APPLICATION OF ALTERNATIVE METHODS IN THE REGULATORY ASSESSMENT OF CHEMICAL SAFETY RELATED TO HUMAN SKIN CORROSION & IRRITATION

CURRENT STATUS AND FUTURE PROSPECTS

Updated Version

Final version

21 December 2016

Dr. Chantra Eskes

SeCAM Services & Consultation on Alternative Methods

Via Campagnora 1, 6983 Magliaso, Switzerland

Sponsored by:

Office Fédéral de la Santé Publique, Confédération Suisse

Division Produits Chimiques

Berne, Suisse

Contract n. 15.0025838 / 444.0000 / -70

LIST OF CONTENTS

LIST OF ABBREVIATIONS

A.I.S.E.	International Association for Soaps Detergents and Maintenance Products
BfR	German Federal Institute for Risk Assessment
BfR-DSS	BfR Decision Support System
Cat.	Category
C&L	Classification & Labelling
UN GHS Cat. 1	Skin Corrosive
UN GHS Cat. 2	Skin Irritant
UN GHS Cat. 3	Mild irritant to skin (optional)
CDS	Chemical Detection System
DEREK	Deductive Estimation of Risk from Existing Knowledge
DPBS	Dulbecco's Phosphate Buffered Saline
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ECVAM	European Centre for the Validation of Alternative Methods
ESAC	ECVAM Scientific Advisory Committee
ET_{50}	Time needed for a test chemical to reduce the viability of the tissues by 50 % using a fixed concentration
EU	European Union
EU CLP	EU Regulation 1272/2008 on the Classification, Labelling and Packaging of Substances and Mixtures
EU DSD	EU Dangerous Substances Directive 67/548/EEC
EU DPD	EU Dangerous Preparation Directive 199/45/EC
EURL-ECVAM	European Union Reference Laboratory for alternatives to animal testing
GCL	Generic concentration limit
GD	Guidance Document
GHS	Globally Harmonized System for Hazard Classification and Labelling of chemicals
GLP	Good Laboratory Practices
HPLC	High-performance liquid chromatography
HPT	Human Patch Test
IATA	Integrated Approaches for Testing and Assessment
IC_{50}	Concentration at which a substance reduces the viability of the tissues by 50 % after a fixed exposure time
ICCVAM	US Interagency Coordinating Committee on the Validation of Alternative Methods
ITS	Integrated Testing Strategy
JRC	Joint Research Centre
LLNA	Local Lymph Node Assay
MAD	Mutual Acceptance of Data
MTT	3-(4,5-Dimethyl-2-thiazol-2-yl)-2,5-diphenyltetrazolium bromide, EINECS number 206-069-5, CAS number 298-93-1
NIH	National Institutes of Health
NSC	Non-Specific Colour
NSMTT	Non-Specific MTT reduction
OD	Optical Density
OECD	Organisation for Economic Co-operation and Development
(Q)SAR	(Quantitative) Structure-Activity Relationship
PBS	Phosphate Buffered Saline
PS	Performance Standards
PTFE	Polytetrafluoroethylene
R34	Causes burns
R35	Causes severe burns
R38	Irritating to skin
REACH	EU Regulation 1907/2006 on the Registration, Evaluation, Authorisation and restriction of Chemicals
RH	Relative humidity
RhE	Reconstructed human Epidermis
RT	Room Temperature
SCL	Specific concentration limit
SCT	Skin Corrosion Test
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SICRET	Skin Irritation Corrosion Rules Estimation Tool
SIT	Skin Irritation Test
SM	Standard Model
SOP	Standard Operating Procedure
TER	Transcutaneous Electrical Resistance
TG	Test Guideline
TOPKAT	TOxicity Prediction by Komputer Assisted Technology
UN	United Nations
UPLC	Ultra Performance Liquid Chromatography
US	United States
US DOT	US Department of Transportation
UVCB	Unknown or variable composition, complex reaction products or biological materials
VRM	Validated Reference Method
WoE	Weight of evidence

1 Regulatory Background

1.1. Mechanisms and definitions of skin irritation and corrosion

The human skin is divided in three distinct regions: the epidermis as the outer region, the dermis and the deeper localized subcutis. The epidermis represents 5% of the full thickness of the skin, and is subdivided into 5 to 6 layers based on cellular characteristics (see figure 1.1). The outer layer represents the *stratum corneum*, whereas the inner layer represents the *stratum basale*, subdivided into the basal layer (outer part) and the basal lamina (inner part).

