* EURL ECVAM European Union Reference Laboratory for Alternatives

to Animal Testing

* GLP Good Laboratory Practice
* ISO International Organization for Standardization
* LogP Log of partition coefficient between an aqueous and lipophilic

phases

* MW Molecular Weight
* OECD Organisation for Economic Co-operation and Development
* SOP Standard Operating Procedure
* TST Test Submission Template

**6. REFERENCES**

Balls M., Blaauboer B.J., Fentem J.H., Bruner L., Combes R.D., Ekwall B., Fielder R.J., Guillouzo A., Lewis R.W., Lovell D.P., Reinhardt C.A., Repetto G., Sladowski D., Spielmann H., Zucco F. (1995). Practical Aspects of the Validation of Toxicity Test Procedures. The Report and Recommendations of ECVAM Workshop 5. Altern. Lab. Anim. 23, 129-147.

Curren R.D., Southee J.A., Spielmann H., Liebsch M., Fentem J.H., Balls M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods. Altern. Lab. Anim. 23, 211-217.

Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A., Zuang V. (2004). A modular approach to the ECVAM principles on test validity. Altern. Lab. Anim. 32, 467-472.

OECD (2005). Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. Series on Testing and Assessment, No. 34. OECD, Paris.

OECD (2018). Guidance document on Good *In Vitro* Method Practices (GIVIMP) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment. Series on Testing and Assessment, *in press*. OECD, Paris.

1. The so-called “DB-ALM” protocols are EURL ECVAM’s reporting format for the dissemination of a test method protocol via the EURL ECVAM database web service on alternative methods. More information can be found at: <https://ecvam-dbalm.jrc.ec.europa.eu/home/contribute>. [↑](#footnote-ref-1)