新規試験法提案書

皮膚腐食性試験代替法
ヒト表皮モデル法

平成29年 6 月

国立医薬品食品衛生研究所

新規試験法提案書

平成 29 年 6 月 1 日

No.　2017-01

皮膚腐食性試験代替法ヒト表皮モデル法に関する提案

平成 29 年 5 月 11 日に東京、国立医薬品食品衛生研究所にて開催された新規試験法評価会議（通称：JaCVAM 評価会議）において以下の提案がなされた。

提案内容： ヒト表皮モデル（EpiSkin™、EpiDerm™、SkinEthic™、epiCS®）を用いた皮膚腐食性試験法において陽性の結果が得られた場合、被験物質を腐食性物質（国連 GHS 分類における区分 1）と判定することは可能であるが、UN GHS 分類の細区分のためには EpiSkin™ のみ利用可能である。なお、本試験法の利用にあたっては、適用範囲を十分に配慮した上で使用されるべきである。

この提案書は、Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 431 *In vitro* skin corrosion: reconstructed human epidermis (RHE) test methodをもとに、皮膚腐食性試験資料編纂委員会によりまとめられた文書を用いて、JaCVAM評価会議が評価および検討した結果、その有用性が確認されたことから作成された。

以上の理由により、行政当局の安全性評価方法として皮膚腐食性試験代替法ヒト表皮モデル法の使用を提案するものである。

大野泰雄

JaCVAM 評価会議　議長

西川秋佳

JaCVAM 運営委員会　委員長

JaCVAM 評価会議

大 野 泰 雄　（公益財団法人　木原記念横浜生命科学振興財団）：座長
五十嵐良明　（国立医薬品食品衛生研究所）
石 井 雄 二　（国立医薬品食品衛生研究所　安全性生物試験研究センター）
井 上 智 彰　（日本免疫毒性学会）
今 井 教 安　（日本動物実験代替法学会）
岩瀬裕美子　（日本製薬工業協会）
篠 田 和 俊　（独立行政法人　医薬品医療機器総合機構）
杉山真理子　（日本化粧品工業連合会）
仲 井 俊 司　（日本化学工業協会）
中村るりこ　（独立行政法人　製品評価技術基盤機構）
西 川 秋 佳　（国立医薬品食品衛生研究所　安全性生物試験研究センター）
沼 澤　　聡　（日本毒性学会）
野 口　真 希　（独立行政法人　医薬品医療機器総合機構）
森 田　　健　（日本環境変異原学会）
横 関 博 雄　（日本皮膚アレルギー・接触皮膚炎学会）

任期：平成 28 年 4 月 1 日～平成 30 年 3 月 31 日

JaCVAM 運営委員会

西 川 秋 佳　（国立医薬品食品衛生研究所　安全性生物試験研究センター）：委員長
川 西　　徹　（国立医薬品食品衛生研究所）
大 原　　拓　（厚生労働省　医薬・生活衛生局　医薬品審査管理課）
小川久美子　（国立医薬品食品衛生研究所　安全性生物試験研究センター　病理部）
加 藤　　篤　（国立感染症研究所）
諫 田 泰 成　（国立医薬品食品衛生研究所　安全性生物試験研究センター　薬理部）
日下部哲也　（厚生労働省　医薬・生活衛生局　医薬品審査管理課　化学物質安全対策室）
篠 田 和 俊　（独立行政法人　医薬品医療機器総合機構）
高 木 篤 也　（国立医薬品食品衛生研究所　安全性生物試験研究センター　毒性部
　　動物管理室）
東 野 正 明　（厚生労働省　医薬・生活衛生局　医薬品審査管理課　化学物質安全対策室）
日 田　　充　（厚生労働省　医薬・生活衛生局　医薬品審査管理課　化学物質安全対策室）
平 林 容 子　（国立医薬品食品衛生研究所　安全性生物試験研究センター　毒性部）
広 瀬 明 彦　（国立医薬品食品衛生研究所　安全性生物試験研究センター　安全性予測評価部）
廣 田 光 恵　（独立行政法人　医薬品医療機器総合機構）
本 間 正 充　（国立医薬品食品衛生研究所　安全性生物試験研究センター　変異遺伝部）
小 島　　肇　（国立医薬品食品衛生研究所　安全性生物試験研究センター　安全性予測評価部
　　第二室）：事務局

JaCVAM Statement on the
In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method

At a meeting held on 11 May 2017 at the National Institute of Health Sciences (NIHS) in Tokyo, Japan, the Japanese Center for the Validation of Alternative Methods (JaCVAM) Regulatory Acceptance Board unanimously endorsed the following statement:

Proposal: Although a positive result in an in vitro skin corrosion test using human skin models such as EpiSkin™, EpiDerm™, SkinEthic™, or epiCS® is generally considered sufficient for predicting a test chemical to cause skin corrosion under UN GHS Category 1, only skin corrosion tests using EpiSkin are considered sufficient for predicting a test chemical to cause skin corrosion under the UN GHS subcategories. Furthermore, thorough consideration must be given to the applicability domain when using this test.

This statement was prepared following a review of the Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 431 In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method together with other materials prepared by the Skin Corrosion Testing JaCVAM Editorial Committee to acknowledge that the results of a review and study by the JaCVAM Regulatory Acceptance Board have confirmed the usefulness of this assay.

Based on the above, we propose the In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method as a useful means for assessing skin corrosion potential during safety assessments by regulatory agencies.

Yasuo Ohno
Chairperson
JaCVAM Regulatory Acceptance Board

Akiyoshi Nishikawa
Chairperson
JaCVAM Steering Committee

June 1, 2017

The JaCVAM Regulatory Acceptance Board was established by the JaCVAM Steering Committee, and is composed of nominees from the industry and academia.

This statement was endorsed by the following members of the JaCVAM Regulatory Acceptance Board:

Mr. Yasuo Ohno (Kihara Memorial Yokohama Foundation for the Advancement of Life Sciences) : Chairperson
Mr. Yoshiaki Ikarashi (National Institute of Health Sciences: NIHS)
Mr. Noriyasu Imai (Japanese Society for Alternatives to Animal Experiments)
Mr. Tomoaki Inoue (Japanese Society of Immunotoxicology)
Mr. Yuji Ishii (Biological Safety Research Center: BSRC, NIHS)
Ms. Yumiko Iwase (Japan Pharmaceutical Manufacturers Association)
Mr. Takeshi Morita (Japanese Environmental Mutagen Society)
Mr. Shunji Nakai (Japan Chemical Industry Association)
Ms. Ruriko Nakamura (National Institute of Technology and Evaluation)
Mr. Akiyoshi Nishikawa (BSRC, NIHS)
Ms. Maki Noguchi (Pharmaceuticals and Medical Devices Agency)
Mr. Satoshi Numazawa (Japanese Society of Toxicology)
Mr. Kazutoshi Shinoda (Pharmaceuticals and Medical Devices Agency)
Ms. Mariko Sugiyama (Japan Cosmetic Industry Association)
Mr. Hiroo Yokozeki (Japanese Society for Dermatoallergology and Contact Dermatitis)

Term: From 1st April 2016 to 31st March 2018

This statement was endorsed by the following members of the JaCVAM steering Committee after receiving the report from JaCVAM Regulatory Acceptance Board:

Mr. Akiyoshi Nishikawa (BSRC, NIHS): Chairperson
Mr. Toru Kawanishi (NIHS)
Mr. Mitsuru Hida (Ministry of Health, Labour and Welfare)
Ms. Yoko Hirabayashi (Division of Toxicology, BSRC, NIHS)
Mr. Akihiko Hirose (Division of Risk Assessment, BSRC, NIHS)
Ms. Mitsue Hirota (Pharmaceutical & Medical Devices Agency)
Mr. Masamitsu Honma (Division of Genetics and Mutagenesis, BSRC, NIHS)
Mr. Yasunari Kanda (Division of Pharmacology, BSRC, NIHS)
Mr. Atsushi Kato (National Institute of Infectious Diseases)
Mr. Tetsuya Kusakabe (Ministry of Health, Labour and Welfare)
Ms. Kumiko Ogawa (Division of Pathology, BSRC, NIHS)
Mr. Taku Oohara (Ministry of Health, Labour and Welfare)
Mr. Kazutoshi Shinoda (Pharmaceuticals and Medical Devices Agency)
Mr. Atsuya Takagi (Animal Management Section of the Division of Toxicology, BSRC, NIHS)
Mr. Masaaki Tsukano (Ministry of Health, Labour and Welfare)
Mr. Hajime Kojima (Division of Risk Assessment, BSRC, NIHS): Secretary

皮膚腐食性試験代替法

ヒト表皮モデル法

目　　次

評価会議報告書 ---------- 1

評価報告書 ---------- 7

OECD GUIDELINE FOR THE TESTING OF CHEMICALS 431,
In vitro skin corrosion: reconstructed human epidermis (RHE) test method ----------35

PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO RECONSTRUCTED HUMAN EPIDERMIS (RHE) TEST METHODS FOR SKIN CORROSION TESTING AS DESCRIBED IN TG 431 ----------61

評価会議報告書

皮膚腐食性試験代替法

ヒト表皮モデル法

JaCVAM 評価会議

平成 29 年（2017 年）5 月 11 日

JaCVAM 評価会議

大 野 泰 雄　（公益財団法人　木原記念横浜生命科学振興財団）：座長
五十嵐良明　（国立医薬品食品衛生研究所　生活衛生化学部）
石 井 雄 二　（国立医薬品食品衛生研究所　安全性生物試験研究センター）
井 上 智 彰　（日本免疫毒性学会）
今 井 教 安　（日本動物実験代替法学会）
岩瀬裕美子　（日本製薬工業協会）
篠 田 和 俊　（独立行政法人　医薬品医療機器総合機構）
杉山真理子　（日本化粧品工業連合会）
仲 井 俊 司　（日本化学工業協会）
中村るりこ　（独立行政法人　製品評価技術基盤機構）
野 口 真 希　（独立行政法人　医薬品医療機器総合機構）
西 川 秋 佳　（国立医薬品食品衛生研究所　安全性生物試験研究センター）
沼 澤　　聡　（日本毒性学会）
森 田　　健　（日本環境変異原学会）
横 関 博 雄　（日本皮膚アレルギー・接触皮膚炎学会）

任期：平成 28 年 4 月 1 日～平成 30 年 3 月 31 日

ヒト表皮モデルを用いた皮膚腐食性試験法は、ウサギを用いる皮膚腐食性試験の代替法として開発された試験法である。本試験法では、腐食性物質が角質層を傷害または角質層に吸収された後拡散し、表皮細胞に到達して細胞毒性を示すという仮説に基づき、被験物質曝露後の細胞生存率を指標に皮膚腐食性を評価する。

本試験法については、EpiSkin™、EpiDerm™、SkinEthic™、epiCS®という４種の表皮モデルがあげられている。いずれのモデルによる試験もバリデーション研究が実施され、ECVAM（European Centre for the Validation of Alternative Methods：欧州代替法評価センター）によりその信頼性と再現性が高いことが確認された[1~4]。EpiSkin™、EpiDerm™においては、ESAC（ECVAM Scientific Advisory Committee : ECVAM 科学諮問会議）での評価後、その結論は ICCVAM（Interagency Coordinating Committee on the Validation of Alternative Methods：米国代替法に関する省庁間連絡会議）においても確認された[5,6]。上記 4 種モデルについては OECD（Organisation for Economic Co-operation and Development：経済協力開発機構）にてテストガイドライン（TG）431 として承認された。この TG は昨今種々の点で改訂されており、現在は 2016 年版となっている[7]。

JaCVAM 評価会議は、皮膚腐食性試験資料編纂委員会により作成された「ヒト表皮モデルを用いた皮膚腐食性試験代替法の評価報告書」[8]を用いて、本試験法の妥当性について検討した。

１．試験法の定義

名称：　ヒト表皮モデルを用いた皮膚腐食性試験代替法

代替する対象毒性試験：　ウサギを用いる皮膚腐食性試験

試験法の概略：　本試験法では、ウサギ皮膚の代わりに角質層を有する 3 次元再構築ヒト表皮モデルを用い、被験物質が角質層下の表皮細胞層に対し細胞毒性を示す能力を評価する。表皮モデル表面に被験物質を一定時間適用した後洗浄し、表皮細胞の生存率を MTT〔3 - (4, 5 - Dimethylthiazol - 2 - yl) - 2, 5-diphenyltetrazolium bromide〕の還元量から求め、皮膚腐食性を判定する。

２．評価に用いた資料及び評価内容の科学的妥当性

本試験法は、4 種の表皮モデルのいずれのモデルも、ECVAM によるバリデーション研究とそれに続く ESAC による第三者評価により、実験動物を用いた皮膚腐食性試験の代替法として科学的に妥当であると報告されており[1~4]、EpiSkin™および EpiDerm™を用いた試験法については ICCVAM においても同様に確認された[6]。上記 4 種の表皮モデルを用いる方法については現在 OECD テストガイドラインとして承認されている[7]。JaCVAM 皮膚腐食性試験資料編纂委員会では、現在まで公開されている情報[1~7,9~11]を基にヒト表皮モデルを用いた皮膚腐食性試験代替法としての科学的妥当性について評価した。その結果、本試験法は、「腐食性物質が角質層の傷害または角質層への吸収を経て下層の表皮細胞に対し毒性を示す」という皮膚腐食性発現機序に基づき、細胞への毒性を指標にしたものであり、原理的にも妥当であると判断された。

３．本試験法の有用性と適用限界

いずれの表皮モデルも動物を使用しておらず、動物福祉面から代替法として妥当である。強い酸性（pH 2.0 以下）またはアルカリ性（pH 11.5 以上）物質は、強い局所傷害を引き起こす可能性があることから、皮膚腐食性と判断しても良いことになっている [9]。しかしながら、これは腐食性についての情報が他にない場合に行われるワーストケースとしての判断であり、例えば緩衝作用が小さい物質や混合物の場合は偽陽性の判断となる可能性も考えられる [10]。このため、このような物質に対しても動物を用いる必要のない本試験法で皮膚腐食性を評価することは有用である。

いずれのモデルも UN GHS (United Nations Globally Harmonized System of Classification and Labelling of Chemicals：国連 GHS) 分類における腐食性（区分 1）について、その有無の予測性は高く、OECD で集計した化学物質では、感度は 95～100%、特異度は 72～79%、正確度は 84～90%であり、偽陰性率は 0～5%と低いレベルであった（表 1）。各モデルによる国連 GHS 分類の細区分（1A～1C）の予測性については、80 物質を 2 または 3 回実験して得られた値を見る限り、EpiSkin™ のみが 1B/1C を 80%程度識別可能であった。その他のモデルは細区分の評価には利用できないと判断した。

MTT 還元物質に対しては、TG431 に対応方法が記載されており、これに準じることで評価可能である。ガス、エアロゾールについてはバリデーションが実施されておらず、適用可否は判断できない。それら以外の物質については、物理状態（液体・固体等）および水溶性の有無にかかわらず適用可能であり、混合物にも適用可能な場合がある。

表 1．化学物質の全セット*における予測性（腐食性 [国連 GHS 区分 1] の有無）[11]

	EpiSkin™	EpiDerm™	SkinEthic™	epiCS®
正確度	89.6%	87.9%	84.6%	84.3%
感度	98.5%	100%	94.6%	95.3%
特異度	79.3%	73.9%	73.0%	71.6%
偽陽性率	20.7%	26.1%	27.0%	28.4%
偽陰性率	1.5%	0%	5.4%	4.7%

*) OECD の集計による。

４．目的とする物質又は製品の毒性を評価する試験法としての、社会的受け入れ性及び行政上の利用の可能性

社会的受け入れ性：

本試験法は、通常の培養技術を習熟した施設であれば実施できる試験法であり、OECD TG431 に承認された表皮モデルは市販されている。また、生きた動物を用いないという点で、3Rs の精神に合致しており、社会的受け入れ性は高い。

行政上の利用性：

本試験法において陽性の結果が得られた場合、被験物質を腐食性物質（国連 GHS 分類における区分 1）と判定することは可能であるが、UN GHS 分類の細区分のためには EpiSkin™ のみ利用可能である。なお、本試験法の利用にあたっては、適用範囲を十分に配慮した上で使用されるべきである。

参考文献

1) ECVAM (1998) Statement on the scientific validity of the EpiSkin test (An in vitro test for skin corrosivity)
2) ECVAM (2000) Statement on the application of the EpiDerm™ human skin model for skin corrosivity testing
3) ECVAM (2006) Statement on the application of the SkinEthic™ human skin model for skin corrosivity testing
4) ECVAM (2009) ESAC Statement on the scientific validity of an in-vitro test for skin corrosivity testing
5) ICCVAM (1999) NIH Publication No.99-4495. Corrositex: An in vitro test method for assessing dermal corrosivity potential of chemicals.
6) ICCVAM (2002) NIH Publication No.02-4502. ICCVAM Evaluation of EPISKIN™, and EpiDerm™ (EPI-200) and rat skin transcutaneous electrical resistance (TER) assay: in vitro test method for assessing dermal corrosivity potential of chemicals.
7) OECD (2016) Guideline for the testing of chemicals. 431, in vitro Skin Corrosion: Human skin model test.
8) JaCVAM 皮膚腐食性資料編纂委員会：皮膚腐食性試験評価報告書　ヒト表皮モデルを用いた皮膚腐食性試験代替法の評価報告書（2017 年 2 月 24 日）.
9) OECD (2014) Series on Testing and Assessment No. 203, Guidance document on an Integrated Approach on Testing and Assessment (IATA) for skin corrosion and irritation
10) Sheel J. et.al., Classification and labeling of industrial products with extreme pH by making use of in vitro methods for the assessment of skin and eye irritation and corrosion in a weight of evidence approach, Toxicology in Vitro, 25, 1435-1447 (2011)
11) OECD (2015) Series on Testing & Assessment No. 219, Performance Standards for the assessment of proposed similar or modified in vitro reconstructed human epidermis (RhE) test methods for skin corrosion testing as described in TG 431