The epidermis layers are formed by keratinocytes having ordered differentiation of cells from the basal layer keratinocytes, which are metabolically active and have the capacity to divide. Some daughter cells of the basal layer move upward and differentiate. The outermost layer, the *stratum corneum* consists of cornified keratinocytes that have elongated and flattened with respect to the basal layer keratinocytes, and have lost their nucleus and all capacity for metabolic activity. The dominant constituent of these cells is keratin. In addition to keratinocytes, the epidermis contains two dendritic cell types, melanocytes and Langerhans cells. Melanocytes produce melanin, the principal pigment of human skin, whereas Langherans cells express Ia (immune recognition) antigen and receptors of IgG and C3 on their surface (for review see Patrick and Maibach, 1994).

The *stratum corneum* represents an effective barrier against a vast number of substances. Apart this, keratinocytes play crucial roles in the immune surveillance of the epidermis, as after stimulation they can trigger inflammatory responses (for review see Welss *et al.*, 2004).

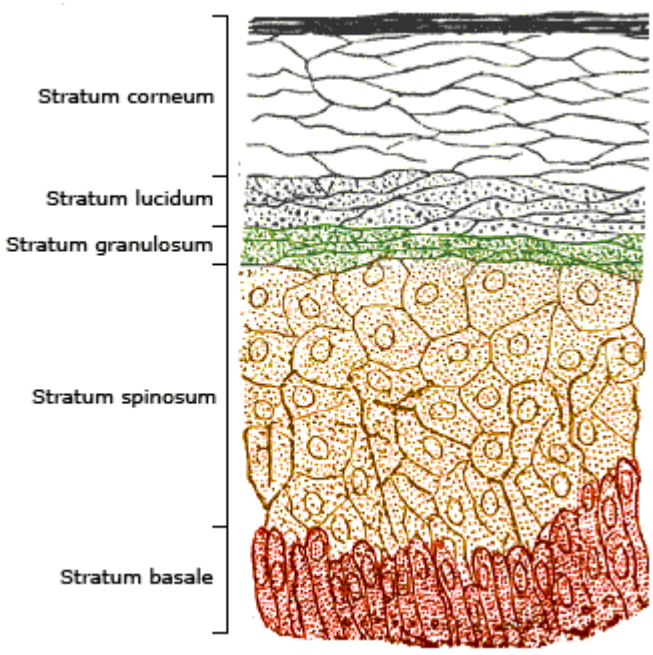

Figure 1.1. The epidermis cell layers

If skin corrosion assesses the potential of a substance to cause visible necrosis through the epidermis and into the dermis that may lead to irreversible damage to skin, acute skin irritation is characterised by the local and reversible non-immunological inflammatory response of the living skin.

Chemical-induced skin irritation manifested by erythema and oedema, is the result of a cascade of events beginning with penetration of the *stratum corneum* and damage to the underlying layers of keratinocytes. The dying keratinocytes release mediators that begin the inflammatory cascade which acts on the cells in the dermis, particularly the stromal and endothelial cells, It is the dilation and increased permeability of the endothelial cells that produce the observed erythema and oedema.

However, the underlying mechanisms of skin irritation, linked to the molecular and cellular responses, are still poorly understood. Probably different pathways may be involved, such as damage to the barrier function of the *stratum corneum*, and the direct effects of irritants on cells of the skin (for review see Welss *et al.*, 2004).

Within the regulatory context, skin corrosion is generally defined, based on the *in vivo* animal test, as "*the production of irreversible damage of the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test substance for up to 4 hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discolouration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions*." (UN, 2015a; OECD, 2014; EC, 2016a).

Furthermore, skin irritation is generally defined, based on the *in vivo* animal test, as "*the production of reversible damage of the skin following the application of a test substance for up to 4 hours*" (UN, 2015a; OECD, 2014; EC, 2016a).

1.2. Regulatory requirements for alternative test methods

At the European Union (EU) level, several legislations promote the use of alternative methods to animal toxicological testing such as the animal welfare regulation, the chemicals regulation (REACH), the classification, labelling and packaging (EU CLP) regulation and the cosmetics regulation. The EU Directive 86/689 updated as Directive 2010/63 on the protection of animals used for scientific purposes promotes the use of alternative methods. Article 1 states that the directive aims, among others, to lay down rules for "*the replacement and reduction of the use of animals in procedures and the refinement of the breeding, accommodation, care and use of animals in procedures*", where procedure means "*any use, invasive or non-invasive, of an animal for experimental or other scientific purposes (…) which may cause the animal a level of pain, suffering, distress or lasting harm equivalent to, or higher than, that caused by the introduction of a needle in accordance with good veterinary practice*". In addition, it states in Article 4(1) that "*Member States shall ensure that, wherever possible, a scientifically satisfactory method or testing strategy, not entailing the use of a live animal, shall be used instead of a procedure*" and in Article 4(2) that "*Member States shall ensure that the number of animals used in projects is reduced to a minimum without compromising the objectives of the project*". Finally, Article 13(1) states that "*Member States shall ensure that a procedure is not carried out if another method or testing strategy for obtaining the result sought, not entailing the use of a live animal, is recognised under the legislation of the Union*" (EC, 2010).

Within the regulation 1907/2006 for the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), only *in vitro* testing shall be conducted for skin corrosion and irritation as standard information for substances manufactured or imported in quantities of 1 to 10 tonnes per year (Annex VII, EC, 2006). Furthermore, for substances manufactured or imported in quantities of 10 tonnes per year or more, an *in vivo* study for skin corrosion/irritation shall be considered only if the *in vitro* studies above mentioned are not applicable, or the results of these studies are not adequate for classification and risk assessment (Annex VIII, EC, 2016b). This is in accordance to Article 25(1) of the REACH regulation, which states that "animal *testing on vertebrate animals shall be undertaken only as a last resort*". Article 13(1) further notes, that information on hazards (regarding positive results) and risks may also be generated by suitable alternative methods that have not yet been taken up as official regulatory test methods, upon the condition that such methods fulfil the requirements of Annex XI (e.g., ECVAM criteria for the entry of a test into the prevalidation process). If such methods are moreover validated, they can be used for identifying positives as well as negative results according to Annex XI (see chapter 1.3).

The 7th amendment to the EU Cosmetics Directive (Directive 2003/15/EC) later incorporated into the EU Cosmetics Regulation (EC, 2009a), went further and prohibited animal testing of finished products since 2004 and of ingredients since 2009. The animal testing ban was reinforced by a marketing ban of cosmetics tested on animals that entered into force since 2004 for the finished products and since 2013 for cosmetics containing ingredients tested on animals (EC, 2003, 2009a).

Finally, the EU regulation on classification, labelling and packaging of substances and mixtures (EU CLP; EC, 2008a, 2016a), which implements the Globally Harmonized System (GHS) for classification and labelling of substances and mixtures in the European Union, also encourages the use of tiered weight-of-evidence evaluations, and makes use of information from *in vitro* testing in its tiered classification approach for skin corrosion and irritation. Article 8(1) requests that before making use of a new test, the rules of section 1 of Annex XI to the REACH Regulation (1907/2006; EC, 2006) are taken into account, which relates to the use of existing data (physical-chemical properties and human data), weight-of-evidence, (Quantitative) Structure-Activity Relationship ((Q)SAR, *in vitro* methods, grouping and read-across approaches. Furthermore Article 8(3) of the EU CLP Regulation (EC, 2008a) states the possibility of conduct testing using "*sound scientific principles that are internationally recognised or methods validated according to international procedures*." Finally, the EU CLP regulation encourages the use of expert judgement and weight-of-evidence approaches (Annex I, part I section 1.1.1.3), including "*the results of suitable* in vitro *tests, relevant animal data, information from the application of the category approach (grouping, read-across), (Q)SAR results, human experience such as occupational data and data from accident databases, epidemiological and clinical studies and well documented case reports and observations"*. Furthermore it states that "*the quality and consistency of the data shall be given appropriate weight. Information on substances or mixtures related to the substance or mixture being classified shall be considered as appropriate, as well as site of action and mechanism or mode of action study results. Both positive and negative results shall be assembled together in a single weight of evidence determination*."

As a consequence, there is a strong regulatory requirement for *in vitro* alternative methods within the European Union. For this reason, the European Chemicals Agency has issued a guidance for the evaluation of available information for REACH (ECHA, 2011), providing further insights on how data derived from *in vitro* studies can be used:

1) information from validated *in vitro* tests: may fully or partly replace an *in vivo* test depending on the purpose for which the test method was validated and adopted. In that case, one of the criteria for acceptance is the adequacy of the information generated using such test(s) for the purpose of classification and labelling and/or risk assessment,
2) information derived from suitable *in vitro* methods: can be used for determining the presence of a certain dangerous property, adapting the standard testing regime as set out in annex XI of the REACH regulation (EC, 2006).
3) Information from *in vitro* tests may be also used to provide mechanistic insights (ECHA, 2011) and aid and inform the risk assessment process (e.g., toxicogenomics, molecular pathways/mode of action).