評価報告書

皮膚腐食性試験代替法

ヒト表皮モデル法

皮膚腐食性試験資料編纂委員会

平成 29 年（2017 年）2 月 24 日

皮膚腐食性試験資料編纂委員会

髙橋　祐次　　国立医薬品食品衛生研究所 毒性部

中村るりこ　　独立行政法人 製品評価技術基盤機構　化学物質管理センター

須方　督夫　　住友化学株式会社 レスポンシブルケア部（化学品安全グループ）
／日本化学工業協会

小島　肇　　国立医薬品食品衛生研究所 安全性予測評価部

要旨

ウサギを用いる皮膚腐食性試験の動物実験代替法（代替法）として経済協力開発機構(OECD: Organisation for Economic Co-operation and Development）で試験ガイドライン(TG: Test Guideline) 431 として承認されたヒト表皮モデルを用いる試験法の有用性を評価した。信頼性と妥当性という視点において、ヒト表皮モデルを用いた試験を評価した結果、TG431 に掲載されているすべてのモデル EpiSkin™、EpiDerm™、SkinEthic™、epiCS®が腐食性の有無を評価できるモデルとして推奨できると考えられた。ただし、国連化学品の分類および表示に関する世界調和システム(UN GHS: United Nations Globally Harmonized System of Classification and Labelling of Chemicals)分類の細区分を考慮する場合は EpiSkin™が皮膚腐食性試験の代替法としてもっとも有用であると結論した。

1. 試験法の科学的および規制面からの妥当性

皮膚腐食性試験は皮膚刺激性試験の一環として行われ、種々のガイドラインではDraizeらにより提唱されたウサギを用いる方法が推奨されてきた[1)]。この方法は被験物質の刺激性や腐食性を検出する試験として長く使用されてきたものの、判定を肉眼で行うため客観性に乏しく実験間や施設間での再現性が乏しい。更に動物に激しい痛みとストレスを与えることが社会的に問題となり、以前より動物を使用しない動物実験代替法（以下、代替法と記す）の開発が切望されていた。

この代替法として、経済協力開発機構(OECD: Organisation for Economic Co-operation and Development）で試験ガイドライン(TG: Test Guideline)431には、皮膚腐食性試験として角質層を有し3次元的に再構築されたヒト表皮モデルを使用した評価方法が記載されている[2)]。この試験法は、腐食性物質が角質層の傷害または角質層に吸収された後拡散することにより、下層の細胞に到達して細胞毒性を示すという仮説に基づき、被験物質曝露後の細胞生存率を指標に皮膚腐食性を評価している。EpiSkin™やEpiDerm™等のヒト表皮モデルは欧米では既にバリデーション研究が実施され、欧州では化学物質の皮膚腐食性評価を目的として承認され、化学物質のリスク表示識別等に利用されている。特に昨今では国連化学品の分類および表示に関する世界調和システム(UN GHS: United Nations Globally Harmonized System of Classification and Labelling of Chemicals)分類に従って評価されるケースが増えている。このTG431は昨今、毎年種々の点で改訂されており、現在は2016年版となっている[3)]。

我が国で既存の化学物質を評価する場合、OECDで承認された試験方法による結果は、一般的に行政的に受け入れられるが、現在まで代替法での結果をもとに行政的に評価された例は多くない。安全性評価における代替法の普及が切望されている現状において、我が国でも科学的に妥当なものは積極的に受け入れることが必要となっている。なお、国内企業からも国際的な評価に耐えうるヒト表皮モデルが開発されている。

本評価書では、OECD TG431に掲載されたヒト表皮モデルEpiSkin™、EpiDerm™、SkinEthic™およびepiCS®を用いる腐食性試験法の有用性を評価した[3)]。

2. 試験プロトコル構成の妥当性[3)]

被験物質が角質層を通過して表皮細胞に曝露され、MTT〔3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide〕の還元量から求めた細胞生存率の割合から皮膚腐食性を判定する。それらの概要を表1にまとめた。実験操作上の違いは、前培養法および染色液の用量、組織からの抽出法等であり、基本的な曝露時間と判定基準については表2を参照されたい。

試験プロトコルとして、EpiDerm™を例に説明する。6 wellプレートの各wellに培養液1 mLを加えた後にヒト表皮モデルを置き、被験物質が液状の場合はピペッターで50 µL、粉末など固形の場合は25 mgを25 µLの水と合わせ、モデル上層に適用する(n=2)。被験物質を3分または60分処理後、プレートから被験物質をデカンテーションにより除去した後、10 mLのPBSで2〜3回軽く洗浄する。洗浄後、ペーパータオル等で水分を切り、破損に注意しながら別の24 wellプレートにヒト表皮モデルを移動する。MTT色素を含む培養液をヒト表皮モデルの下方に0.3 mL添加する。37℃、CO_2インキュベータ中に3時間静置した後、以下の手順にて抽出を行う。イソプロパノールを2 mL添加し、一晩室温放置後、96 wellプレートに抽出液を200 µLずつ移し（1物

質あたり 2〜3 well）、マイクロプレートリーダーを用いて 540 nm あるいは 570 nm 領域での吸光度を測定する。イソプロパノールのみを加えた well をブランクとし、実測値とブランク値の差を求める。溶媒対照の吸光度を 100%とし各検体の 3 または 60 分間処理時の吸光度を%として算出し細胞の生存率とする。3 分間処理したときの生存率が 50%未満、あるいは 3 分間では生存率が 50%以上であるが 60 分間処理したときに 15%未満の結果を示す物質を“腐食性”と判定する。一方、3 分間処理したときに生存率が 50%以上、60 分間処理したときに 15%以上の物質は“非腐食性”と判定する。試験は 1 回とし、well 間で異なる結果が得られた場合のように明確な評価ができなかった場合は追加試験を実施し、最終評価とする。

表１．皮膚腐食性試験のために確認された RhE 試験方法からなる主な試験方法 [3)]

±で示した数字は、ばらつきの許容範囲を示す。

テスト方法 要素	EpiSkin™	EpiDerm™SCT	SkinEthic™ RHE	epiCS®
モデル表面積	0.38cm^2	0.63 cm^2	0.5 cm^2	0.6 cm^2
組織数	曝露時間毎に少なくとも 2	曝露時間毎に 2-3	曝露時間毎に少なくとも 2	曝露時間毎に少なくとも 2
使用量と適用	液体および粘性物：50±3 µL (131.6µL/cm^2) 固体：20±2 mg (52.6 mg/cm^2)+100 ±5 µL の NaCl 溶液(9 g/L) ロウ様/粘着性物質：ナイロンメッシュを用いて 50±2 mg (131.6 mg/cm^2)	液体：ナイロンメッシュを用いた状態で、もしくは用いない場合でも 50 µL (79.4 µL/cm^2) 被験物質とナイロンメッシュとの親和性は予試験で確認する。 半固体：50 µL (79.4 µL/cm^2) 固体： 25 mg (39.7 mg/cm^2)＋25 µL の水(必要であればそれ以上) ロウ様物：15 µL の水で湿らせた直径約 8 mm のフラットなディスク様の片を上に乗せる。	液体および粘性物：ナイロンメッシュを使用して 40±3 µL (80 µL/cm^2) 被験物質とナイロンメッシュとの親和性は予試験で確認する。 固体： 20±3 mg (40 mg/cm^2)＋20±2 µL の水 ロウ様/粘着性物質：ナイロンメッシュを用いた状態で 20±3 mg (40 mg/cm^2)	液体：ナイロンメッシュを用いた状態で 50 µL (83.3 µL/cm^2) 被験物質とナイロンメッシュとの親和性は予試験で確認する。 半固体：50 µL (83.3 µL/cm^2) 固体：25 mg (41.7 mg/cm^2)+ 25 µL の水(必要であればそれ以上) ロウ様物：15 µL の水で湿らせたフラットな、直径約 8 mm のクッキー様の片を上に乗せる。
直接 MTT 還元性の事前確認	50 µL(液体)もしくは 20 mg(固体)に 0.3 mg/mL MTT 溶液 2 mL を加えて、37℃、CO_2 濃度 5%、湿度 95%で 180±5 分培養 →溶液の色が青/紫に変わった場合、水処理でモデル構成細胞を死滅させたものに被験物質を処置する対照もとる。	50 µL(液体)もしくは 25 mg(固体)に 1 mg/mL MTT 溶液 1 mL を加えて、37℃、CO_2 濃度 5%、湿度 95%で 60 分培養 →溶液の色が青/紫に変わった場合、凍結処理でモデル構成細胞を死滅させたものに被験物質を処置する対照もとる。	40 µL(液体)もしくは 20 mg(固体)に 1 mg/mL MTT 溶液 1 mL を加えて、37℃、CO_2 濃度 5%、湿度 95%で 180±15 分培養 →溶液の色が青/紫に変わった場合、凍結処理でモデル構成細胞を死滅させたものに被験物質を処置する対照もとる。	50 µL(液体)もしくは 25 mg(固体)に 1 mg/mL MTT 溶液 1 mL を加えて 37℃、CO_2 濃度 5%、湿度 95%で 60 分培養 →溶液の色が青/紫に変わった場合、凍結処理でモデル構成細胞を死滅させたものに被験物質を処置する対照もとる。

テスト方法 要素	EpiSkin™	EpiDerm™SCT	SkinEthic™ RHE	epiCS®
着色障害の事前確認	10 µL(液体)もしくは 10 mg(固体)に 90 µL の水を加えて室温で 15 分攪拌する。 →溶液が着色した場合、MTT のみを加えない対照をとる。	50 µL(液体)もしくは 25 mg(固体)に 300 µL の水を加えて 37℃、CO_2 濃度 5%、湿度 95%で 60 分攪拌する。 →溶液が着色した場合、MTT のみを加えない対照をとる。	40 µL(液体)もしくは 20 mg(固体)に 300 µL の水を加えて室温で 60 分攪拌する。 →被験物質の色が着色している場合、MTT のみを加えない対照をとる。	50 µL(液体)もしくは 25 mg(固体)に 300 µL の水を加えて 37℃、CO_2 濃度 5%、湿度 95%で 60 分攪拌する。 →溶液の色が着色している場合、MTT のみを加えない対照をとる。
曝露時間と温度	室温（18-28℃）で 3 分、60 分（±5 分）および 240 分（±10 分）換気されたキャビネット内	室温で 3 分、および 37℃、CO_2 濃度 5%、湿度 95%で 60 分	室温で 3 分、および 37℃、CO_2 濃度 5%、湿度 95%で 60 分	室温で 3 分、および 37℃、CO_2 濃度 5%、湿度 95%で 60 分
PBS によるすすぎ	PBS 25 mL(すすぎ一回毎に 2 mL)	PBS を一定した弱流で 20 回	PBS を一定した弱流で 20 回	PBS を一定した弱流で 20 回
陰性対照	50 µL の塩化ナトリウム溶液（9g/L) 曝露時間毎に	50 µL の水 曝露時間毎に	40 µL の水 曝露時間毎に	50 µL の水 曝露時間毎に
陽性対照	50 µL の氷酢酸で 4 時間曝露時のみ	8 N 水酸化カリウム 50 µL で曝露時間毎に	8 N 水酸化カリウム 40 µL で 1 時間曝露時のみ	8 N 水酸化カリウム 5 0µL で曝露時間毎に
MTT 溶液	濃度 0.3 mg/mL 2 mL	濃度 1 mg/mL 300 µL	濃度 1 mg/mL 300 µL	濃度 1 mg/mL 300 µL
MTT 溶液での培養時間および温度	37℃、CO_2 濃度 5%、湿度 95%で 180 分(±15 分)	37℃、CO_2 濃度 5%、湿度 95%で 180 分	37℃、CO_2 濃度 5%、湿度 95%で 180 分(±15 分)	37℃、CO_2 濃度 5%、湿度 95%で 180 分
抽出溶媒	500 µL の酸性化イソプロパノール (0.04 N 塩酸を含むイソプロパノール) (分離した組織を十分に浸漬させる)	2 mL のイソプロパノール (インサート全体から抽出する)	1.5 mL のイソプロパノール (インサート全体から抽出する)	2 mL のイソプロパノール (インサート全体から抽出する)

テスト方法 要素	EpiSkin™	EpiDerm™SCT	SkinEthic™ RHE	epiCS®
抽出時間および温度	遮光し、室温で一晩	室温で振とうせずに一晩、もしくは室温で振とうした状態で（～120rpm）120 分	室温で振とうせずに一晩、もしくは室温で振とうした状態で（～120rpm）120 分	室温で振とうせずに一晩、もしくは室温で振とうした状態で（～120rpm）120 分
OD 測定条件	参照フィルターなしで 570nm (545-595nm)	参照フィルターなしで 570nm (もしくは 540nm)	参照フィルターなしで 570nm (545-600nm)	参照フィルターなしで 540-570nm
組織の品質確認	SDS で 18 時間処理 1.0 mg/mL≦IC_{50}≦3.0 mg/mL	1%Triton X-100 で処理 4.08 時間≦ET_{50}≦8.7 時間	1%Triton X-100 で処理 4.0 時間≦ET_{50}≦10.0 時間	1%Triton X-100 で処理 2.0 時間≦ET_{50}≦7.0 時間
試験回数	1 回、明確な結果が得られなかった時は 2 回	1 回、明確な結果が得られなかった時は 2 回	1 回、明確な結果が得られなかった時は 2 回	1 回、明確な結果が得られなかった時は 2 回
適合判定基準	1. 陰性対照(塩化ナトリウム溶液)で処理された組織の OD 値の平均は、曝露時間毎に 0.6 以上 1.5 以下。 2. 陽性対照(氷酢酸)で 4 時間処置された組織の生存率の平均は 20%以下。 3. 生存率の幅が 20-100%および OD 値が 0.3 以上の場合、2 つの組織間の生存率の差は 30%をこえない。	1. 陰性対照(水)で処置された組織の OD 値の平均は、全ての曝露時間において 0.8 以上 2.8 以下。 2. 陽性対照（8N 水酸化カリウム）に 1 時間曝露した組織複製物の測定値の生存率は 15%未満。 3. 生存率が 20-100%の場合において、組織間の変動係数(CV)は 30%以下。	1. 陰性対照(水)で処置された組織の OD 値の平均は、曝露時間毎に 0.8 以上 3.0 以下。 2. 1 時間(もし可能であれば 4 時間)陽性対照(8N 水酸化カリウム)で処置された組織複製物の測定値の生存率は 15%未満。 3. 20-100%の生存率および 0.3 以上の OD 値において、2 つの組織間の生存率の差は 30%を超えない。	1. 陰性対照(水)で処置した組織の OD 値の平均は、各曝露時間毎に 0.8 以上 2.8 以下。 2. 1 時間(もし可能であれば 4 時間)陽性対照(8N 水酸化カリウム)で処置された組織複製物の測定値の生存率は 15%未満。 3. 20-100%の生存率および 0.3 以上の OD 値において、2 つの組織間の生存率の差は 30%を超えない。

表 2-1　EpiSkin™ の予測モデル [3)]

3 分、60 分および 240 分曝露後の生存率	予測性の評価
3 分曝露後の生存率が 35%未満	腐食性 ・国連 GHS 細区分　1A*
3 分曝露後の生存率が 35%以上で、かつ 60 分曝露後の生存率が 35%未満の場合、 もしくは 60 分曝露後の生存率が 35%以上で、かつ 240 分曝露後の生存率が 35%未満	腐食性 ・国連 GHS 細区分 1B あるいは 1C
240 分曝露後の生存率が 35%以上	非腐食性

*) 腐食性の細区分における RhE 試験法の有用性を評価するために作成したデータによると、EpiSkin™ 試験法により区分 1A に分類された物質／混合物の約 22%が、実際には区分 1B または 1C に属するものである可能性がある （すなわち、過大評価）。

表 2-2 EpiDerm™SCT、SkinEthic™RHE および epiCS®の予測モデル [3)]

3 分および 60 分曝露後の生存率	予測性の評価
3 分曝露後の生存率が 50%未満	腐食性 ・国連 GHS 細区分　1A*
3 分曝露後の生存率が 50%以上で、かつ 60 分曝露後の生存率が 15%未満	腐食性 ・国連 GHS 細区分 1B あるいは 1C
3 分曝露後の生存率が 50%以上で、かつ 60 分曝露後の生存率が 15%以上	非腐食性

*) 腐食性の細区分における RhE 試験法の有用性を評価するために作成したデータによると、EpiSkin™ 試験法により区分 1A に分類された物質／混合物の約 22%が、実際には区分 1B または 1C に属するものである可能性がある （すなわち、過大評価）。

3. 開発および評価に使われた物質の分類、選択理由の妥当性、*in vitro* および参照データの有無

EpiSkin™ の再現性は、60 の被験物質により調べられている。それらの物質が掲載された論文の抜粋を ANNEX 1 に示す [4)]。EpiDerm™の再現性は、24 の被験物質により調べられている。それらの物質が掲載された論文の抜粋を ANNEX 2 に示す [5)]。SkinEthic™の再現性は、12 の被験物質により調べられている。それらの物質が掲載された論文の抜粋を ANNEX 3 に示す [6)]。EpiCS®の再現性は、12 の被験物質により調べられている。それらの物質の抜粋を ANNEX 4 に示す [2)]。さらに、上記 4 つのモデルの予測性が 80 被験物質で調べられている。それらの物質が掲載された論文の抜粋を ANNEX 5 に示す [7)]。評価に使用された被験物質の多くは欧州代替法評価センター（ECVAM：European Centre for the Validation of Alternative Methods）主導の皮膚腐食性試験バリデーションで使用された物質である。これら被験物質の評価結果は下記の資料として提出されている。

Liebsh et al., ATLA 2000[5)], Barratt et al., Toxicol. In Vitro 1998[8)], Fentem et al, Toxicol. In Vitro 1998[4)], Worth et al., ATLA 1998[9)], Botham et al., ATLA 1995[10)] , ICCVAM(1999) NIH Publication No:99-4495[11)], ICCVAM (2002) NIH Publication No: 02-4502[12)]