As such, the scientific validation of *in vitro* methods certify their level of relevance and reliability to be used in the regulatory framework for detecting both positive and negative results, as full replacement or partial replacement of the animal testing.

The area of skin corrosion and irritation represents one of the pioneering areas for the validation of alternative test methods, in which replacement alternatives have been validated according to internationally agreed principles (OECD, 2005) and adopted by the Organisation for Economic Co-operation and Development (OECD) since 2004 for skin corrosion and since 2010 for skin irritation. As a consequence they fall under the OECD international Mutual Acceptance of Data (MAD), in which test data generated in any OECD member country in accordance with these OECD Test Guidelines and following the Principles of Good Laboratory Practice (GLP) should be accepted in other OECD member countries for assessment purposes and other uses relating to the protection of human health and the environment.

The present document describes how to derive skin corrosion and irritation hazard identification from the currently recommended approaches and adopted *in vitro* methods for regulatory skin corrosion and skin irritation testing, which minimizes to the extent possible the need for animal testing whilst ensuring human safety.

1.3. Regulatory requirements for skin corrosion and irritation testing

In the EU, it is the chemicals policy 1907/2006 adopted in 2006 for the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) that establishes standard information requirements that need be submitted for the registration and evaluation of new and existing chemicals. Such information requirements are specified in details in the REACH Annexes VI to XI (EC, 2006). According to Annex VI, the registrant should gather and evaluate all available information before considering further testing. These include physico-chemical properties, (Q)SAR), grouping, *in vitro* data, animal studies, and human data (for details see chapter 2). Information on exposure, use and risk management measures should also be collected and evaluated. If these data are inadequate for hazard and risk assessment, further testing should be carried out in accordance with the requirements of REACH Annexes VII to X, which are based on the tonnage levels of the manufactured or imported chemicals.

The standard toxicological information requirements for substances manufactured or imported in quantities between one tonne and 10 tonnes per year are laid down in Annex VII of the REACH regulation (EC, 2006, 2016b). If new testing data on skin irritation and corrosion are necessary, these must be derived from *in vitro* methods only. Annex VII does not foresee *in vivo* testing for skin corrosion and irritation. The standard information required at this tonnage level for skin corrosion and irritation can be satisfied by following four consecutive steps:

1) *in vitro* study for skin corrosion,
2) *in vitro* study for skin irritation.

Specific rules for adaptation state that the study/ies do(es) not need to be conducted if:

- the substance is a strong acid (pH ≤ 2,0) or base (pH ≥ 11,5) and the available information indicates that it should be classified as skin corrosion (Category 1), or
- the substance is spontaneously flammable in air or in contact with water or moisture at room temperature, or
- the substance is classified as acute toxicity by the dermal route (Category 1), or
- an acute toxicity study by the dermal route does not indicate skin irritation up to the limit dose level (2 000 mg/kg body weight).

Furthermore it states that if results from one of the two *in vitro* studies for skin corrosion and/or irritation already allow a conclusive decision on the classification of a substance or on the absence of skin irritation potential, the second study need not to be conducted.

For substances manufactured or imported in quantities of ≥ 10 tonnes per year, the toxicological information requirements are laid down in Annex VIII of the REACH regulation, which has been recently revised (EC, 2016b). Such information is additional to that required in annex VII, and specific rules for adaptation state that:

"*An* in vivo *study for skin corrosion/irritation shall be considered only if the* in vitro *studies (…) in Annex VII are not applicable, or the results of these studies are not adequate for classification and risk assessment"*.

Moreover, it states that the study does not need to be conducted if:

- the substance is a strong acid (pH ≤ 2,0) or base (pH ≥ 11,5), or
- the substance is spontaneously flammable in air or in contact with water or moisture at room temperature, or
- the substance is classified as acute toxicity by the dermal route (Category 1), or
- an acute toxicity study by the dermal route does not indicate skin irritation up to the limit dose level (2 000 mg/kg body weight).

Annex VI of the REACH regulation also states that "*new tests on vertebrates shall only be conducted or proposed as a last resort when all other data sources have been exhausted*". In particular it states that the *in vivo* testing requirement of Annex VIII can be adapted by the rules laid down in Annex XI allowing to avoid unnecessary animal testing. Annex XI establishes amongst others, the conditions in which the standard testing may not be scientifically necessary. *In vitro* test methods fall within this category. Annex XI states that "*results obtained from a suitable* in vitro *method*" may be used to "*indicate the presence of a certain dangerous property, or may be important in relation to a*

mechanistic understanding which may be important for the assessment". "*Suitable*" *in vitro* methods "*means sufficiently well developed according to internationally agreed test development criteria (e.g., criteria from the European Centre for the Validation of Alternative Methods (ECVAM) for the entry of a test into the pre-validation process)*". However, depending on the potential risk, immediate (or proposed) confirmation may be necessary requiring tests beyond the information foreseen in Annexes VII or VIII (or Annexes IX or X according to the respective the tonnage levels). In addition, it states that "*if the results obtained from the use of such* in vitro *methods do not indicate a certain dangerous properties*", a confirmatory test according to annex VII to X "*may be waived if the following conditions are met:*

1. *results are derived from an* in vitro *method whose scientific validity has been established by a validation study, according to internationally agreed principles*
2. *results are adequate for the purpose of classification and labelling and/or risk assessment, and,*
3. *adequate and reliable documentation of the applied method is provided*."

In order to apply the information testing requirements as laid down in REACH, the European Chemicals Agency (ECHA) has issued an Endpoint Specific Guidance on the REACH Information Requirements and Chemical Safety Assessment (ECHA, 2015a). For skin corrosion and irritation, a testing and assessment strategy is proposed to be followed, including a step-wise approach that takes into consideration information from the physico-chemical properties of the test substance, existing human data, existing animal data, (Q)SAR and read-across, existing *in vitro* data on the test chemical, a weight-of-evidence evaluation, the generation of new *in vitro* data and only as a last resort, the generation of new *in vivo* testing. This approach is based on the OECD Guidance Document 203 on Integrated Approaches for Testing and Assessment (IATA) for skin corrosion and irritation (OECD, 2014). Chapter 2 describes in more details the purposes and uses of the currently recommended testing approaches.

1.4. Classification for Skin Corrosion & Irritation

In principle, the *in vivo* testing for skin corrosion shall no longer be used or only in specific cases as a last resort, due to the use of integrated approaches to testing and assessment (see chapter 2) and the availability of replacement *in vitro* tests for skin corrosion and skin irritation (see chapter 3). However, as the animal *in vivo* study has been traditionally used to classify for potential skin corrosion and skin irritation hazard effects, and that validation studies on alternative methods for skin irritation and corrosion testing have used the animal test as the reference test method, the *in vivo* observations used for such classification according to the UN GHS and EU CLP criteria are described here below. The Swiss Ordinance on protection against dangerous substances and preparations n. 813.11 (Swiss Federal Council, 2015) when introducing the GHS classification system according to UN (2015a) also recalls the EU CLP Regulation (EC, 2016a), which are both (the UN GHS and EU CLP) explained here after.

1.4.1. The UN Globally Harmonised System (GHS) for classification & labelling

The UN has published in 2003 the Globally Harmonized System for classification and labelling to favour the harmonized classification of hazards across the world, which is now in its 6th revision (UN, 2015a). Although the main criteria of this classification system are still based on the traditional *in vivo* animal test method adopted by the OECD (OECD TG 404, 2015a) originally developed by Draize (1944), the UN GHS additionally contains a tiered evaluation approach that should be considered for the final classification. The tiered evaluation approach includes the use of *in vitro* data as well as other information sources such as existing human and animal data, physico-chemical properties and weight of evidence evaluation (see chapter 2).

Skin Corrosion

A substance is considered to be corrosive to skin in the *in vivo* animal test when it produces destruction of skin tissue (i.e., visible necrosis through the epidermis and into the dermis) in at least one tested animal after exposure for up to 4 hours. Corrosive substances should be classified in

Category 1 where sub-categorization is not required by a competent authority or where data are not sufficient for sub-categorization. When data are sufficient and where required by a competent authority substances may be classified in one of the three Sub-categories (see table 1.1): Sub-category 1A, where responses are noted following up to 3 minutes exposure and up to 1 hour observation; Sub-category 1B, where responses are described following exposure between 3 minutes and 1 hour and observations up to 14 days; and Sub-category 1C, where responses occur after exposures between 1 hour and 4 hours and observations up to 14 days.

Table 1.1. UN GHS skin corrosive category and Sub-categories based on *in vivo* animal test data

Authorities not requiring Sub-categories or data not sufficient for sub-categorization	Sufficient data available and required by a competent authority	*In vivo* corrosive in at least one tested animal	
		Exposure	Observation
Category 1	Sub-category 1A	≤ 3 minutes	≤ 1 hour
	Sub-category 1B	> 3 minutes, ≤ 1 hour	≤ 14 days
	Sub-category 1C	> 1 hour, ≤ 4 hours	≤ 14 days

In case of evaluations based on 4 to 6 animal study, the *in vivo* skin corrosion Category 1 should be used if destruction of skin tissue occurs in at least one animal (out of up to 6 animals) after exposure up to 4 hours in duration.