4. 試験法の正確性（再現性）

EpiSkin™においては、60 物質を用いた 3 施設での 2-way の ANOVA 解析を用い、施設内および施設間の変動については、それらの間に有意な差はないと判断された(Fentem et al 1998)[4]。60 物質（27 腐食性物質および 33 非腐食性物質）のうち、42 物質は 3 施設とも施設内および施設間再現性が良好であった。残る 18 物質では何等かの結果が異なっていたが、ECVAM は本試験法の信頼性と再現性は高いと判断した[13]。この結論は、ECVAM 科学諮問会議（ESAC: ECVAM Scientific Advisory Committee)[13] および米国の代替法に関する省庁間連絡会議（ICCVAM：Interagency Coordinating Committee on the Validation of Alternative Methods)[12] での評価においても確認された。

EpiDerm™においては、24 物質を用いた 3 施設での 2 回の実験において、21 物質の腐食性を 3 施設すべてで正しく予測できた（Liebsch et al 2000）[5]。ECVAM は本試験法の信頼性と再現性は高いと判断した。この結論は、ESAC の評価後[14]、ICCVAM においても確認された[11,12]。

SkinEthic™においては、12 物質を用いた 3 施設での 3 回の実験において、93.2%の結果が一致し（Kandarova et al 2006）[6]、ESAC において施設内および施設間再現性が高いと判断された[15]。

EpiCS®においては、TG431（2004）の参照 12 物質を用いた 4 施設での 3 回の実験において[2]、テトラクロロエチレンを除く 11 物質の結果が一致し、ESAC において施設内および施設間再現性が認められた[16]。

5. 試験法の信頼性[7、17、18]

いずれのヒト表皮モデルも腐食性の有無の予測性は高く、正確度[1]は 84～90%、感度[2]は 95～100%、特異度[3]は 72～79%であり、偽陰性率は 0～5%と低いレベルであった（表 3)。これらの値から、OECD の性能標準では正確度、感度および特異度をバリデーションに必要な基準として 82.5、95 および 70%と定めている。

さらに、各モデルの細区分の予測性を表 4 まとめた。80 物質を 2 または 3 回実験して得られた値を見る限り、細区分である 1 B/1C を 80%程度で評価できているモデルは EpiSkin™ のみであった。その他のモデルは国連 GHS 細区分の評価には利用できないと判断した。

これらモデルは MTT 還元物質への対応方法も TG431 に記載されている。適用できない物質としてはガス、エアロゾールのみが記載されている（バリデーション未実施)。これらモデルは物理状態（液体・固体等）および水溶性の有無にかかわらず適用可能であり、ガス、エアロゾールを除く混合物でも適用可能とされている。なお、特定の種類の物質や混合物においてこれらモデルの適用性を否定するような明確な根拠が得られた場合には、適用範囲から除外するべきであるとされている。

---

1 正確度：被験物質について試験法で得られる結果が、被験物質の既定の参照値と類似している程度。
2 感度：試験された陽性物質の中で、試験法で正しく陽性と判定されたものの割合。
3 特異度：試験された陰性物質の中で、試験法で正しく陰性と判定されたものの割合。

表 3. OECD で集計した全セットの化学物質を用いた予測性の計算結果（腐食性 [国連 GHS 区分１] の有無）[17)]

	EpiSkin™	EpiDerm™	SkinEthic™	epiCS®
正確度	89.6%	87.9%	84.6%	84.3%
感度	98.5%	100%	94.6%	95.3%
特異度	79.3%	73.9%	73.0%	71.6%
偽陽性率	20.7%	26.1%	27.0%	28.4%
偽陰性率	1.5%	0%	5.4%	4.7%

表 4.　OECD で集計した全セットの化学物質を用いた予測性の計算結果（細区分：国連 GHS 区分１A、１B/1C、非腐食性）[3)]

化学品全体の統計データ（%） （80 種類の化学品を 2 または 3 回試験、すなわち、159*または 24 回の分類結果） *1 品目は、入手不可能であったため 1 回のみ試験				
	EpiSkin™	EpiDerm™	SkinEthic™	EpiCS®
過大評価				
区分 1BC の化学品の１A への過大評価	21.5	29.0	31.2	32.8
区分 NC の化学品の１B/１C への過大評価	20.7	23.4	27.0	28.4
区分 NC の化学品の１A への過大評価	0.0	2.7	0.0	0.0
過大評価された区分 NC 化学品	20.7	26.1	27.0	28.4
全区分での過大評価率	17.9	23.3	24.5	25.8
過小評価				
区分１A の１B/1C への過小評価	16.7	16.7	16.7	12.5
区分１A の NC への過小評価	0.0	0.0	0.0	0.0
区分１B/1C の NC への過小評価	2.2	0.0	7.5	6.6
全区分での過小評価率	3.3	2.5	5.4	4.4
正確な分類				
正しく分類された１A 化学品	83.3	83.3	83.3	87.5
正しく分類された１B/1C 化学品	76.3	71.0	61.3	60.7
正しく分類された NC 化学品	79.3	73.9	73.0	71.6
一致度(予測能)	78.8	74.2	70.0	69.8

NC:非腐食性

6. 他の科学的な報告との比較の有無

OECD の腐食性試験代替法ガイドラインとして、TG431 の他に「TG430 TER（Transcutaneous Electrical Resistance Test Method：経皮電気抵抗性試験）」[19)]および「TG435　（*In Vitro* Membrane Barrier Test Method for Skin Corrosion）: *in vitro* 膜バリア試験」[20)] が承認されている。これらはいずれも ECVAM にてバリ

デーション研究が実施され、ICCVAM はこれらの試験法（Rat Skin TER, EpiSkin™, EpiDerm™および Corrositex®）の正確度、感度および特異度について比較している（表 5）[11,12]。

これらの比較においては同じ物質を用いて評価されておらず、試験物質の数量や選択物質の種類が異なっているため結果の数値だけをもって、単純にヒト表皮モデルの優越性を比較評価することは困難であるが、いずれの試験法も同等の予測性を有すると思われる。

表 5. 試験法の比較結果 [11,12]

	TER	EpiSkin	EpiDerm	Corrositex
物質数	122	60	24	163
正確度	81% (99/122)	83% (50/60)	92% (22/24)	79% (128/163)
感度	94%(51/54)	82%(23/28)	92%(11/12)	85%(76/89)
特異度	71%(48/68)	84%(27/32)	83%(10/12)	72%(52/74)

7. 3Rs 原則との関係（動物福祉面からの妥当性）

いずれの表皮モデルも動物を使用しておらず、動物福祉面から代替法として妥当である。

8. 試験法の有用性と限界（コスト、時間からの妥当性など）

強い酸性（pH 2.0 以下）またはアルカリ性（pH 11.5 以上）物質は、強い局所作用を有する可能性が高いことから、皮膚腐食性と判断しても良いことになっている [21]。しかしながら、これは腐食性についての情報が他にない場合に行われるワーストケースとしての判断であり、例えば酸や塩基の添加により pH が変わり易い物質や混合物の場合は偽陽性の判断となる可能性も考えられる [22]。このため、このような物質に対して動物を用いる必要のない TG431 で皮膚腐食性を評価することは有用である。

表 6 に示す習熟度確認物質を正しく分類できるか否か試験することにより、専門技術の習熟について確認することができるとの記載が TG431 にある。

9. その他

OECD に承認されたヒト表皮モデルは海外で開発された製品であり、コスト面でやや割高である。日本製のヒト表皮モデルとして、これまでに LabCyte EPI-Model[24] や Vitrolife-Skin が報告されている [25]。特に、Vitrolife-Skin は厚生労働科学研究補助金事業でバリデートされ [26]、JaCVAM においても評価がなされている [27]。ただし、OECD の性能標準に準じたバリデーションは実施されておらず、OECD に新たな試験法として推奨できるモデルではない。よって、これらのモデルの評価は本評価書では行わなかった。

特許についての情報は今回の検討資料に示されていない。なお、上記したヒト表皮モデルは既に市販されており、いずれも購入可能である。

10. 結論

信頼性と妥当性という視点において、ヒト表皮モデルを用いた皮膚腐食性試験を評価した結果、被験物質の皮膚腐食性を評価する試験法として EpiSkin™、EpiDerm™、SkinEthic™および epiCS®が推奨できるモデルとして挙げられた。ただし、国連 GHS 細区分を考慮する場合は、EpiSkin™が皮膚腐食性試験の代替法としてもっとも有用であると結論した。

表 6. 習熟度確認物質 [3)]

化学物質 [1]	CASRN	化学物質分類 [2]	UN GHS（in vivo 試験）による区分 [3]	VRM（in vitro 試験）による区分 [4]	MTT 還元剤 [5]	物理的状態
区分 1A の *in vivo* 腐食性物質						
ブロモ酢酸	79-08-3	有機酸	1A	(3) 1A	--	固体
3 フッ化ボロン二水和物	13319-75-0	無機酸	1A	(3) 1A	--	液体
フェノール	108-95-2	フェノール類	1A	(3) 1A	--	固体
ジクロロアセチルクロリド	79-36-7	求電子剤	1A	(3) 1A	--	液体
区分 1B/1C の *in vivo* 腐食性物						
グリオキシル酸一水和物	563-96-2	有機酸	1B/1C	(3) 1B/1C	--	固体
乳酸	598-82-3	有機酸	1B/1C	(3) 1B/1C	--	液体
エタノールアミン	141-43-5	有機塩基	1B	(3) 1B/1C	□	粘稠性
塩酸（14.4%）	7647-01-0	無機酸	1B/1C	(3) 1B/1C	--	液体
***in vivo* 非腐食性物質**						
臭化フェネチル	103-63-9	求電子剤	NC	(3) NC	□	液体
4-アミノ-1,2,4-トリアゾール	584-13-4	有機塩基	NC	(3) NC	--	固体
4-（メチルチオ）ベンズアルデヒド	3446-89-7	求電子剤	NC	(3) NC	□	液体
ラウリン酸	143-07-7	有機酸	NC	(3) NC	--	固体

略語：CASRN＝CAS 登録番号；UN GHS＝国連勧告「化学品の分類および表示に関する世界調和システム」(1)；VRM＝バリデーション済み標準試験法；NC＝非腐食性

[1] これらの化学物質は、まず腐食性であるか非腐食性であるかにより分類し、さらに腐食性の細区分および化学物質の種類により細分類した。表に記載した化学物質は、ECVAM による EpiSkin™ および EpiDerm™ のバリデーション試験で使用された物質 [4,5,8)]、ならびに EpiSkin™ [23)]、EpiDerm™、SkinEthic™ および epiCS® [7)]の開発者により提供されたデータに基づいたバリデーション後試験より選択したものである。特に記載のない限り、市販されている化学物質の購入時の純度において試験を行った [5,8)]。この選択にあたっては、可能な限り以下のような物質を含めることとした：(i) VRM により測定または予測可能な腐食性反応の範囲（例えば、非腐食性、弱腐食性ないし強腐食性)を代表する化学物質、(ii)バリデーション試験で用いられた化学物質分類を代表する物質、
(iii)化学構造が明確に同定されている物質、(iv)VRM において再現性の高い結果が得られる物質、(v)*in vivo* での標準試験法において確実な結果が得られる物質、(vi)市販されている物質、ならびに(vii) 非常に高額な廃棄コストがかからない物質。

[2]Barratt ら(1998)による化学物質分類 [8)]。

[3]UN GHS の細区分 1A、1B および 1C には、国連包装等級 I、II および III がそれぞれ対応する。

[4] 表に記載した VRM による *in vitro* での分類予測は、試験法の開発者が行ったバリデーション後試験において EpiSkin™ および EpiDerm™ 試験法（VRM）により得られたものである。

[5]ECVAM による皮膚腐食性試験で得られた生存率の値は、直接的な MTT 還元を考慮した補正を行っていない。バリデーション後試験のデータは、適切な対照を用いて得たものである。

11．文献

1) OECD (2015) Guideline for the testing of chemicals. 404, Acute Dermal Irritation/Corrosion
2) OECD (2004) Guideline for the testing of chemicals. 431, in vitro Skin Corrosion: Human skin model test.
3) OECD (2016) Guideline for the testing of chemicals. 431, in vitro Skin Corrosion: Human skin model test.
4) Fentem, J.H. et al. (1998) The ECVAM International Validation Study on in vitro tests for skin corrosivity. 2. Results and evaluation by the management team. Toxicology in Vitro, 12, 483-524.
5) Liebsch, M., et al. (2000) The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. ATLA, 28, 371-401.
6) Kandarova, H. et al. (2006) Assessment of the human epidermis model SkinEthic RHE for in vitro skin corrosion testing of chemicals according to new OECD TG 431, Toxicology in Vitro, 20, 547-559.
7) Desprez, B. et al. (2015) Two novel prediction models improve prediction of skin corrosive sub-category by test methods of OECD Test Guideline No. 431, Toxicol. In Vitro, 29,2055-2080.
8) Barratt, M.D. et al. (1998) The ECVAM International Validation Study on in vitro tests for skin corrosivity. 1. Selection and Distribution of the test chemicals. Toxicology in Vitro, 12, 471-482.
9) Worth, A.P. et al. (1998) An evaluation of the proposed OECD testing strategy for skin corrosion. ATLA, 26, 709-720.
10) Botham, P.A. et al. (1995) A pre-validation study on in vitro skin corrosivity testing: The report and recommendations of ECVAM workshop 6. ATLA, 23, 219-255.
11) ICCVAM (1999) NIH Publication No.99-4495. Corrositex: An in vitro test method for assessing dermal corrosivity potential of chemicals.
12) ICCVAM (2002) NIH Publication No.02-4502. ICCVAM Evaluation of EPISKIN, and EpiDerm (EPI-200) and rat skin transcutaneous electrical resistance (TER) assay: in vitro test method for assessing dermal corrosivity potential of chemicals.
13) ECVAM (1998) Statement on the scientific validity of the Episkin test (An in vitro test for skin corrosivity)
14) ECVAM (2000) Statement on the application of the EpiDerm™ human skin model for skin corrosivity testing
15) ECVAM (2006) Statement on the application of the SkinEthic™ human skin model for skin corrosivity testing
16) ECVAM (2009) ESAC Statement on the scientific validity of an in-vitro test for skin corrosivity testing
17) OECD (2015) Series on Testing & Assessment No. 219, Performance Standards for the assessment of proposed similar or modified in vitro reconstructed human epidermis (RhE) test methods for skin corrosion testing as described in TG 431
18) OECD (2013) Series on Testing & Assessment No. 190, Summary document on the Statistical Performance of Methods in OECD Test Guideline 431 for Sub-Categorisation
19) OECD (2015) Guideline for the testing of chemicals. 430, in vitro Skin Corrosion: Transcutaneous electrical resistance test (TER).

20) OECD (2015) Guideline for the testing of chemicals. 435, in vitro membrane barrier test method for Skin Corrosion.
21) OECD (2014) Series on Testing and Assessment No. 203, Guidance document on an Integrated Approach on Testing and Assessment (IATA) for skin corrosion and irritation
22) Sheel J. et.al. (2011) Classification and labeling of industrial products with extreme pH by making use of in vitro methods for the assessment of skin and eye irritation and corrosion in a weight of evidence approach, Toxicology in Vitro, 25, 1435-1447
23) Alépée N., Grandidier M.H., and Cotovio J. (2014). Sub-Categorisation of Skin Corrosive Chemicals by the EpiSkinTM Reconstructed Human Epidermis Skin Corrosion Test Method according to UNGHS: Revision of OECD Test Guideline 431. *Toxicol. In Vitro* 28, 131-145.
24) Katoh, M., et al. (2010) Assessment of the human epidermal model LabCyte EPI-MODEL for In vitro skin corrosion testing according to the OECD test guideline 431. J Toxicol Sci., 35(3):411-7.
25) Morikawa, N., et al. (2005) Experimental study on a novel chemical application procedure for in vitro skin corrosivity testing using the Vitrolife-Skin ™ human skin model, AATEX, 11, 1, 68-78.
26) Kojima H et al. (2008) Validation of human skin models for skin corrosivity tests in Japan, AATEX 13(1), 36-44.
27) JaCVAM (2008) ヒト皮膚モデルを用いた皮膚腐食性試験代替法の第三者評価報告書（Vitrolif-Skin™）

ANNEX

1) Fentem J.H. et al. (1998), The ECVAM International Validation Study on in vitro tests for skin corrosivity. 2. Results and evaluation by the management team. Toxicology in Vitro, 12, 483-524 抜粋

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

3) Kandarova, H. et al. (2006), Assessment of the human epidermis model SkinEthic RHE for in vitro skin corrosion testing of chemicals according to new OECD TG 431, Toxicology in Vitro, 20, 547-559抜粋

4) OECD (2004) Guideline for the testing of chemicals. 431, in vitro Skin Corrosion: Human skin model test. 抜粋

5) OECD (2013) GD190 Summary document on the Statistical Performance of Methods in OECD Test Guideline 431 for Sub-Categorisation 抜粋

6) Desprez, B. et al. (2015), Two novel prediction models improve prediction of skin corrosive sub-category by test methods of OECD Test Guideline No. 431, Toxicol. In Vitro, 29,2055-2080 抜粋