Skin Irritation

One main irritant category is defined by the UN GHS classification system, i.e., Category 2 (see table 1.2). However, an additional optional category for mild irritants (i.e., Category 3) is also defined for those authorities wanting to have more than one skin irritant category.

The major criterion for the irritant category based on the *in vivo* animal test is that at least 2 out of 3 tested animals have a mean score for either erythema/eschar or oedema of ≥ 2.3 and ≤ 4.0 (see table 1.2). For the optional mild irritant category, the *in vivo* mean score cut-off values are ≥ 1.5 and < 2.3 for at least 2 out of 3 tested animals. Test chemicals in the irritation category are excluded from the mild irritation category.

In addition to severity of effects, reversibility of skin lesions is another consideration when evaluating *in vivo* irritant responses. When inflammation persists to the end of the observation period in 2 or more test animals, taking into consideration alopecia (limited area), hyperkeratosis, hyperplasia and scaling, then a test chemical shall be considered to be an *in vivo* irritant.

The *in vivo* animal irritant responses within a test can be quite variable, as they are with corrosion. Therefore, a separate irritant *in vivo* criterion accommodates cases when there is a significant irritant response but less than the *in vivo* mean score criterion for a positive test. For example, a test chemical might be designated as an irritant *in vivo* if at least 1 out of 3 tested animals shows a very elevated mean score throughout the study, including lesions persisting at the end of an observation period of normally 14 days. Other responses could also fulfil this criterion. However, it should be ascertained that the responses are the result of chemical exposure. Addition to this criterion increases the sensitivity of the classification system.

In case of evaluations based on a 4 to 6 animal study, the following *in vivo* criteria should be applied:

- Skin irritation Category 2 should be used if at least 3 out of 4 animals, 3 out of 5 animals, or 4 out of 6 animals, show a mean score per animal of ≥ 2.3 and ≤ 4.0 for either erythema/eschar or for oedema;
- Skin irritation Category 3 should be used if at least 3 out of 4 animals, 3 out of 5 animals, or 4 out of 6 animals, show a mean score per animal of ≥ 1.5 and < 2.3 for either erythema/eschar or for oedema.

Table 1.2. UN GHS skin irritation categories based on *in vivo* animal test data

Categories	***In vivo* criteria**[a]
Irritant Category 2	(1) Mean value of ≥ 2,3 ≤ 4,0 for erythema/eschar or for oedema in at least 2 of 3 tested animals from gradings at 24, 48 and 72 hours after patch removal or, if reactions are delayed, from grades on 3 consecutive days after the onset of skin reactions; or (2) Inflammation that persists to the end of the observation period normally 14 days in at least 2 animals, particularly taking into account alopecia (limited area), hyperkeratosis, hyperplasia, and scaling; or (3) In some cases where there is pronounced variability of response among animals, with very definite positive effects related to chemical exposure in a single animal but less than the criteria above.
Optional mild irritant Category 3	Mean value of ≥ 1.5 < 2.3 for erythema/eschar or for oedema from gradings in at least 2 of 3 tested animals from grades at 24, 48 and 72 hours or, if reactions are delayed, from grades on 3 consecutive days after the onset of skin reactions (when not included in the irritant category above).

[a] Grading criteria are understood as described in the OECD Test Guideline 404 (OECD, 2015a)

1.4.2. The EU CLP classification system

The EU CLP regulation (1272/2008; EC, 2008a) implements in the European Union the UN GHS classification and labelling system. This EU CLP classification system replaces since December 2010 the former EU Dangerous Substances Directive 67/548/EEC establishing criteria for classification of substances (EU DSD; EC, 2001), and since 2015 the EU Dangerous Preparation Directive 1999/45/EC establishing criteria for classification of mixtures (EU DPD; EC, 1999). The latest EU CLP criteria for the classification and labelling of *in vivo* skin corrosion and irritant effects are described here after (EC, 2016a).

Skin corrosion

The EU CLP is equivalent to the UN GHS, and requires the use of the three Sub-categories within the corrosive category (1A, 1B and 1C).

Skin irritation

The EU CLP is equivalent to the UN GHS, but makes use of a single category for skin irritation (Category 2) only, whereas the mild irritant category 3 is not required. Substances falling in the UN GHS category 3, require No Category classification under the EU CLP.

1.4.3. Comparison of classification systems

In addition to the UN GHS and EU CLP classification systems, skin corrosives can also be classified for transport purposes according to the UN model regulations for the transport of dangerous goods (UN, 2015b), which is based on three packaging groups (PG I, II, III). As with the UN GHS, the main criteria of this classification system are still based on the traditional *in vivo* animal test adopted by the OECD (OECD TG 404, 2015a), although it recommends the use of other information sources such as human experience from accidental exposure and the use of *in vitro* test methods. Table 1.3 provides with an overview of the UN transport packaging groups as compared to the UN GHS / EU CLP classification systems, as well as with the former EU classification systems for dangerous substances based on the traditional *in vivo* animal test method (UN, 2015a,b; EC, 2001, 2016a).

Table 1.3. Corrosion classification based on *in vivo* animal test data according to the UN GHS/EU CLP, the UN transportation and the EU DSD classification systems.

UN GHS[(a)] **& EU CLP**[(b)]		**UN transport class 8 packaging group** [(c)]	*In vivo* corrosive in at least one tested animal	
			Exposure	Observation
Category 1	**Sub-cat. 1A**	**I**	≤ 3 minutes	≤ 1 hour
	Sub-cat. 1B	**II**	> 3 minutes, ≤ 1 hour	≤ 14 days
	Sub-cat. 1C	**III** [(d)]	> 1 hour, ≤ 4 hours	≤ 14 days

[(a)] Recommends considering the use of a tiered approach for the evaluation of initial information and classification where applicable, that includes existing human or animal data, existing *in vitro* data, pH-based assessment, SAR methods and a weight-of-evidence approach (UN, 2015a).

[(b)] Recommends the use of a testing and assessment strategy for classification of substances based on physico-chemical properties (organic (hydro) peroxide, pH), existing human and animal data, read-across or (Q)SAR methods and results from OECD adopted, validated or suitable *in vitro* test methods (ECHA, 2015a).

[(c)] In the UN model regulations for the transport of dangerous goods allocation of substances listed in the Dangerous Goods List to the packing groups in Class 8 has been made on the basis of experience taking into account such additional factors as inhalation risk and reactivity with water (including the formation of dangerous decomposition products). In the absence of an entry in the dangerous goods list, it is recommended to take into account human experience from accidental exposure. In absence of human experience, grouping shall be based on data obtained from experiments in accordance with OECD TG 404 or 435. A substance determined not to be corrosive according to OECD TG 430 or 431 may be considered not to be corrosive to skin for the purpose of the UN transport regulation without further testing (UN, 2015b).

[(d)] An additional alternative criterion for PG III is metal corrosion: corrosion rate on either steel or aluminium surfaces exceeding 6.25 mm a year at a test temperature of 55°C when tested on both materials (UN, 2015b).

Regarding skin irritation, figure 1.2 provides with a comparative overview of the *in vivo* animal criteria applied in the UN GHS, EU CLP and EU DSD classification systems (UN, 2015; EC, 2001, 2016a).

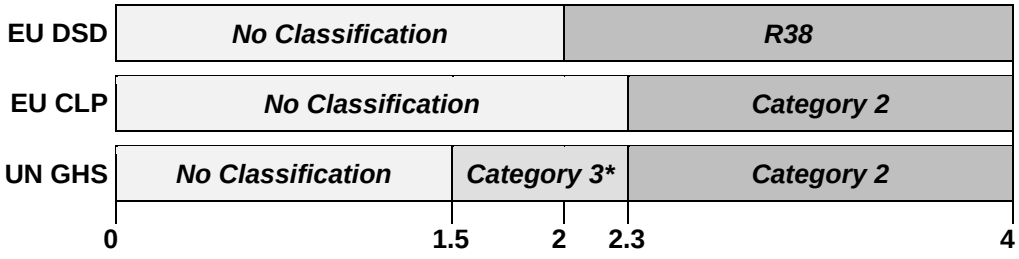

Figure 1.2: Erythema/oedema Draize score criteria used in the EU DSD, EU CLP and UN GHS classification for skin irritation hazard. Scores refer to the mean value from gradings at 24, 48 and 72 hours observed in at least two out of three animals (or in case of more than 3 animals: as described in 14.41 for UN GHS & EU CLP, and as mean value over all tested animals for the EU DSD). * Category 3 is an optional category available for those authorities wanting to have more than one skin irritant category.