ANNEX 1 ：Fentem et al (1998)[4]のバリデーションで使われた 60 物質

Table 4. Test chemicals

No.	Chemical	C/NC	EU risk phrase	UN packing group	PII*
Organic acids					
1	Hexanoic acid	C	R34	II/III	—
29	65/35 Octanoic/decanoic (capric) acids	C	R34	II/III	NPC†
36	2-Methylbutyric acid	C	R34	II/III	> 4
40	Octanoic (caprylic) acid	C	R34	II/III	4.44
47	60/40 Octanoic/decanoic acids	C	R34	II/III	NPC
50	55/45 Octanoic/decanoic acids	C	R34	II/III	5.11
7	3,3′-Dithiodipropionic acid	NC			0
12	Dodecanoic (lauric) acid	NC			0.44
26	Isostearic acid	NC			4.33
34	70/30 Oleine/octanoic acid	NC			3.78
58	10-Undecenoic acid	NC			2.42
Organic bases					
2	1,2-Diaminopropane	C	R35	I	—
15	Dimethyldipropylenetriamine	C	R35	I	NPC
38	Tallow amine	C	R35	II	NPC
55	1-(2-Aminoethyl)piperazine	C	R34	II	—
13	3-Methoxypropylamine	C	R34	II/III	6.67
17	Dimethylisopropylamine	C	R34	II/III	5.61
45	*n*-Heptylamine	C	R34	II/III	6.67
10	2,4-Xylidine (2,4-dimethylaniline)	NC			1.44
35	Hydrogenated tallow amine	NC			3.56
59	4-Amino-1,2,4-triazole	NC			0
Neutral organics					
8	Isopropanol	NC			0.78
11	2-Phenylethanol (phenylethylalcohol)	NC			0.92/2.22
16	Methyl trimethylacetate	NC			0
19	Tetrachloroethylene	NC			5.67
22	*n*-Butyl propionate	NC			1.08
27	Methyl palmitate	NC			4.56
44	Benzyl acetone	NC			1.21
51	Methyl laurate	NC			3.89
56	1,9-Decadiene	NC			3.0
Phenols					
3	Carvacrol	C	R34	II/III	> 4
23	2-*tert*-Butylphenol	C	R34	II/III	5.67
9	*o*-Methoxyphenol (Guaiacol)	NC			2.38
30	4,4-Methylene-bis-(2,6-di-*tert*-butylphenol)	NC			0
49	Eugenol	NC			2.92
Inorganic acids					
4	Boron trifluoride dihydrate	C	R35	I	—
28	Phosphorus tribromide	C	R35	I	—
32	Phosphorus pentachloride	C	R35	I	—
25	Sulfuric acid (10% wt)	C	R34/R35‡	I/II/III	—
57	Phosphoric acid	C	R34	II	—
43	Hydrochloric acid (14.4% wt)	C	R34	II/III	—
53	Sulfamic acid	NC			—
Inorganic bases					
18	Potassium hydroxide (10%, aq.)	C	R34/R35‡	I/II/III	NPC
42	2-Mercaptoethanol, Na salt (45%, aq.)	C	R34	II/III	NPC
21	Potassium hydroxide (5%, aq.)	NC			5.22
24	Sodium carbonate (50%, aq.)	NC			2.33
Inorganic salts					
20	Iron (III) chloride	C	R34	II	—
52	Sodium bicarbonate	NC			0.11
54	Sodium bisulfite	NC			1.0
Electrophiles					
5	Methacrolein	C	R34	II/III	4.11
14	Allyl bromide	C	R34	II/III	7.17
48	Glycol bromoacetate (85%)	C	R34	II/III	7.67
6	Phenethyl bromide	NC			0
31	2-Bromobutane	NC			2.44
33	4-(Methylthio)-benzaldehyde	NC			0.89
39	2-Ethoxyethyl methacrylate	NC			1.56
46	Cinnamaldehyde	NC			3.71
Soaps/surfactants					
37	Sodium undecylenate (33%, aq.)	NC			1.67
41	20/80 Coconut/palm soap	NC			2.67
60	Sodium lauryl sulfate (20%, aq.)	NC			6.78

*PII = primary irritation index (Bagley *et al.*, 1996; ECETOC, 1995); †NPC = not possible to calculate; ‡ = the animal data and other supporting information are not sufficiently comprehensive to enable unequivocal classification as R34/II & III or R35/I; however, it is more probable that an R34/II & III label is appropriate, and this is the classification which has been used in the analysis of the results obtained in the validation study. The numbers are for the identification of individual chemicals (Barratt *et al.*, 1998).

ANNEX 2: Liebsch et al (2000)[5]のバリデーションで使われた 24 物質

Table VIII: Comparison of predictions obtained with EpiDerm and EPISKIN for the 24 chemicals tested blind in three laboratories, either in the present study, or in the ECVAM skin corrosivity validation study

No.	Chemical name	*In vivo*	EPISKIN	EpiDerm
1	4-Amino-1,2,4-triazole	NC	NC	NC
2	Benzylacetone	NC	NC	NC
3	1,9-Decadiene	NC	NC	NC
4	Dodecanoic (lauric) acid	NC	NC	NC
5	Eugenol	NC	NC	NC
6	Hydrogenated tallow amine	NC	NC	NC
7	Isostearic acid	NC	NC	NC
8	Methyl 2,2-dimethylpropanoate	NC	NC	C[a]
9	Sodium carbonate (50% aq.)	NC	NC	NC
10	Sodium lauryl sulphate (20% aq.)	NC	NC	NC
11	Sulphamic acid	NC	C[a]	C[a]
12	Tetrachloroethylene	NC	NC	NC
13	Boron trifluoride dihydrate	C	C	C
14	2-*tert*-Butylphenol	C	C	C
15	1,2-Diaminopropane	C	C	C
16	Dimethyldipropylenetriamine	C	C	C
17	Dimethylisopropylamine	C	C	C
18	Glycol bromoacetate (85%)	C	C	C
19	*n*-Heptylamine	C	NC[a]	C
20	Methacrolein	C	C	NC[a]
21	Octanoic (caprylic) acid	C	C	C
22	60/40 Octanoic/decanoic acids	C	C	C
23	Phosphorus tribromide	C	C	C
24	Potassium hydroxide (10% aq.)	C	C	C

Each classification represents six independent tests in the case of EpiDerm and nine independent tests in the case of EPISKIN. For clarity, in the few cases of conflicting results obtained in repeated tests or in different laboratories, the classification shown represents the majority of classifications obtained.

[a]*Misclassification.*

ANNEX 3: Kandarova et al (2006)[6]のバリデーションで使われた 12 物質

Table 2
OECD reference chemicals—phase II

No.	Chemical name	CAS no.	In vivo class (C/NC)	Remarks on data supporting classification/ general comments
1	1,2-Diaminopropane	78-90-0	C*	Interaction with MTT was observed
2	Acrylic acid	79-10-7	C*	Published data with EpiDerm or EPISKIN model missing (corrosive on EpiDerm after 3 min at ZEBET—unpublished experiment)
3	2-tert-Butylphenol	88-18-6	C	Borderline C/NC chemical, as judged from the proximity of the chemical to the classification boundary (SAR analysis) (Barratt et al., 1998). Interaction with MTT
4	Potassium hydroxide (10% aq)	1310-58-3	C	C, but supporting data do not enable unequivocal classification as either R34 (II/III) or R 35 (I); more probable to be R 34 (II/III) (Barratt et al., 1998)
5	Octanoic acid (caprylic acid)	124-07-02	C	Borderline C/NC chemical, as judged from the proximity of the chemical to the classification boundary (SAR analysis) (Barratt et al., 1998)
6	Sulfuric acid (10% wt.)	7664-93-9	C	According to the classification mentioned in OECD Guideline 431, the chemical is classified as corrosive According to Annex I of the Directive 67/548/EEC in range of concentration 5–15% the chemical is classified as irritant
7	4-Amino-1,2,4-triazole	584-13-4	NC	Non-irritant
8	Eugenol	97-53-0	NC	Borderline NC/C chemical, as judged from the proximity of the chemical to the classification boundary (SAR analysis) (Barratt et al., 1998) Interaction with MTT
9	Phenethyl bromide	103-63-9	NC	Interaction with MTT
10	Tetrachloroethylene	127-18-4	NC	Classified as C in one of three EPISKIN laboratories in the validation study (Fentem et al., 1998) Very high scores for irritation in vivo in rabbits (ECETOC, 1995)
11	Isostearic acid	30399-84-9	NC	Non-irritant
12	4-(Methylthio)-benzaldehyde	3446-89-7	NC	Interaction with MTT

C*—severely corrosive; C—corrosive; NC—non-corrosive.

ANNEX 4:OECDTG431 (2004)[3]の参照物質

Table 1: Reference Chemicals

1,2-Diaminopropane	CAS-No. 78-90-0	Severely Corrosive
Acrylic Acid	CAS-No. 79-10-7	Severely Corrosive
2-tert-Butylphenol	CAS-No. 88-18-6	Corrosive
Potassium hydroxide (10%)	CAS-No. 1310-58-3	Corrosive
Sulfuric acid (10%)	CAS-No. 7664-93-9	Corrosive
Octanoic acid (caprylic acid)	CAS-No. 124-07-02	Corrosive
4-Amino-1,2,4-triazole	CAS-No. 584-13-4	Not corrosive
Eugenol	CAS-No. 97-53-0	Not corrosive
Phenethyl bromide	CAS-No. 103-63-9	Not corrosive
Tetrachloroethylene	CAS-No. 127-18-4	Not corrosive
Isostearic acid	CAS-No. 30399-84-9	Not corrosive
4-(Methylthio)-benzaldehyde	CAS-No. 3446-89-7	Not corrosive

ANNEX 5: Desprez et al(2015)[7] 80 物質抜粋

EpiSkin™–results of EpiSkin™ are presented for the purpose of the ROC analysis performed on the basis of PMvar2																						
					In vitro results Run 1						*In vitro* results Run 2						*In vitro* results Run 3					
Internal reference	Chemical name	CASRN	*In vivo*	Physical state	v3	v60	V240	vfin	Current PM	PMvar2 (y = 40, z = 38.0	v3	v60	v240	vfin	Current PM	PMvar2 (y = 40, z = 38.0)	v3	v60	v240	vfin	Current PM	PMvar2 (y = 18, z = 130)
1	o-Methoxyphenol (guaiacol)	90-05-1	NC	L	109.5	64.8	36.1	538.9	NC	NC	98.6	49.9	37.3	481.6	NC	NC	83.7	35.5	15.7	134.9	1BC	1BC
2	2,4-Xylidine (2,4-dimethylaniline)	95-68-1	NC	L	109.8	101.2	31.2	242.2	1BC	1BC	124.5	98.3	77.7	674	NC	NC	139.5	93.8	50.8	702.6	NC	NC
3	Phenethyl bromide (2-bromoethy benzene)	103-63-9	NC	L	141	127.2	117.2	808.4	NC	NC	144.7	137.5	139.4	855.7	NC	NC	147.6	173	182.7	946.1	NC	NC
4	Butyl carbamate	592-35-8	NC	S	110.9	113.9	91.1	648.6	NC	NC	108.5	112.1	86.2	632.3	NC	NC	91.6	107.6	93.5	567.5	NC	NC
5	L-Glutamic acid hydrochloride	138-15-8	NC	S	106.7	84.2	33.6	224.5	1BC	1BC	95.3	90.2	68.6	540	NC	NC	94.5	94	42.8	514.8	NC	NC
6	1-(o-Tolyl)biguanide	93-69-6	NC	S	115.6	110.3	104.9	677.6	NC	NC	94.6	98.5	98.9	575.8	NC	NC	94.6	107.1	104	589.5	NC	NC
7	Butyl glycolate (polysolvan)	7397-62-8	NC	L	110.9	108.2	82.6	634.4	NC	NC	88.7	92.6	74.5	521.9	NC	NC	96.1	105.7	94.1	584.2	NC	NC
8	2-Hydroxyisobutyiric acid	594-61-6	NC	S	94.7	42.6	5.4	142.7	1BC	1BC	95.5	41.4	3.7	140.6	1BC	1BC	103	35	6.7	144.7	1BC	1BC
9	Oxalic acid dihydrate	6153-56-6	NC	S	103.6	38.9	4.2	146.7	1BC	1BC	91.5	51.9	17.5	160.9	1BC	1BC	109.6	50.3	6.4	166.3	1BC	1BC
10	alpha-Ketoglutaric acid	328-50-7	NC	S	71.5	18.1	4.1	89.6	1BC	1BC	101.7	19.8	6.6	121.5	1BC	1BC	101.7	10.5	5.7	112.2	1BC	1BC
11	Sulphamic acid	5329-14-6	NC	S	102.9	26.6	20.9	129.5	1BC	1BC	97.9	19.4	0.6	117.3	1BC	1BC	111.2	34.2	2.5	145.4	1BC	1BC
12	Dodecanoic acid (lauric acid)	143-07-7	NC	S	102	117.4	120.8	646.2	NC	NC	104	111.5	141.7	669.2	NC	NC	104.6	94.4	108.4	621.2	NC	NC
13	Sodium lauryl sulphate (20%)	151-21-3	NC	L	104.9	100.1	79	598.7	NC	NC	114	114.6	94.2	664.8	NC	NC	89	90.2	80.6	526.8	NC	NC
14	Methyl trimethylacetate	598-98-1	NC	L	111.3	117.4	75.3	637.9	NC	NC	106.5	100.9	94	620.9	NC	NC	116.2	109.3	97.2	671.3	NC	NC
15	4-Amino-4H-1,2,4-triazole	584-13-4	NC	S	116.8	120.6	79.6	667.4	NC	NC	105.9	106	99.4	629	NC	NC	105	105.6	97.2	622.8	NC	NC
16	1,9-Decadiene	1647-16-1	NC	L	109.5	112.4	87.8	638.2	NC	NC	114.3	136.3	126.7	720.2	NC	NC	137.6	128.6	110.2	789.2	NC	NC
17	Sodium carbonate (50%)	497-19-8	NC	L	103.8	65.2	60.7	541.1	NC	NC	140	107.5	44.7	712.2	NC	NC	73.8	79	43.5	417.7	NC	NC

(continued on next page)

Table 9 (*continued*)

18	Benzylacetone (4–phenyl–2–butanone)	2550–26–7	NC	L	133.2	141.2	137.1	811.1	NC	NC	134.7	151.3	149.9	840	NC	NC	150.3	142.2	143.4	886.8	NC	NC
19	Eugenol	97–53–0	NC	L	188.4	122.3	42.7	918.6	NC	NC	121.7	86	12.1	219.8	1BC	1BC	132.9	74.8	14.2	221.9	1BC	1BC
20	Tetrachloroethylene	127–18–4	NC	L	106.5	88.4	68.5	582.9	NC	NC	113.1	116.3	56.2	624.9	NC	NC	112	97.4	75	620.4	NC	NC
21	Sodium undecylenate (33%)	3398–33–2	NC	L	137.6	36	10.7	184.3	1BC	1BC	127.7	40.2	11	178.9	1BC	1BC	149.7	55	16.5	221.2	1BC	1BC
22	4–Amino–5–methoxy–2–methylbenzensulphonic acid	6471–78–9	NC	S	99.9	115.6	98.6	613.8	NC	NC	105.7	92.3	85.1	600.2	NC	NC	98.2	112.6	101.2	606.6	NC	NC
23	Potassium hydroxide (5%)	1310–58–3	NC	L	72.3	24.6	16.5	96.9	1BC	1BC	68.5	30.1	14.6	98.6	1BC	1BC	94.8	18.9	26.8	113.7	1BC	1BC
24	3,3–Dithiopropionic acid	1119–62–6	NC	S	113.4	117.5	105.8	676.9	NC	NC	107.9	108.9	95.7	636.2	NC	NC	110.3	102.1	113.1	656.4	NC	NC
25	Isopropanol	67–63–0	NC	L	98.9	84.4	88.5	568.5	NC	NC	91.8	87.8	80.8	535.8	NC	NC	98.1	100.2	94.7	587.3	NC	NC
26	2–Phenylalcohol (2–Phenetyl etanol)	60–12–8	NC	L	102.1	98.1	91.5	598	NC	NC	110.5	108.3	126.1	676.4	NC	NC	88.1	87.6	100.1	540.1	NC	NC
27	n–Butyl propionate	590–01–2	NC	L	106.3	80.5	58.7	564.4	NC	NC	111.5	105.7	63.1	614.8	NC	NC	95.3	70.4	47.9	499.5	NC	NC
28	Methyl palmitate	112–39–0	NC	S	108.3	105.3	97.2	635.7	NC	NC	115.7	109.3	92.1	664.2	NC	NC	96.5	80.1	103.1	569.2	NC	NC
29	Methyl laurate	111–82–0	NC	L	100.9	100.2	93.6	597.4	NC	NC	102.4	95.8	100.5	605.9	NC	NC	105.8	117.9	110.8	651.9	NC	NC
30	Sodium bicarbonate	144–55–8	NC	S	94.3	95.1	90.7	563	NC	NC	105	102.1	115.3	637.4	NC	NC	102.8	92.3	97.4	600.9	NC	NC
31	2–Bromobutane	78–76–2	NC	L	105.6	85.5	35.3	543.2	NC	NC	101.8	95.1	103.2	605.5	NC	NC	133.8	104	54.3	693.5	NC	NC
32	4–(Methylthio)–benzaldehyde	3446–89–7	NC	L	136.7	150.4	138.1	835.3	NC	NC	143.7	150.3	150.7	875.8	NC	NC	142.2	158.3	154.2	881.3	NC	NC
33	2–Ethoxyethyl methacrylate	2370–63–0	NC	L	132	133.2	125.8	787	NC	NC	142	139.6	164.5	872.1	NC	NC	133.1	139.7	154.6	826.7	NC	NC
34	Cinnamaldehyde	14371–10–9	NC	L	142.1	125.1	99.3	792.8	NC	NC	134.5	97.3	80	715.3	NC	NC	138.5	94.2	48.8	697	NC	NC
35	4,4′–Methylene–bis–(2,6–ditert–butylphenol)	118–82–1	NC	S	109.5	100.9	102.7	641.6	NC	NC	110.3	104.8	100.7	646.7	NC	NC	110.6	100.5	95.4	638.3	NC	NC
36	Sodium bisulfite	7631–90–5	NC	S	94.9	67.6	42.3	489.5	NC	NC	89.4	92.8	93.8	544.2	NC	NC	71.3	54.2	47.4	386.8	NC	NC
37	10–Undecenoic acid	112–38–9	NC	S	118.2	67.4	60	600.2	NC	NC	114.6	134	102.3	694.7	NC	NC	96.7	93.8	101.6	582.2	NC	NC
38	N,N–Dimethylbenzylamine	103–83–3	1BC	L	97.4	50	20.6	168	1BC	1BC	98	38.5	19.7	156.2	1BC	1BC	85.9	44.1	14.9	144.9	1BC	1BC
39	Fluoboric acid (hydrogentetrafluoroborate) (48%)	16872–11–0	1BC	L	11.5	4.1	3.5	11.5	1A	1A	18.9	4.1	4.8	18.9	1A	1A	9.6	2.5	2.7	9.6	1A	1A
40	Maleic anhydride	108–31–6	1BC	S	78.8	13	2.7	91.8	1BC	1BC	72.5	10.1	4.3	82.6	1BC	1BC	80.2	6	5.5	86.2	1BC	1BC
41	60/40 octanoic/decanoc acid	68937–75–7	1BC	L	77.4	7.4	2	84.8	1BC	1BC	55	12.4	3.4	67.4	1BC	1BC	103.6	18	7.3	121.6	1BC	1BC
42	55/45 octanoic/decanoc acid	68937–75–7	1BC	L	59.3	18	4.1	77.3	1BC	1BC	68.8	13.6	4.6	82.4	1BC	1BC	103.2	7.6	3	110.8	1BC	1BC

43	65/35 octanoic/decanoic acid	68937–75–7	1BC	L	62.4	8.9	3.4	71.3	1BC	1BC	84.4	7.4	2.7	91.8	1BC	1BC	100.7	8.5	3.5	109.2	1BC	1BC
44	N,N-dimethylisopropylamine	996–35–0	1BC	L	94.3	12.9	8.9	107.2	1BC	1BC	87.7	14.2	8.3	101.9	1BC	1BC	77.6	13.7	9.4	91.3	1BC	1BC
45	Hydrochloric acid (14.4%)	7647–01–0	1BC	L	69.3	5.7	1.9	75	1BC	1BC	63.3	2.2	4.6	65.5	1BC	1BC	65.3	2.5	6.8	67.8	1BC	1BC
46	n-Heptylamine	111–68–2	1BC	L	36.2	50.7	31.9	118.8	1BC	1BC	43.9	25.5	9.7	69.4	1BC	1BC	26.2	22.9	13.3	26.2	1A	1A
47	Octanoic acid (caprylic acid)	124–07–2	1BC	L	16.5	4.1	5.3	16.5	1A	1A	25	6.7	5.3	25	1A	1A	18.2	3.7	3.7	18.2	1A	1A
48	Carvacrol	499–75–2	1BC	L	48.9	23.4	12.1	72.3	1BC	1BC	57.3	26.5	12.5	83.8	1BC	1BC	73.4	31.1	14.7	104.5	1BC	1BC
49	2-tert-Butylphenol	88–18–6	1BC	L	86.3	7.5	12.3	93.8	1BC	1BC	52.9	20.8	9.2	73.7	1BC	1BC	60.7	7.2	9.7	67.9	1BC	1BC
50	Methacrolein	78–85–3	1BC	L	85.4	20.6	0	106	1BC	1BC	122.3	32.6	45.4	154.9	1BC	1BC	92.3	42.9	27	162.2	1BC	1BC
51	Lactic acid	598–82–3	1BC	L	80.2	9.4	6.8	89.6	1BC	1BC	72.3	6.3	4	78.6	1BC	1BC	93.6	8.9	4.2	102.5	1BC	1BC
52	Sodium bisulphate monohydrate	10034–88–5	1BC	S	108	51.5	12.2	171.7	1BC	1BC	102.3	40.9	27.2	170.4	1BC	1BC	88.5	44.3	18.8	151.6	1BC	1BC
53	Glyoxylic acid monohydrate	563–96–2	1BC	S	110.4	22.5	24.6	132.9	1BC	1BC	93.6	40.5	12.3	146.4	1BC	1BC	84.9	27.6	10.8	112.5	1BC	1BC
54	Sodium bisulphate	7681–38–1	1BC	S	100.1	31.7	4	131.8	1BC	1BC	103.7	44.1	10.7	158.5	1BC	1BC	83.4	35.9	14.6	133.9	1BC	1BC
55	Cyclohexylamine	108–91–8	1BC	L	89.8	13.1	8.4	102.9	1BC	1BC	46.8	3.4	18.3	50.2	1BC	1BC	73.9	9.2	10.3	83.1	1BC	1BC
56	2-Methylbutyric acid	600–07–7	1BC	L	80.4	2.2	3.2	82.6	1BC	1BC	35.9	3.9	5.6	39.8	1BC	1A	83	5	3.6	88	1BC	1BC
57	Glycol bromoacetate (85%)	3785–34–0	1BC	L	106.3	33.8	27	140.1	1BC	1BC	70.6	56	30.6	157.2	1BC	1BC	90.1	71.3	16.9	178.3	1BC	1BC
58	3-Methoxypropylamine	5332–73–0	1BC	L	32.8	20.1	12.9	32.8	1A	1A	23.8	15.8	7.9	23.8	1A	1A	38	30	29.8	68	1BC	1BC
59	Allyl bromide	106–95–6	1BC	L	113.9	37.9	8.2	160	1BC	1BC	119.4	17.4	21.9	136.8	1BC	1BC	108.7	21.1	7.7	129.8	1BC	1BC
60	1-(2-Aminoethyl)piperazine	140–31–8	1BC	L	89.5	69.7	26.7	185.9	1BC	1BC	90.9	41.8	5.5	138.2	1BC	1BC	87.2	57.6	27.3	172.1	1BC	1BC
61	Iron(III) chloride	7705–08–0	1BC	S	77.6	–	43.1	–	NC	–	80.4	58.6	50.6	430.8	NC	NC	78	89.6	28.3	195.9	1BC	1BC
62	Phosphoric acid	7664–38–2	1BC	L	65.6	20.6	1.7	86.2	1BC	1BC	113.1	9.8	2.3	122.9	1BC	1BC	67.4	20.8	9	88.2	1BC	1BC
63	Propionic acid	79–09–4	1BC	L	3.2	2.5	3.9	3.2	1A	1A	5.5	11.8	4.3	5.5	1A	1A	5.8	4.3	9.3	5.8	1A	1A
64	Butyric acid	107–92–6	1BC	L	3.8	2.4	6.3	3.8	1A	1A	8	2.8	3.4	8	1A	1A	14.8	4.7	4.6	14.8	1A	1A
65	Boron trifluoride–acetic acid complex	373–61–5	1BC	L	29.1	4	5.9	29.1	1A	1A	71.6	3	2.3	74.6	1BC	1BC	28.9	4	3.1	28.9	1A	1A
66	Ethanolamine	141–43–5	1BC	V	66.2	40.3	20.8	127.3	1BC	1BC	105.7	52.3	20.2	178.2	1BC	1BC	78.7	67.4	10.8	156.9	1BC	1BC
67	Hydrobromic acid (48%)	10035–10–6	1BC	L	15.8	16.4	5.9	15.8	1A	1A	4.1	3.2	2.9	4.1	1A	1A	2.8	4.2	4.6	2.8	1A	1A
68	HCl + sulphuric acid + citric acid (5, 5, 5 wt.%)	–	1BC	L	84.6	1.6	4.1	86.2	1BC	1BC	95.4	19.3	3.3	114.7	1BC	1BC	79.4	32.4	4.5	111.8	1BC	1BC
69	1,2-Diaminopropane	78–90–0	1A	L	37.2	21.2	11.8	58.4	1BC	1BC	33	14.7	8.3	33	1A	1A	32	14.4	13	32	1A	1A
70	Phosphorus tribromide	7789–60–8	1A	L	5.5	2.5	8.1	5.5	1A	1A	8.4	8	1.8	8.4	1A	1A	9.4	15.1	8.6	9.4	1A	1A
71	Boron trifluoride dihydrate	13319–75–0	1A	L	2.4	4.2	2.7	2.4	1A	1A	2.5	2.6	1.9	2.5	1A	1A	4.5	2.9	2.6	4.5	1A	1A

(continued on next page)

Table 9 (*continued*)

72	Acrylic acid	79–10–7	1A	L	1.8	2.7	3.2	1.8	1A	1A	2.4	3.8	3.4	2.4	1A	1A	2.8	2.5	2.9	2.8	1A	1A
73	Formic acid	64–18–6	1A	L	4.3	5.6	9.6	4.3	1A	1A	5.7	4.4	5.8	5.7	1A	1A	7.8	4.8	9.9	7.8	1A	1A
74	Dichloroacetyl chloride	79–36–7	1A	L	5.6	6.3	8.3	5.6	1A	1A	5.8	8.5	10.2	5.8	1A	1A	6.2	10.5	8.1	6.2	1A	1A
75	Silver nitrate	7761–88–8	1A	S	12.1	13.4	14.5	12.1	1A	1A	80.6	4.4	1.2	85	1BC	1BC	66.9	2.5	6.4	69.4	1BC	1BC
76	Phenol	108–95–2	1A	S	29.8	21.8	23.1	29.8	1A	1A	22	18.4	18.5	22	1A	1A	21.4	17.2	17.2	21.4	1A	1A
77	Acetic acid	64–19–7	1A	L	2.4	5.6	3	2.4	1A	1A	4.5	4.7	2.8	4.5	1A	1A	2.9	4.1	2.6	2.9	1A	1A
78	Bromoacetic acid	79–08–3	1A	S	3	2.8	3.5	3	1A	1A	3	2.5	3.7	3	1A	1A	2	2.7	4.3	2	1A	1A
79	N,N–dimethyl–dipropylenetriamine	10563–29–8	1A	L	93.5	55.1	23.3	171.9	1BC	1BC	90.8	45	32	167.8	1BC	1BC	74.4	70.6	28.7	173.7	1BC	1BC
80	Sulphuric acid (98%)	7664–93–9	1A	L	8	13.8	9.7	8	1A	1A	10.5	9.9	10.1	10.5	1A	1A	14.1	13.9	14.4	14.1	1A	1A

ANNEX 6: OECD (2013)[18)]予測性データ抜粋

表 A-1. EpiDerm^{TM} の腐食性有無の予測性

EpiDerm with entire set of chemicals (80 chemicals tested over 3 runs, i.e. 240 classifications)				
	Test method: EpiDerm			
In vivo categories	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	Sum
In vivo Cat. 1A	33	3	0	36
In vivo Cat. 1BC	39	54	0	93
In vivo Cat. NC	3	26	82	111
Sum	75	83	82	240

Within Corrosive: Cat. 1A Versus Cat. 1BC	
Sensitivity for 1A	91.67 %
Sensitivity for 1BC	58.06 %
Accuracy	67.44 %
Cat. Corr. Versus Cat. Non Corr	
Sensitivity for Corr.	100.00 %
Sensitivity for Non-Corr.	73.87 %
Accuracy	87.92 %

Calculations over the 3 chemicals categories:			
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:	
70.42%		**% OverClass 1BC as 1A**	**%OverClassNC as 1BC**	**% OverClass NC as 1A**
		41.94%	23.42%	2.70%
		% UnderClass 1A as 1BC	**% underClass 1A as NC**	**% UnderClass 1BC as NC**
		8.33%	0.00%	0.00%
Correctly Classified		**% of OverClass**	**% of UnderClass**	
Cat.1A	91.67%	28.33%	1.25%	
Cat.1BC	58.06%			
Cat. NC	73.87%			

Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
%OverClass NC as Corr. 26.13%	**%OverClass 1BC as 1A** 41.94%	If % OverClass NC as Corr≥50%	If OverClass 1BC as 1A≥50%
%UnderClass 1BC as NC 0.00%	**%UnderClass 1A as NC** 0.00%	If % UnderClass. 1BC as NC≥30%	If UnderClass. 1A as NC≥30%

表 A-2. EpiSkin^{TM} の腐食性有無の予測性

EpiSkin with entire set of chemicals (80 chemicals tested over 3 runs, i.e. 240 classifications)				
	Test method: EpiSkin			
In vivo categories	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	Sum
In vivo Cat. 1A	30	6	0	36
In vivo Cat. 1BC	20	71	2	93
In vivo Cat. NC	0	23	88	111
Sum	50	100	90	240

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	83.33 %
Sensitivity for 1BC	78.02 %
Accuracy	79.53 %
Cat. Corr. Versus Cat. Non Corr	

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
78.75%		**% OverClass 1BC as 1A**	**%OverClassNC as 1BC**	**% OverClass NC as 1A**
		21.51%	20.72%	0.00%
		% UnderClass 1A as 1BC	**% underClass 1A as NC**	**% UnderClass 1BC as NC**
		16.67%	0.00%	2.15%
Correctly Class		**% of OverClass**	**% of UnderClass**	
Cat.1A	83.33%	17.92%	3.33%	
Cat.1BC	76.34%			
Cat. NC	79.28%			

Sensitivity for Corr.	98.45 %
Sensitivity for Non Corr.	79.28 %
Accuracy	89.58 %

Fentem's criteria		According to Fentem, tests are UNACCEPTABLE if...	
%OverClass NC as Corr. 20.72%	**%OverClass 1BC as 1A** 21.51%	If % OverClass NC as Corr≥50%	If OverClass 1BC as 1A≥50%
%UnderClass 1BC as NC 2.15%	**%UnderClass 1A as NC** 0.00%	If % UnderClass. 1BC as NC≥30%	If UnderClass. 1A as NC≥30%

表 A-3．SkinEthic^{TM} の腐食性有無の予測性

SkinEthic with entire set of chemicals (80 chemicals tested over 3 runs, i.e. 240 classifications)				
	Test method: SkinEthic			
In vivo categories	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	Sum
In vivo Cat. 1A	31	5	0	36
In vivo Cat. 1BC	43	43	7	93
In vivo Cat. NC	3	27	81	111
Sum	77	75	88	240

Within Corrosive: Cat. 1A Versus Cat. 1BC	
Sensitivity for 1A	86.11 %
Sensitivity for 1BC	50.00 %
Accuracy	60.66 %
Corr. Versus Non Corr	
Sensitivity for Corr	94.57 %
Sensitivity for Non Corr	72.97 %
Accuracy	84.58 %

Calculations over the 3 chemicals categories:			
Accuracy (Pred. C)	Checking the misclassifications over the 3 categories:		
64.58%	**% OverClass 1BC as 1A**	**%OverClass NC as 1BC**	**% OverClass NC as 1A**
	46.24%	24.32%	2.70%
	% UnderClass 1A as 1BC	**% underClass 1A as NC**	**% UnderClass 1BC as NC**
	13.89%	0.00%	7.53%
Correctly Class			
Cat.1A 86.11%	**% of OverClass**	**% of UnderClass**	
Cat.1BC 46.24%	30.42%	5.00%	
Cat. NC 72.97%			
Fentem's criteria	**According to Fentem, tests are UNACCEPTABLE if...**		

%OverClass NC as Corr.	**%OverClass 1BC as 1A**	If % OverClass NC as Corr≥50%	If OverClass 1BC as 1A≥50%
27.03%	46.24%		
%UnderClass 1BC as NC	**%UnderClass 1A as NC**	If % UnderClass. 1BC as NC≥30%	If UnderClass. 1A as NC≥30%
7.53%	0.00%		

表 A-4．EpiCS®腐食性有無の予測性

epiCS with entire set of chemicals				
	Test method: epiCS			
In vivo categories	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	Sum
In vivo Cat. 1A	22	2	0	24
In vivo Cat. 1BC	28	29	4	61
In vivo Cat. NC	0	21	53	74
Sum	50	52	57	159

Within 1A Vs. 1BC	
Sensitivity for 1A	91.67%
Specificity = Sensitivity for 1BC	50.88%
Accuracy	62.96%
Corr. Vs. Non Corr	
Sensitivity for Corr	95.29%
Sp for Coor = Se for Non Corr	71.62%
Accuracy	84.28%

Calculations over the 3 chemicals categories:			
Accuracy (Pred. C)	Checking the misclassifications over the 3 categories:		
65.41%	**% OverClass 1BC as 1A**	**%OverClassNC as 1BC**	**% OverClass NC as 1A**
	45.90%	28.38%	0.00%
	% UnderClass 1A as 1BC	**% underClass 1A as NC**	**% UnderClass 1BC as NC**
	8.33%	0.00%	6.56%
Correctly Class			
Cat.1A 91.67%	**% of OverClass**	**% of UnderClass**	
Cat.1BC 47.54%	30.82%	3.77%	
Cat. NC 71.62%			

Adopted:
29 July 2016

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

In vitro skin corrosion: reconstructed human epidermis (RHE) test method

INTRODUCTION

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

2. The assessment of skin corrosion potential of chemicals has typically involved the use of laboratory animals (OECD Test Guideline 404 (TG 404); originally adopted in 1981 and revised in 1992, 2002 and 2015) (2). In addition to the present TG 431, two other *in vitro* test methods for testing corrosion potential of chemicals have been validated and adopted as OECD Test Guidelines 430 (3) and 435 (4). Furthermore the *in vitro* OECD TG 439 (5) has been adopted for testing skin irritation potential. A document on Integrated Approaches to Testing and Assessment (IATA) for Skin Corrosion and Irritation describes several modules which group information sources and analysis tools, and provides guidance on (i) how to integrate and use existing testing and non-testing data for the assessment of skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed (6).

3. This Test Guideline addresses the human health endpoint skin corrosion. It makes use of reconstructed human *epidermis* (RhE) (obtained from human derived non-transformed epidermal keratinocytes) which closely mimics the histological, morphological, biochemical and physiological properties of the upper parts of the human skin*, i.e.* the *epidermis*. This Test Guideline was originally adopted in 2004 and updated in 2013 to include additional test methods using the RhE modelsand the possibility to use the methods to support the sub-categorisation of corrosive chemicals, and updated in 2015 to refer to the IATA guidance document and introduce the use of an alternative procedure to measure viability.

4. Four validated test methods using commercially available RhE models are included in this Test Guideline. Prevalidation studies (7), followed by a formal validation study for assessing skin corrosion (8) (9) (10) have been conducted (11) (12) for two of these commercially available test methods, EpiSkin™ Standard Model (SM) and EpiDerm™ Skin Corrosivity Test (SCT) (EPI-200) (referred to in the following text as the Validated Reference Methods – VRMs). The outcome of these studies led to the recommendation that the two VRMs mentioned above could be used for regulatory purposes for distinguishing corrosive (C) from non-corrosive (NC) substances, and that the EpiSkin™ could moreover

This Guideline was adopted by the OECD Council by written procedure on 29 July 2016 [C(2016)103].

be used to support sub-categorisation of corrosive substances (13) (14) (15). Two other commercially available *in vitro* skin corrosion RhE test methods have shown similar results to the EpiDerm™ VRM according to PS-based validation (16) (17) (18). These are the SkinEthic™ RHE[1] and epiCS® (previously named EST-1000) that can also be used for regulatory purposes for distinguishing corrosive from non-corrosive substances (19) (20). Post validation studies performed by the RhE model producers in the years 2012 to 2014 with a refined protocol correcting interferences of unspecific MTT reduction by the test chemicals improved the performance of both discrimination of C/NC as well as supporting sub-categorisation of corrosives (21) (22). Further statistical analyses of the post-validation data generated with EpiDerm™ SCT, SkinEthic™ RHE and EpiCS® have been performed to identify alternative predictions models that improved the predictive capacity for sub-categorisation (23).

5. Before a proposed similar or modified *in vitro* RhE test method for skin corrosion other than the VRMs can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure its similarity to the VRMs, in accordance with the requirements of the Performance Standards (PS) (24) set out in accordance with the principles of Guidance Document No.34 (25). The Mutual Acceptance of Data will only be guaranteed after any proposed new or updated test method following the PS have been reviewed and included in this Test Guideline. The test methods included in this Test Guideline can be used to address countries' requirements for test results on *in vitro* test method for skin corrosion, while benefiting from the Mutual Acceptance of Data.

DEFINITIONS

6. Definitions used are provided in Annex 1.

INITIAL CONSIDERATIONS

7. This Test Guideline allows the identification of non-corrosive and corrosive substances and mixtures in accordance with the UN GHS (1). This Test Guideline further supports the sub-categorisation of corrosive substances and mixtures into optional Sub-category 1A, in accordance with the UN GHS (1), as well as a combination of Sub-categories 1B and 1C (21) (22) (23). A limitation of this Test Guideline is that it does not allow discriminating between skin corrosive Sub-category 1B and Sub-category 1C in accordance with the UN GHS (1) due to the limited set of well-known *in vivo* corrosive Sub-category 1C chemicals. EpiSkin™, EpiDerm™ SCT, SkinEthic™ RHE and epiCS® test methods are able to sub-categorise (i.e. 1A versus 1B-and-1C versus NC)

8. A wide range of chemicals representing mainly individual substances has been tested in the validation supporting the test methods included in this Test Guideline when they are used for identification of non-corrosives and corrosives; the empirical database of the validation study amounted to 60 chemicals covering a wide range of chemical classes (8) (9) (10). Testing to demonstrate sensitivity, specificity, accuracy and within-laboratory-reproducibility of the assay for sub-categorisation was performed by the test method developers and results were reviewed by the OECD (21) (22) (23). On the basis of the overall data available, the Test Guideline is applicable to a wide range of chemical classes and physical states including liquids, semi-solids, solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other prior treatment of the sample is required. In cases where evidence can be demonstrated on the non-applicability of test methods included in the Test Guideline to a specific category of test chemicals, these test methods should not be used for that specific category of test chemicals. In

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

addition, this Test Guideline is assumed to be applicable to mixtures as an extension of its applicability to substances. However, due to the fact that mixtures cover a wide spectrum of categories and composition, and that only limited information is currently available on the testing of mixtures, in cases where evidence can be demonstrated on the non-applicability of the Test Guideline to a specific category of mixtures (*e.g.* following a strategy as proposed in (26)), the Test Guideline should not be used for that specific category of mixtures. Before use of the test guideline on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture. Gases and aerosols have not been assessed yet in validation studies (8) (9) (10). While it is conceivable that these can be tested using RhE technology, the current Test Guideline does not allow testing of gases and aerosols.

9. Test chemicals absorbing light in the same range as MTT formazan and test chemicals able to directly reduce the vital dye MTT (to MTT formazan) may interfere with the tissue viability measurements and need the use of adapted controls for corrections. The type of adapted controls that may be required will vary depending on the type of interference produced by the test chemical and the procedure used to measure MTT formazan (see paragraphs 25-31).

10. While this Test Guideline does not provide adequate information on skin irritation, it should be noted that OECD TG 439 specifically addresses the health effect skin irritation *in vitro* and is based on the same RhE test system, though using another protocol (5). For a full evaluation of local skin effects after a single dermal exposure, the Guidance Document No. 203 on Integrated Approaches for Testing Assessment should be consulted (6). This IATA approach includes the conduct of *in vitro* tests for skin corrosion (such as described in this Test Guideline) and skin irritation before considering testing in living animals. It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

PRINCIPLE OF THE TEST

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

12. The RhE test method is based on the premise that corrosive chemicals are able to penetrate the *stratum corneum* by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. Cell viability is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (27). Corrosive chemicals are identified by their ability to decrease cell viability below defined threshold levels (see paragraphs 35 and 36). The RhE-based skin corrosion test methods have shown to be predictive of *in vivo* skin corrosion effects assessed in rabbits according to the OECD guideline 404 (2).

DEMONSTRATION OF PROFICIENCY

13. Prior to routine use of any of the four validated RhE test methods that adhere to this Test Guideline, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances listed in Table 1. In case of the use of a method for sub-classification, also the correct sub-categorisation should be demonstrated. In situations where a listed substance is unavailable or where justifiable, another substance for which adequate *in vivo* and *in vitro* reference data are available may be used (e.g. from the list of reference chemicals (24)) provided that the same selection criteria as described in Table 1 is applied.

Table 1: List of Proficiency Substances[1]

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

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

[1]The proficiency substances, sorted first by corrosives versus non-corrosives, then by corrosive sub-category and then by chemical class, were selected from the substances used in the ECVAM validation studies of EpiSkin™ and EpiDerm™ (8) (9) (10) and from post-validation studies based on data provided by EpiSkin™ (22), EpiDerm™, SkinEthic™ and epiCS® developers (23). Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (8) (10). The selection includes, to the extent possible, substances that: (i) are representative of the range of corrosivity responses (*e.g.* non-corrosives; weak to strong corrosives) that the VRMs are capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation studies; (iii) have chemical structures that are well-defined; (iv) induce reproducible results in the VRM; (v) induce definitive results in the *in vivo* reference test method; (vi) are commercially available; and (vii) are not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt *et al.* (8).

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

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

[5]The viability values obtained in the ECVAM Skin Corrosion Validation Studies were not corrected for direct MTT reduction (killed controls were not performed in the validation studies). However, the post-validation data generated by the test method developers that are presented in this table were acquired with adapted controls (23).

14. As part of the proficiency exercise, it is recommended that the user verifies the barrier properties of the tissues after receipt as specified by the RhE model manufacturer. This is particularly important if tissues are shipped over long distance/time periods. Once a test method has been successfully established

and proficiency in its use has been demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties in regular intervals.

PROCEDURE

15. The following is a generic description of the components and procedures of the RhE test methods for skin corrosion assessment covered by this Test Guideline. The RhE models endorsed as scientifically valid for use within this Test Guideline, *i.e.* the EpiSkin™ (SM), EpiDerm™ (EPI-200), SkinEthic™ RHE and epiCS® models (16) (17) (19) (28) (29) (30) (31) (32) (33), can be obtained from commercial sources. Standard Operating Procedures (SOPs) for these four RhE models are available (34) (35) (36) (37), and their main test method components are summarised in Annex 2. It is recommended that the relevant SOP be consulted when implementing and using one of these methods in the laboratory. Testing with the four RhE test methods covered by this Test Guideline should comply with the following:

RHE TEST METHOD COMPONENTS

General Conditions

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

Functional Conditions

Viability

17. The assay used for quantifying tissue viability is the MTT-assay (27). The viable cells of the RhE tissue construct reduce the vital dye MTT into a blue MTT formazan precipitate, which is then extracted from the tissue using isopropanol (or a similar solvent). The OD of the extraction solvent alone should be sufficiently small, i.e., OD < 0.1. The extracted MTT formazan may be quantified using either a standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure (38). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control. An acceptability range (upper and lower limit) for the negative control OD values should be established by the RhE model developer/supplier. Acceptability ranges for the negative control OD values for the four validated RhE test methods included in this Test Guideline are given in Table 2. An HPLC/UPLC-Spectrophotometry user should use the negative control OD ranges provided in Table 2 as the acceptance

criterion for the negative control. It should be documented that the tissues treated with negative control are stable in culture (provide similar OD measurements) for the duration of the exposure period.

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

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

Barrier function

18. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of certain cytotoxic benchmark chemicals (*e.g.* SDS or Triton X-100), as estimated by IC_{50} or ET_{50} (Table 3). The barrier function of each batch of the RhE model used should be demonstrated by the RhE model developer/vendor upon supply of the tissues to the end user (see paragraph 21).

Morphology

19. Histological examination of the RhE model should be performed demonstrating multi-layered human *epidermis*-like structure containing *stratum basale*, *stratum spinosum*, *stratum granulosum* and *stratum corneum* and exhibits lipid profile similar to lipid profile of human epidermis. Histological examination of each batch of the RhE model used demonstrating appropriate morphology of the tissues should be provided by the RhE model developer/vendor upon supply of the tissues to the end user (see paragraph 21).

Reproducibility

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

Quality control (QC)

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

Table 3: QC batch release criteria

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM)	IC_{50} = 1.0 mg/mL	IC_{50} = 3.0 mg/mL

(18 hours treatment with SDS)(33)		
EpiDerm™ SCT (EPI-200) (1% Triton X-100)(34)	ET_{50} = 4.0 hours	ET_{50} = 8.7 hours
SkinEthic™ RHE (1% Triton X-100)(35)	ET_{50} = 4.0 hours	ET_{50} = 10.0 hours
epiCS®(1% Triton X-100)(36)	ET_{50} = 2.0 hours	ET_{50} = 7.0 hours

Application of the Test Chemical and Control Substances

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

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

Cell Viability Measurements

24. The MTT assay, which is a quantitative assay, should be used to measure cell viability under this Test Guideline (27). The tissue sample is placed in MTT solution of appropriate concentration (0.3 or 1 mg/mL) for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (*e.g.* isopropanol, acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm, or by an HPLC/UPLC-spectrophotometry procedure (see paragraphs 30 and 31) (38).

25. Test chemicals may interfere with the MTT assay, either by direct reduction of the MTT into blue formazan, and/or by colour interference if the test chemical absorbs, naturally or due to treatment procedures, in the same OD range of formazan (570 ± 30 nm, mainly blue and purple chemicals). Additional controls should be used to detect and correct for a potential interference from these test chemicals such as the non-specific MTT reduction (NSMTT) control and the non-specific colour (NSC) control (see paragraphs 26 to 30). This is especially important when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*, and is therefore present in the tissues when the MTT viability test is performed. Detailed description of how to correct direct MTT

reduction and interferences by colouring agents is available in the SOPs for the test methods (34) (35) (36) (37).

26. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT medium (34) (35) (36) (37). If the MTT mixture containing the test chemical turns blue/purple, the test chemical is presumed to directly reduce the MTT, and further functional check on non-viable epidermis should be performed, independently of using the standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure. This additional functional check employs killed tissues that possess only residual metabolic activity but absorb the test chemical in similar amount as viable tissues. Each MTT reducing chemical is applied on at least two killed tissue replicates per exposure time, which undergo the whole skin corrosion test. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the MTT reducer minus the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT).

27. To identify potential interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol and decide on the need for additional controls, spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, further colorant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required (see paragraphs 30 and 31). When performing the standard absorbance (OD) measurement, each interfering coloured test chemical is applied on at least two viable tissue replicates per exposure time, which undergo the entire skin corrosion test but are incubated with medium instead of MTT solution during the MTT incubation step to generate a non-specific colour (NSC_{living}) control. The NSC_{living} control needs to be performed concurrently per exposure time per coloured test chemical (in each run) due to the inherent biological variability of living tissues. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the interfering test chemical and incubated with MTT solution minus the percent non-specific colour obtained with living tissues exposed to the interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected ($\%NSC_{living}$).

28. Test chemicals that are identified as producing both direct MTT reduction (see paragraph 26) and colour interference (see paragraph 27) will also require a third set of controls, apart from the NSMTT and NSC_{living} controls described in the previous paragraphs, when performing the standard absorbance (OD) measurement. This is usually the case with darkly coloured test chemicals interfering with the MTT assay (e.g., blue, purple, black) because their intrinsic colour impedes the assessment of their capacity to directly reduce MTT as described in paragraph 26. These test chemicals may bind to both living and killed tissues and therefore the NSMTT control may not only correct for potential direct MTT reduction by the test chemical, but also for colour interference arising from the binding of the test chemical to killed tissues. This could lead to a double correction for colour interference since the NSC_{living} control already corrects for colour interference arising from the binding of the test chemical to living tissues. To avoid a possible double correction for colour interference, a third control for non-specific colour in killed tissues (NSC_{killed}) needs to be performed. In this additional control, the test chemical is applied on at least two killed tissue replicates per exposure time, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. A single NSC_{killed} control is sufficient per test chemical regardless of the number of independent tests/runs performed, but should be performed concurrently to the NSMTT control and, where possible, with the same tissue batch. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT minus $\%NSC_{living}$ plus the percent non-specific colour obtained with killed tissues exposed to the interfering test chemical and incubated with medium without MTT, calculated relative to the negative control run concurrently to the test being corrected ($\%NSC_{killed}$).

29. It is important to note that non-specific MTT reduction and non-specific colour interferences may increase the readoutsof the tissue extract above the linearity range of the spectrophotometer. On this basis, each laboratory should determine the linearity range of their spectrophotometer with MTT formazan (CAS # 57360-69-7) from a commercial source before initiating the testing of test chemicals for regulatory purposes. In particular, the standard absorbance (OD) measurement using a spectrophotometer is appropriate to assess direct MTT-reducers and colour interfering test chemicals when the ODs of the tissue extracts obtained with the test chemical without any correction for direct MTT reduction and/or colour interference are within the linear range of the spectrophotometer or when the uncorrected percent viability obtained with the test chemical already defined it as a corrosive (see paragraphs 35 and 36). Nevertheless, results for test chemicals producing %NSMTT and/or $\%NSC_{living} \geq 50\%$ of the negative control should be taken with caution.

30. For coloured test chemicals which are not compatible with the standard absorbance (OD) measurement due to too strong interference with the MTT assay, the alternative HPLC/UPLC-spectrophotometry procedure to measure MTT formazan may be employed (see paragraph 31) (37). The HPLC/UPLC-spectrophotometry system allows for the separation of the MTT formazan from the test chemical before its quantification (38). For this reason, NSC_{living} or NSC_{killed} controls are never required when using HPLC/UPLC-spectrophotometry, independently of the chemical being tested. NSMTT controls should nevertheless be used if the test chemical is suspected to directly reduce MTT or has a colour that impedes the assessment of the capacity to directly reduce MTT (as described in paragraph 26). When using HPLC/UPLC-spectrophotometry to measure MTT formazan, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. For test chemicals able to directly reduce MTT, true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT. Finally, it should be noted that direct MTT-reducers that may also be colour interfering, which are retained in the tissues after treatment and reduce MTT so strongly that they lead to ODs (using standard OD measurement) or peak areas (using UPLC/HPLC-spectrophotometry) of the tested tissue extracts that fall outside of the linearity range of the spectrophotometer cannot be assessed, although these are expected to occur in only very rare situations.

31. HPLC/UPLC-spectrophotometry may be used also with all types of test chemicals (coloured, non-coloured, MTT-reducers and non-MTT reducers) for measurement of MTT formazan (38). Due to the diversity of HPLC/UPLC-spectrophotometry systems, qualification of the HPLC/UPLC-spectrophotometry system should be demonstrated before its use to quantify MTT formazan from tissue extracts by meeting the acceptance criteria for a set of standard qualification parameters based on those described in the U.S. Food and Drug Administration guidance for industry on bio-analytical method validation (38) (39). These key parameters and their acceptance criteria are shown in Annex 4. Once the acceptance criteria defined in Annex 4 have been met, the HPLC/UPLC-spectrophotometry system is considered qualified and ready to measure MTT formazan under the experimental conditions described in this Test Guideline.

Acceptability Criteria

32. For each test method using valid RhE models, tissues treated with the negative control should exhibit OD reflecting the quality of the tissues as described in table 2 and should not be below historically established boundaries. Tissues treated with the PC, *i.e.* glacial acetic acid or 8N KOH, should reflect the ability of the tissues to respond to a corrosive chemical under the conditions of the test method (see Annex 2). The variability between tissue replicates of test chemical and/or control substances should fall within

the accepted limits for each valid RhE model requirements (see Annex 2) (*e.g.* the difference of viability between the two tissue replicates should not exceed 30%). If either the negative control or PC included in a run fall out of the accepted ranges, the run is considered as not qualified and should be repeated. If the variability of test chemicals falls outside of the defined range, its testing should be repeated.

Interpretation of Results and Prediction Model

33. The OD values obtained for each test chemical should be used to calculate percentage of viability relative to the negative control, which is set at 100%. In case HPLC/UPLC-spectrophotometry is used, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. The cut-off percentage cell viability values distinguishing corrosive from non-corrosive test chemical (or discriminating between different corrosive sub-categories) are defined below in paragraphs 35 and 36 for each of the test methods covered by this Test Guideline and should be used for interpreting the results.

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

35. The prediction model for the EpiSkin™ skin corrosion test method (9) (34) (22), associated with the UN GHS (1) classification system, is shown in Table 4:

Table 4:EpiSkin™ prediction model

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

*) According to the data generated in view of assessing the usefulness of the RhE test methods for supporting sub-categorisation, it was shown that around 22 % of the Sub-category 1A results of the EpiSkin™ test method may actually constitute Sub-category 1B or Sub-category 1C substances/mixtures (i.e. over classifications) (see Annex 3).

36. The prediction models for the EpiDerm™ SCT (10) (23) (35), the SkinEthic™ RHE (17) (18) (23) (36), and the epiCS® (16) (23) (37) skin corrosion test methods, associated with the UN GHS (1) classification system, are shown in Table 5:

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

Viability measured after exposure time points (t=3 and 60 minutes)	Prediction to be considered
STEP 1 for EpiDerm™ SCT, for SkinEthic™ RHE and epiCS®	
< 50% after 3 min exposure	**Corrosive**
≥ 50% after 3 min exposure **AND** < 15% after 60 min exposure	**Corrosive**
≥ 50% after 3 min exposure **AND** ≥ 15% after 60 min exposure	**Non-corrosive**
STEP 2 for EpiDerm™ SCT – for substances/mixtures identified as Corrosive in step 1	
< 25% after 3 min exposure	Optional Sub-category 1A *
≥ 25 % after 3 min exposure	A combination of optional Sub-categories 1B-and-1C
STEP 2 for SkinEthic™ RHE – for substances/mixtures identified as Corrosive in step 1	
< 18 % after 3 min exposure	Optional Sub-category 1A *
≥ 18 % after 3 min exposure	A combination of optional Sub-categories 1B-and-1C
STEP 2 for epiCS® – for substances/mixtures identified as Corrosive in step 1	
< 15 % after 3 min exposure	Optional Sub-category 1A *
≥ 15 % after 3 min exposure	A combination of optional Sub-categories 1B-and-1C

* According to the data generated in view of assessing the usefulness of the RhE test methods for supporting sub-categorisation, it was shown that around 29%, 31% and 33% of the Sub-category 1A results of the EpiDerm™ SCT,

SkinEthic™ RHE and epiCS® test methods, respectively, may actually constitute Sub-category 1B or Sub-category 1C substances/mixtures (i.e. over-classifications) (see Annex 3).

DATA AND REPORTING

Data

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

Test report

38. The test report should include the following information:

Test Chemical and Control Substances:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Multi-constituent substance, UVCB and mixture: characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents;
- Physical appearance, water solubility, and any additional relevant physicochemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control substance prior to testing, if applicable (*e.g.* warming, grinding);
- Stability of the test chemical, limit date for use, or date for re-analysis if known;
- Storage conditions.

-

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

Test Conditions:

- RhE model used (including batch number);
- Calibration information for measuring device (*e.g.* spectrophotometer), wavelength and band pass (if applicable) used for quantifying MTT formazan, and linearity range of measuring device;
- Description of the method used to quantify MTT formazan;
- Description of the qualification of the HPLC/UPLC-spectrophotometry system, if applicable;
- Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to:
  i) Viability;
  ii) Barrier function;
  iii) Morphology;
  iv) Reproducibility and predictive capacity;

v) Quality controls (QC) of the model;

- Reference to historical data of the model. This should include, but is not limited to acceptability of the QC data with reference to historical batch data;
- Demonstration of proficiency in performing the test method before routine use by testing of the proficiency substances.

Test Procedure:

- Details of the test procedure used (including washing procedures used after exposure period);
- Doses of test chemical and control substances used;
- Duration of exposure period(s) and temperature(s) of exposure;
- Indication of controls used for direct MTT-reducers and/or colouring test chemicals, if applicable;
- Number of tissue replicates used per test chemical and controls (PC, negative control, and NSMTT, NSCliving and NSCkilled, if applicable), per exposure time;
- Description of decision criteria/prediction model applied based on the RhE model used;
- Description of any modifications of the test procedure (including washing procedures).

Run and Test Acceptance Criteria:

- Positive and negative control mean values and acceptance ranges based on historical data;
- Acceptable variability between tissue replicates for positive and negative controls;
- Acceptable variability between tissue replicates for test chemical.

Results:

- Tabulation of data for individual test chemicals and controls, for each exposure period, each run and each replicate measurement including OD or MTT formazan peak area, percent tissue viability, mean percent tissue viability, differences between replicates, SDs and/or CVs if applicable;
- If applicable, results of controls used for direct MTT-reducers and/or colouring test chemicals including OD or MTT formazan peak area, %NSMTT, %NSCliving, %NSCkilled, differences between tissue replicates, SDs and/or CVs (if applicable), and final correct percent tissue viability;
- Results obtained with the test chemical(s) and control substances in relation to the defined run and test acceptance criteria;
- Description of other effects observed;
- The derived classification with reference to the prediction model/decision criteria used.

Discussion of the results

Conclusions

LITERATURE

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

(2) OECD. (2015). Guideline for Testing of Chemicals. (No. 404.): Acute Dermal Irritation, Corrosion, Organisation for Economic Cooperation and Development, Paris.

(3) OECD. (2015). Guideline for the Testing of Chemicals (No. 430.): *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance (TER). Organisation for Economic Cooperation and Development, Paris.

(4) OECD. (2015). Guideline for the Testing of Chemicals (No. 435.): *In Vitro* Membrane Barrier Test Method. Organisation for Economic Cooperation and Development, Paris. (5) OECD. (2015). Guideline for the Testing of Chemicals (No. 439.): *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Cooperation and Development, Paris.

(6) OECD. (2014). Guidance Document on Integrated Approaches to Testing and Assessment of Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment, (No. 203.) Organisation for Economic Cooperation and Development, Paris.

(7) Botham P.A., Chamberlain M., Barratt M.D., Curren R.D., Esdaile D.J., Gardner J.R., Gordon V.C., Hildebrand B., Lewis R.W., Liebsch M., Logemann P., Osborne R., Ponec M., Regnier J.F., Steiling W., Walker A.P., and Balls M. (1995). A Prevalidation Study on *In Vitro* Skin Corrosivity Testing. The report and Recommendations of ECVAM Workshop 6.*ATLA* 23, 219-255.

(8) Barratt M.D., Brantom P.G., Fentem J.H., Gerner I., Walker A.P., and Worth A.P. (1998). The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 1. Selection and distribution of the Test Chemicals. *Toxicol.In Vitro* 12, 471-482.

(9) Fentem J.H., Archer G.E.B., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J., Holzhutter H.-G., and Liebsch M. (1998). The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 2. Results and Evaluation by the Management Team. *Toxicol.in Vitro* 12, 483-524.

(10) Liebsch M., Traue D., Barrabas C., Spielmann H., Uphill, P., Wilkins S., Wiemann C., Kaufmann T., Remmele M. and Holzhütter H. G. (2000). The ECVAM Prevalidation Study on the Use of EpiDerm for Skin Corrosivity Testing, *ATLA* 28, pp. 371-401.

(11) 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. et Zucco F. (1995). Practical Aspects of the Validation of Toxicity Test Procedures. The Report and Recommendations of ECVAM Workshops, *ATLA* 23, 129-147.(12) ICCVAM (Interagency Coordinating Committee on the

Validation of Alternative Methods). (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. NIH Publication No. 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. Available at: [http://www.iccvam.niehs.nih.gov/docs/guidelines/validate.pdf].

(13) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (2002). ICCVAM evaluation of EpiDerm™ (EPI-200), EPISKIN™ (SM), and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: *In Vitro* Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 02-4502. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.Available at:[http://www.iccvam.niehs.nih.gov/methods/epiddocs/epis_brd.pdf]

(14) EC-ECVAM. (1998). Statement on the Scientific Validity of the EpiSkin™ Test (an *In Vitro* Test for Skin Corrosivity), Issued by the ECVAM Scientific Advisory Committee (ESAC10), 3 April 1998. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

(15) EC-ECVAM. (2000). Statement on the Application of the EpiDerm™ Human Skin Model for Skin Corrosivity Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC14), 21 March 2000. Available at: [http://ecvam.jrc.ec.europa.eu].

(16) Hoffmann J., Heisler E., Karpinski S., Losse J., Thomas D., Siefken W., Ahr H.J., Vohr H.W. and Fuchs H.W. (2005). Epidermal-Skin-Test 1000 (EST-1000)-A New Reconstructed Epidermis for *In Vitro* Skin Corrosivity Testing. *Toxicol.In Vitro* 19, 925-929.

(17) Kandárová H., Liebsch M., Spielmann,H., Genschow E., Schmidt E., Traue D., Guest R., Whittingham A., Warren N, Gamer A.O., Remmele M., Kaufmann T., Wittmer E., De Wever B., and Rosdy M. (2006). Assessment of the Human Epidermis Model SkinEthic RHE for *In Vitro* Skin Corrosion Testing of Chemicals According to New OECD TG 431. *Toxicol.In Vitro* 20, 547–559.

(18) Tornier C., Roquet M. and Fraissinette A.B. (2010). Adaptation of the Validated SkinEthic™ Reconstructed Human Epidermis (RHE) Skin Corrosion Test Method to 0.5 cm^2 Tissue Sample. *Toxicol. In Vitro* 24, 1379-1385.

(19) EC-ECVAM. (2006). Statement on the Application of the SkinEthic™ Human Skin Model for Skin Corrosivity Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC25), 17 November 2006. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

(20) EC-ECVAM. (2009). ESAC Statement on the Scientific Validity of an *In-Vitro* Test Method for Skin Corrosivity Testing: the EST-1000, Issued by the ECVAM Scientific Advisory Committee (ESAC30), 12 June 2009. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

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

(22) Alépée N., Grandidier M.H., and Cotovio J. (2014). Sub-Categorisation of Skin Corrosive Chemicals by the EpiSkin™ Reconstructed Human Epidermis Skin Corrosion Test Method According to UN GHS: Revision of OECD Test Guideline 431. *Toxicol. In Vitro* 28, 131-145.

(23) Desprez B., Barroso J., Griesinger C., Kandárová H., Alépée N., and Fuchs, H. (2015). Two Novel

Prediction Models Improve Predictions of Skin Corrosive Sub-categories by Test Methods of OECD Test Guideline No. 431. *Toxicol. In Vitro* 29, 2055–2080.

(24) OECD. (2015). Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Reconstructed Human Epidermis (RHE) Test Methods For Skin Corrosion in Relation to OECD TG 431. Environmental Health and Safety Publications, Series on Testing and Assessment (No. 219). Organisation for Economic Cooperation and Development, Paris

(25) OECD. (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. .Environment, Health and Safety Publications, Series on Testing and Assessment (No. 34.), Organisation for Economic Cooperation and Development, Paris.

(26) Eskes C. et al. (2012). Regulatory Assessment of *In Vitro* Skin Corrosion and Irritation Data Within the European Framework: Workshop Recommendations. *Regul.Toxicol.Pharmacol.* 62, 393-403.

(27) Mosmann T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. *J. Immunol. Methods* 65, 55-63.

(28) Tinois E., et al. (1994). The Episkin Model: Successful Reconstruction of Human *Epidermis In Vitro*. In: *In Vitro* Skin Toxicology. Rougier A.,. Goldberg A.M and Maibach H.I. (Eds): 133-140.

(29) Cannon C. L., Neal P.J., Southee J.A., Kubilus J. and Klausner M. (1994), New Epidermal Model for Dermal Irritancy Testing. *Toxicol.in Vitro* 8, 889 - 891.

(30) Ponec M., Boelsma E, Weerheim A, Mulder A, Bouwstra J and Mommaas M. (2000). Lipid and Ultrastructural Characterization of Reconstructed Skin Models. Inter. J. Pharmaceu. 203, 211 - 225.

(31) Tinois E., Tillier, J., Gaucherand, M., Dumas, H., Tardy, M. and Thivolet J. (1991). *In Vitro* and Post – Transplantation Differentiation of Human Keratinocytes Grown on the Human Type IV Collagen Film of a Bilayered Dermal Substitute. *Exp. Cell Res.* 193: 310-319.

(32) Parenteau N.L., Bilbo P, Nolte CJ, Mason VS and Rosenberg M. (1992). The Organotypic Culture of Human Skin Keratinocytes and Fibroblasts to Achieve Form and Function. *Cytotech*. 9, 163-171.

(33) Wilkins L.M., Watson SR, Prosky SJ, Meunier SF and Parenteau N.L. (1994). Development of a Bilayered Living Skin Construct for Clinical Applications. *Biotech.Bioeng.*43/8, 747-756.

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

(35) EpiDerm™ SOP. (February 2012). Version MK-24-007-0024Protocol for: *In Vitro* EpiDerm™ Skin Corrosion Test (EPI-200-SCT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

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

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

(38) Alépée N., Barroso J., De Smedt A., De Wever B., Hibatallah J., Klaric M., Mewes K.R., Millet M., Pfannenbecker U., Tailhardat M., Templier M., and McNamee P. Use of HPLC/UPLC-spectrophotometry for Detection of MTT Formazan in In Vitro Reconstructed Human Tissue (RhT)-based Test Methods Employing the MTT Assay to Expand their Applicability to Strongly Coloured Test Chemicals. Manuscript in Preparation.

(39) US FDA. (2001). Guidance for Industry: Bioanalytical Method Validation. U.S. Department of Health and Human Services, Food and Drug Administration. (May 2001). Available at: [http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf].

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (25).

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

Chemical: means a substance or a mixture.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (25).

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

GHS (Globally Harmonized System of Classification and Labelling of Chemicals): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

HPLC: High Performance Liquid Chromatography.

IATA: Integrated Approach on Testing and Assessment.

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

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

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

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80% (w/w).

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

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration ≥ 10% (w/w) and < 80% (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

NC: Non corrosive.

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

NSC_{living}control : Non-Specific Colour control in living tissues.

NSMTT: Non-Specific MTT reduction.

OD: Optical Density

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

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of reliability and accuracy, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (25).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (25).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (25).

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

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (25).

Skin corrosion *in vivo:* The production of irreversible damage of the skin; namely, visible necrosis through the *epidermis* and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (25).

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

Test chemical: means what is being tested.

UPLC: Ultra-High Performance Liquid Chromatography.

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

ANNEX 2

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

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

Test Method Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Pre-check for colour interference	10 µL (liquid) or 10 mg (solid) + 90 µL H_2O mixed for 15 min at RT → if solution becomes coloured, living adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 300 µL H_2O for 60 min at 37°C, 5% CO_2, 95% RH → if solution becomes coloured, living adapted controls should be performed	40 µL (liquid) or 20mg (solid) + 300 µL H_2O mixed for 60 min at RT → if test chemical is coloured, living adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 300 µL H_2O for 60 min at 37°C, 5% CO_2, 95% RH → if solution becomes coloured, living adapted controls should be performed
Exposure time and temperature	3 min, 60 min (± 5 min) and 240 min (± 10 min) In ventilated cabinet Room Temperature (RT, 18-28°C)	3 min at RT, and 60 min at 37°C, 5% CO_2, 95% RH	3 min at RT, and 60 min at 37°C, 5% CO_2, 95% RH	3 min at RT, and 60 min at 37°C, 5% CO_2, 95% RH
Rinsing	25 mL 1x PBS (2 mL/throwing)	20 times with a constant soft stream of 1x PBS	20 times with a constant soft stream of 1x PBS	20 times with a constant soft stream of 1x PBS
Negative control	50 µL NaCl solution (9 g/L) Tested with every exposure time	50 µL H_2O Tested with every exposure time	40 µL H_2O Tested with every exposure time	50 µL H_2O Tested with every exposure time
Positive control	50 µL Glacial acetic acid Tested only for 4 hours	– 50 µL 8N KOH Tested with every exposure time	40 µL 8N KOH Tested only for 1 hour	50 µL 8N KOH Tested with every exposure time
MTT solution	2 mL 0.3 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL
MTT incubation time and temperature	180 min (± 15 min) at 37°C, 5% CO_2, 95% RH	180 min at 37°C, 5% CO_2, 95% RH	180 min (± 15 min) at 37°C, 5% CO_2, 95% RH	180 min at 37°C, 5% CO_2, 95% RH

Test Method Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Extraction solvent	500 µL acidified isopropanol (0.04 N HCl in isopropanol) (isolated tissue fully immersed)	2 mL isopropanol (extraction from top and bottom of insert)	1.5 mL isopropanol (extraction from top and bottom of insert)	2 mL isopropanol (extraction from top and bottom of insert)
Extraction time and temperature	Overnight at RT, protected from light	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT
OD reading	570 nm (545 - 595 nm) without reference filter	570 nm (or 540 nm) without reference filter	570 nm (540 - 600 nm) without reference filter	540 - 570 nm without reference filter
Tissue Quality Control	18 hours treatment with SDS 1.0 mg/mL ≤ IC_{50} ≤ 3.0 mg/mL	Treatment with 1% Triton X-100 4.08 hours ≤ ET_{50} ≤ 8.7 hours	Treatment with 1% Triton X-100 4.0 hours ≤ ET_{50} ≤ 10.0 hours	Treatment with 1% Triton X-100 2.0 hours ≤ ET_{50} ≤ 7.0 hours
Acceptability Criteria	1. Mean OD of the tissue replicates treated with the negative control (NaCl) should be ≥ 0.6 and ≤ 1.5 for every exposure time 2. Mean viability of the tissue replicates exposed for 4 hours with the positive control (glacial acetic acid), expressed as % of the negative control, should be ≤ 20% 3. In the range 20-100% viability and for ODs≥ 0.3, difference of viability between the two tissue replicates should not exceed 30%.	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 2.8 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be < 15% 3. In the range 20 - 100% viability, the Coefficient of Variation (CV) between tissue replicates should be ≤ 30%	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 3.0 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour (and 4 hours, if applicable) with the positive control (8N KOH), expressed as % of the negative control, should be < 15% 3. In the range 20-100% viability, and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not exceed 30%	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 2.8 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be < 20% 3. In the range 20-100% viability, and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not exceed 30%

ANNEX 3

PERFORMANCE OF TEST METHODS FOR SUB-CATEGORISATION

The table below provides the performances of the four test methods calculated based on a set of 80 chemicals tested by the four test developers. Calculations were performed by the OECD Secretariat, reviewed and agreed by an expert subgroup (21) (23).

EpiSkin™, EpiDerm™ , SkinEthic™ and epiCS® test methods are able to sub-categorise (i.e. 1A versus 1B-and-1C versus NC)

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

STATISTICS ON PREDICTIONS OBTAINED ON THE ENTIRE SET OF CHEMICALS **(n= 80 chemicals tested over 2 independent runs for epiCS® or 3 independent runs for EpiDerm™ SCT, EpiSkin™ and SkinEthic™ RHE, i.e. respectively 159* or 240 classifications)** *one chemical was tested once in epiCS® because of no availability (23)				
	EpiSkin™	**EpiDerm™**	**SkinEthic™**	**epiCS®**
Overclassifications:				
1B-and-1C overclassified 1A	21.50%	29.0%	31.2%	32.8%
NC overclassified 1B-and-1C	20.7%	23.4%	27.0 %	28.4 %
NC overclassified 1A	0.00%	2.7%	0.0 %	0.00%
overclassified Corr.	20.7%	26.1%	27.0%	28.4%
Global overclassification rate (all categories)	17.9%	23.3%	24.5%	25.8%
Underclassifications:				
1A underclassified 1B-and-1C	16.7%	16.7 %	16.7%	12.5 %
1A underclassified NC	0.00%	0.00%	0.00%	0.00%
1B-and-1C underclassified NC	2.2%	0.00%	7.5%	6.6%
Global underclassification rate (all categories)	3.3%	2.5%	5.4%	4.4%
Correct Classifications:				
1A correctly classified	83.3%	83.3%	83.3%	87.5%
1B-and-/1C correctly classified	76.3%	71.0%	61.3%	60.7%
NC correctly classified	79.3%	73.9%	73.0%	71.62%
Overall Accuracy	78.8%	74.2%	70%	69.8%

NC: Non-corrosive

ANNEX 4

Key parameters and acceptance criteria for qualification of an HPLC/UPLC-spectrophotometry system for measurement of MTT formazan extracted from RhE tissues

Parameter	Protocol Derived from FDA Guidance (37)(38)	Acceptance Criteria
Selectivity	Analysis of isopropanol, living blank (isopropanol extract from living RhE tissues without any treatment), dead blank (isopropanol extract from killed RhE tissues without any treatment)	$Area_{interference} \leq 20\%$ of $Area_{LLOQ}$[1]
Precision	Quality Controls (i.e., MTT formazan at 1.6 µg/mL, 16 µg/mL and 160 µg/mL) in isopropanol (n=5)	CV ≤ 15% or ≤ 20% for the LLOQ
Accuracy	Quality Controls in isopropanol (n=5)	%Dev ≤ 15% or ≤ 20% for LLOQ
Matrix Effect	Quality Controls in living blank (n=5)	85% ≤ Matrix Effect % ≤ 115%
Carryover	Analysis of isopropanol after an ULOQ[2] standard	$Area_{interference} \leq 20\%$ of $Area_{LLOQ}$
Reproducibility (intra-day)	3 independent calibration curves (based on 6 consecutive 1/3 dilutions of MTT formazan in isopropanol starting at ULOQ, i.e., 200 µg/mL); Quality Controls in isopropanol (n=5)	Calibration Curves: %Dev ≤ 15% or ≤ 20% for LLOQ Quality Controls: %Dev ≤ 15% and CV ≤ 15%
Reproducibility (inter-day)	Day 1: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 2: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 3: 1 calibration curve and Quality Controls in isopropanol (n=3)	
Short Term Stability of MTT Formazan in RhE Tissue Extract	Quality Controls in living blank (n=3) analysed the day of the preparation and after 24 hours of storage at room temperature	%Dev ≤ 15%
Long Term Stability of MTT Formazan in RhE Tissue Extract, if required	Quality Controls in living blank (n=3) analysed the day of the preparation and after several days of storage at a specified temperature (e.g., 4ºC, -20ºC, -80ºC)	%Dev ≤ 15%

[1] LLOQ: Lower Limit of Quantification, defined to cover 1-2% tissue viability, i.e., 0.8 µg/mL.

[2] ULOQ: Upper Limit of Quantification, defined to be at least two times higher than the highest expected MTT formazan concentration in isopropanol extracts from negative controls i.e., 200 µg/mL.

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

23-Jul-2015

English - Or. English

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

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

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

JT03380410

Complete document available on OLIS in its original format

ENV/JM/MONO(2015)26
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 219

PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* RECONSTRUCTED HUMAN *EPIDERMIS* (RHE) TEST METHODS FOR SKIN CORROSION TESTING AS DESCRIBED IN TG 431[1]

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

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

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

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

About the OECD

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

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

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

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

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

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

or contact:

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

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

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

INTRODUCTION

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

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

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

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

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

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

ESSENTIAL TEST METHOD COMPONENTS

6. The Essential Test Method Components consist of essential structural, functional, and procedural elements of scientifically valid test methods (the VMRs) that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding VRMs (1) (2). The essential test method components to be considered for similar or modified test methods related to TG 431 are described in detail in the following paragraphs.

7. For specific parameters (*e.g.*, for Tables 1, 2, 3 and 4) or modified procedures, adequate values or procedures should be provided for the proposed similar or modified test method, these specific values or procedures may vary depending on the specific test method and/or its modification.

General Conditions

8. Non-transformed human keratinocytes should be used to reconstruct the epithelium. The RhE model is prepared in inserts with a porous synthetic membrane through which nutrients can pass to the cells. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum, stratum granulosum*) should be present under a functional *stratum corneum*. The test chemical is applied topically to the three-dimensional RhE model, which should have a surface in direct contact with air so as to allow for an exposure similar to the *in vivo* situation. The *stratum corneum* should be multi-layered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic benchmark chemicals, *e.g.* sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET_{50}) upon application of the benchmark chemical at a specified, fixed concentration (see paragraph 8). The containment properties of the RhE model should prevent the passage of test chemical around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, and fungi.

Functional Conditions

Viability

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

control. It should be documented that the tissues treated with negative control are stable in culture (provide similar OD measurements) for the duration of the exposure period.

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

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

Barrier function

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

Morphology

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

Reproducibility

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

Quality control (QC)

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

Table 2: QC batch release criteria of the VRMs

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

Procedural Conditions

Application of the Test Chemical and Control Substances

14. At least two tissue replicates should be used for each test chemical and each control substance for each exposure time in each run. For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the epidermis surface while avoiding an infinite dose (*i.e.* a minimum of 70 μL/cm^2 or 30 mg/cm^2 should be used). Whenever possible, solids should be tested as a fine powder.

15. Concurrent negative and positive controls (PC) should be used in each run to demonstrate that viability (with negative controls), and sensitivity (with the PC) of the tissues are within a defined historical acceptance range. The concurrent negative control also provides the baseline (100% tissue viability) to calculate the relative percent viability of the tissues treated with the test chemical. The positive control suggested for the VRMs are glacial acetic acid or 8N KOH depending upon the RhE model used. It should be noted that 8N KOH is a direct MTT reducer that might require adapted controls as described in paragraphs 15 and 16. The suggested VRMs negative controls are 0.9% (w/v) NaCl or water.

Cell Viability Measurements

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

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

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

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

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

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

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

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

Acceptability Criteria

22. For each run, tissues treated with the negative control should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes and should not be outside of the historically established boundaries (see paragraph 7 and table 1). Similarly, tissues treated with the positive control, should show a mean tissue viability (relative to the negative control) within an historically established range, thus reflecting the ability of the tissues to respond to a corrosive chemical under the conditions of the test method. The variability between tissue replicates of test chemicals and/or control substances should fall within the accepted limits also established from historical values (*e.g.* the difference of viability between the two tissue replicates should not exceed 30%). If either the negative control or PC included in a run fall outside of the accepted ranges, the run is considered non-qualified and should be repeated. If the variability between tissue replicates of test chemicals falls outside of the accepted range, the test chemical should be re-tested. Paragraph 33 provides more details on re-testing in case of non-qualified runs during validation studies. Importantly, an increased frequency of non-qualified runs may indicate problems with either the test system (e.g. the intrinsic RhE tissue quality) or with the handling (e.g. shipment, SOP execution). Therefore, occurrence of non-qualified runs in validation studies should be carefully monitored and all non-qualified runs need to be reported.

Interpretation of Results and Prediction Model

23. The OD values obtained for each test chemical should be used to calculate percentage of viability relative to the negative control, which is set at 100%. In case HPLC/UPLC-spectrophotometry is used, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. The cut-off value of percentage cell viability distinguishing corrosive from non-corrosive test chemical (and/or discriminating between different corrosive sub-categories), and the statistical procedure(s) used to evaluate the results should be clearly defined, documented, and proven to be appropriate. The cut-offs defined for the VRMs are defined below in paragraphs 23 and 24.

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

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

Table 3: Prediction model of the VRM EpiSkin™

Viability measured after exposure time points (t=3, 60 and 240 minutes)	Prediction to be considered
< 35% after 3 min exposure	**Corrosive:**

	• Optional Sub-category 1A
≥ 35% after 3 min exposure **AND** < 35% after 60 min exposure **OR** ≥ 35% after 60 min exposure **AND** < 35% after 240 min exposure	**Corrosive:** • A combination of optional Sub-categories 1B-and-1C
≥ 35% after 240 min exposure	Non-corrosive

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

Table 4: Prediction model of the VRM EpiDerm™ SCT

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

MINIMUM LIST OF REFERENCE CHEMICALS

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

28. The 30 Reference Chemicals listed in Table 5 represent the minimum number of chemicals that should be used to evaluate the reliability and predictive capacity of a proposed similar or modified test method able to discriminate between Subcategory 1A, a combination of Sub-categories 1B and 1C as well as non-corrosive substances and mixtures in accordance with the UN GHS (4) (1A vs. 1B-and-1C vs. NC). For similar or modified test methods able to discriminate corrosive from non-corrosive substances and mixtures but not able to support sub-categorisation of corrosive chemicals (C vs. NC), only 20 of the 30 chemicals listed in Table 5 (the ones not in *italics*) need to be evaluated: 5 Sub-category 1A chemicals, 5 chemicals of the combined Sub-categories 1B and 1C as well as 10 non-corrosive chemicals. The exclusive use of these Reference Chemicals for the development/optimization of new similar test methods should be

avoided to the extent possible. In situations where a listed Reference Chemical is unavailable, or cannot be used for other justified reasons, another chemical could be used provided it fulfils the selection criteria as described in Table 5 (foot-note 1) and adequate *in vivo* reference data are available, e.g. preferentially from the test chemicals used during the validation studies of the VRMs (6) (7) (8) (14). To gain further information on the predictive capacity of the proposed test method, additional chemicals representing other chemical classes and for which adequate *in vivo* reference data are available may be tested in addition to the minimum list of Reference Chemicals.

Table 5: Minimum list of Reference Chemicals for determination of Reproducibility and Predictive Capacity of similar or modified *in vitro* RhE-based skin corrosion test methods. The 20 chemicals NOT in *italics* should be tested with similar or modified test methods proposed to discriminate Corrosive from Non-Corrosive chemicals (without sub-categorization). Additional reference chemicals should be tested with similar or modified test methods proposed to identify Sub-category 1A, a combination of Category 1B and 1C (referred to as 1B/1C below) and non-corrosive test chemicals. These additional reference chemicals are indicated *in italics*.

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

Chemical[1]	CASRN	Chemical Class[2]	Physical State	EpiSkin™ [4]	EpiDerm™ [4]	SkinEthic™ [4]	epiCS® [4]
*2-tert-Butylphenol**	*88-18-6*	*Phenol*	*L*	*(3) 1B/1C*	*(3) 1A*	*(3) 1A*	*(2) 1A*
*Cyclohexyl amine**	*108-91-8*	*Organic base*	*L*	*(3) 1B/1C*	*(3) 1A*	*(3) 1A*	*(2) 1A*
UN GHS Sub-category 1A based on *in vivo* results[3]							
Acrylic acid	*79-10-7*	*Organic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Bromoacetic acid	79-08-3	Organic acid	S	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Boron trifluoride dehydrate	13319-75-0	Inorganic acid	L	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Phenol	108-95-2	Phenol	S	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Phosphorus tribromide	*7789-60-8*	*Inorganic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Silver nitrate	*7761-88-8*	*Inorganic salt*	*S*	*(1) 1A (2) 1B/1C*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Formic acid	*64-18-6*	*Organic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
Dichloroacetyl chloride	79-36-7	Electrophile	L	(3) 1A	(3) 1A	(3) 1A	(2) 1A
Sulphuric acid (98%)	*7664-93-9*	*Inorganic acid*	*L*	*(3) 1A*	*(3) 1A*	*(3) 1A*	*(2) 1A*
N,N-Dimethyl dipropylene triamine*	10563-29-8	Organic base	L	(3) 1B/1C	(3) 1B/1C	(3) 1B/1C	(2) 1B/1C

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; UN GHS = United Nations Globally Harmonized System (4); NC = Not Corrosive

[1]The reference chemicals, sorted first by corrosives versus non-corrosives, then by corrosive sub-category, were selected from the test chemicals used in the ECVAM validation studies of EpiSkin™ and EpiDerm™ SCT (6) (7) (14) and from post-validation studies based on data generated by EpiSkin™ (8), EpiDerm™, SkinEthic™ and epiCS® developers. Unless otherwise indicated, these chemicals were tested at the purity level obtained when purchased from a commercial source (6) (7). The selection includes, to the extent possible, chemicals that: (i) are representative of the range of corrosivity responses (*e.g.* non-corrosives; weak to strong corrosives) that the VRMs are capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation studies; (iii) reflect the performance characteristics of the VRM; (iv) have chemical structures that are well-defined; (v) induce reproducible results in the VRM; (vi) induce definitive results in the *in vivo* reference test method; (vii) are commercially available; and (viii) are not associated with prohibitive disposal costs. Chemicals marked with an * are potential direct MTT reducers.

[2]Chemical class assigned by Barratt *et al.* (14).

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

[4]The *in vitro* predictions reported in this table were obtained with the various test methods during post-validation testing performed by the test method developers. The numbers in brackets indicate, for each chemical, the number of the corresponding type of *in vitro* predictions for the test method considered. These predictions were corrected for direct MTT reduction using killed control tissues.

DEFINED RELIABILITY AND PREDICTIVE CAPACITY VALUES

29. For purposes of establishing the reliability (i.e., within- and between laboratory reproducibility) and predictive capacity (i.e., sensitivity, specificity and accuracy) of proposed similar or modified RhE test methods to be used by several independent laboratories, all 30 (or 24 for methods not able to sub-categorize corrosive chemicals) Reference Chemicals listed in Table 5 should be tested in at least three laboratories. In each laboratory, all relevant Reference Chemicals should be tested for each exposure time in three independent runs performed with different tissue batches and at sufficiently spaced time points. Each run should consist of at least two concurrently tested tissue replicates per exposure time for each test chemical, negative control, positive control and adapted controls for direct MTT reduction and/or colour interference.

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

1. Within-laboratory reproducibility (WLR) should be calculated based on concordance of classifications using only qualified tests obtained with Reference Chemicals for which at least two qualified tests are available. In addition, it should be reported the number and identity of the Reference Chemicals which per laboratory have none or only one qualified test (not considered for WLR calculations), as well as how many and which Reference Chemicals per laboratory have two or three qualified tests (used for WLR calculations).

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

3. The calculation of predictive capacity (e.g. sensitivity, specificity and accuracy for corrosive vs. non-corrosive) as well as, in case of subcategorisation, over- and under-prediction rates, should be done using all qualified tests obtained for each Reference Chemical in each laboratory. The calculations should be based on the individual predictions of each qualified test for each Reference Chemical in each laboratory and not on the arithmetic mean values of viability over the different qualified tests performed (15).

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

Within-laboratory reproducibility

31. An assessment of within-laboratory reproducibility for similar or modified test method proposed to discriminate corrosive from non-corrosive chemicals (but not to sub-categorize corrosive chemicals), should show in every laboratory, a concordance of predictions (corrosive or non-corrosive) obtained in different, independent tests of the 24 relevant Reference Chemicals equal or higher (≥) than 90% (actual for EpiSkin™: 100%, 100% and 96% in each laboratory, respectively).

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

Between-laboratory reproducibility

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

Predictive capacity

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

Table 6: Required sensitivity, specificity and accuracy for similar or modified RhE test methods to be considered valid to discriminate corrosive from non-corrosive chemicals (C vs. NC) but not able to support sub-categorisation of corrosive chemicals.

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

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

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

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

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

* Depending on the results obtained with a similar or modified RhE test method for the 30 Reference Chemicals, it may be considered similar to EpiSkin™ or similar to EpiDerm™ for the purpose of this Test Guideline. The EpiSkin™ and EpiDerm™ test methods are able to sub-categorize (i.e. 1A versus 1B-and-1C versus NC) but differences are observed (SkinEthic™ and epiCS® are considered similar to EpiDerm™). For RhE test methods that demonstrate similarity to EpiSkin™, results can be directly used based on the outcoming predictions. For RhE test methods that demonstrate similarity to EpiDerm™, chemicals that are classified as Sub-category 1B-and-1C can be considered as Sub-category 1B-and-1C, whereas chemicals for which cell viability at 3 minutes is below 50% should be considered as Category 1, since the Sub-category 1A predictions of these three test methods contain a high rate of over-predictions of chemicals of Sub-categories 1B-and-1C (see also paragraph 7 of the Test Guideline 431 (2)). The regulatory framework in member countries will decide how this Test Guideline will be used, e.g. acknowledging the significant probability of overclassification, a Sub-category 1A classification may still be accepted or further testing may be conducted to confirm the result.

Study Acceptance Criteria

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

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

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

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

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

LITERATURE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (1).

Between-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Between-laboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as inter-laboratory reproducibility (1).

C: Corrosive.

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

Chemical: means a substance or a mixture.

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

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (1).

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

GHS (Globally Harmonized System of Classification and Labelling of Chemicals): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (4).

HPLC: High Performance Liquid Chromatography.

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

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

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation (1). The term is interchangeably used with similar test method.

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

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

NC: Non corrosive.

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

NSC: Non-Specific Colour in living tissues.

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

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

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of reliability and accuracy, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (1).

Prediction Model: a formula or algorithm (*e.g.,* formula, rule or set of rules) used to convert the results generated by a test method into a prediction of the (toxic) effect of interest. Also referred to as decision criteria. A prediction model contains four elements: (i) a definition of the specific purpose(s) for which the test method is to be used; (ii) specifications of all possible results that may be obtained, (iii) an algorithm that converts each study result into a prediction of the (toxic) effect of interest, and (iv) specifications as to the accuracy of the prediction model (*e.g.,* sensitivity, specificity, and false positive and false negative rates). Prediction models are generally not used in *in vivo* ecotoxicological tests (1).

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

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

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

Reference Chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should

represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses (1).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (1).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (1).

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

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

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (1).

Skin corrosion *in vivo:* The production of irreversible damage of the skin; namely, visible necrosis through the *epidermis* and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions (5).

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (1).

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

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

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

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

Within-laboratory reproducibility: determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times, also referred to as intra-laboratory reproducibility (1).