It is important to note that the three classification systems also consider a substance irritant based on *in vivo* animal test data if effects persist at the end of the observation period (d14) in two or more test animals, and other effects such as hyperplasia, scaling, discoloration, fissures, scabs and alopecia are also taken into account.

Furthermore, the EU DSD and the ECHA guidance on the application of the EU CLP criteria consider organic peroxides as skin irritants and organic hydroperoxides as skin corrosives, except where evidence to the contrary is available (EC, 2001; ECHA, 2015b). Finally, the EU CLP & UN GHS may make use in some cases, where there is pronounced variability of the *in vivo* animal test data, of a separate irritant criterion when there is a significant irritant response but less than the *in vivo* mean score criterion for a positive test.

2 Integrated Approaches for Testing and Assessment (IATA)

Current internationally agreed approaches (OECD, EU and UN) recommend the use of integrated approaches and strategies for the assessment of skin irritation and corrosion effects. If the principles of these approaches are similar, small differences exist in their application. The present chapter provides with an overview of the integrated approach for testing and assessment (IATA) agreed by the OECD member countries (OECD GD 203, 2014), the testing and assessment strategy recommended within the EU (ECHA, 2015a,b), and the UN GHS tiered evaluation approach for initial information (UN, 2015a).

2.1. The OECD GD 203 on an IATA for skin corrosion and irritation

The OECD published in 2014 the first Guidance Document (GD No. 203) on an IATA adopted at an international level for skin corrosion and irritation (OECD, 2014). The IATA aims at hazard identification of the skin corrosion or irritation potential of chemicals (or the absence thereof) and to provide adequate information for classification and labelling according to the UN GHS classification system. It also aims at minimizing the use of animals to the extent possible, whilst ensuring human safety. The IATA comprises three main parts arranged in a sequential way including: Part 1 on the use of existing information, physico-chemical properties and non-testing methods; Part 2 on a weigh-of-evidence evaluation of the existing data, and Part 3, if needed, on the conduct of prospective testing.

2.1.1. Overview of information sources available

The possible individual information sources integrating the IATA have been grouped in eight "Modules" according to the type of information provided, each of which containing one to several individual information sources of similar type. These modules were subsumed in the three major parts of the IATA as described in table 2.1.

Table 2.1. Parts and modules of the IATA for skin corrosion and irritation

Part (*)	**Module**	**Data**
Part 1 (existing information, physico-chemical properties and non-testing methods)	1	- Existing human data a) Non-standardised human data on local skin effects b) Human Patch Test (HPT)
	2	- *In vivo* skin irritation and corrosion data (OECD TG 404)
	3	- *In vitro* skin corrosion data a) OECD TG 430 b) OECD TG 431 c) OECD TG 435
	4	- *In vitro* skin irritation data (OECD TG 439)
	5	- Other *in vivo* and *in vitro* data a) *In vitro* skin corrosion or irritation data from test methods not adopted by the OECD b) Other *in vivo* and *in vitro* dermal toxicity data
	6	Physico-chemical properties (existing, measured or estimated) - e.g., pH, acid/alkaline reserve
	7	Non-testing methods - for substances: (Q)SAR, read-across, grouping and prediction systems; - for mixtures: bridging principles and theory of additivity
Part 2 (WoE analysis)	8	Phases and elements of Weight of evidence (WoE) approaches
Part 3 (Additional testing)	(5b)	*Other in vivo* and/or *in vitro* dermal toxicity testing (if required by other regulations)
	(3)	*In vitro* skin corrosion testing
	(4)	*In vitro* skin irritation testing
	(5a)	*In vitro* skin irritation testing in test method not adopted by the OECD
	(2)	*In vivo* skin irritation and corrosion testing

(*) While the three Parts are considered as a sequence, the order of Modules 1 to 7 of Part 1 might be arranged as appropriate.

The OECD GD 203 also provides an overview on the strengths and limitations as well as the potential role and contribution of each Module and their individual components in the IATA for skin irritation and corrosion. Modules 3 and 4 on the use of *in vitro* test methods will be described in details in chapter 3, whereas all other modules will be described in chapter 4 of the present document.

2.1.2. The IATA framework

A schematic outline of the IATA for skin corrosion and irritation classification and labelling is presented in figure 2.1. While the three Parts are considered as a sequence, Modules 1 to 7 of Part 1 might be arranged as appropriate. Ideally, the IATA should be universally applicable to ensure human safety, whilst making maximum use of existing data, being resource efficient and minimising or eliminating the requirement for animal experiments.

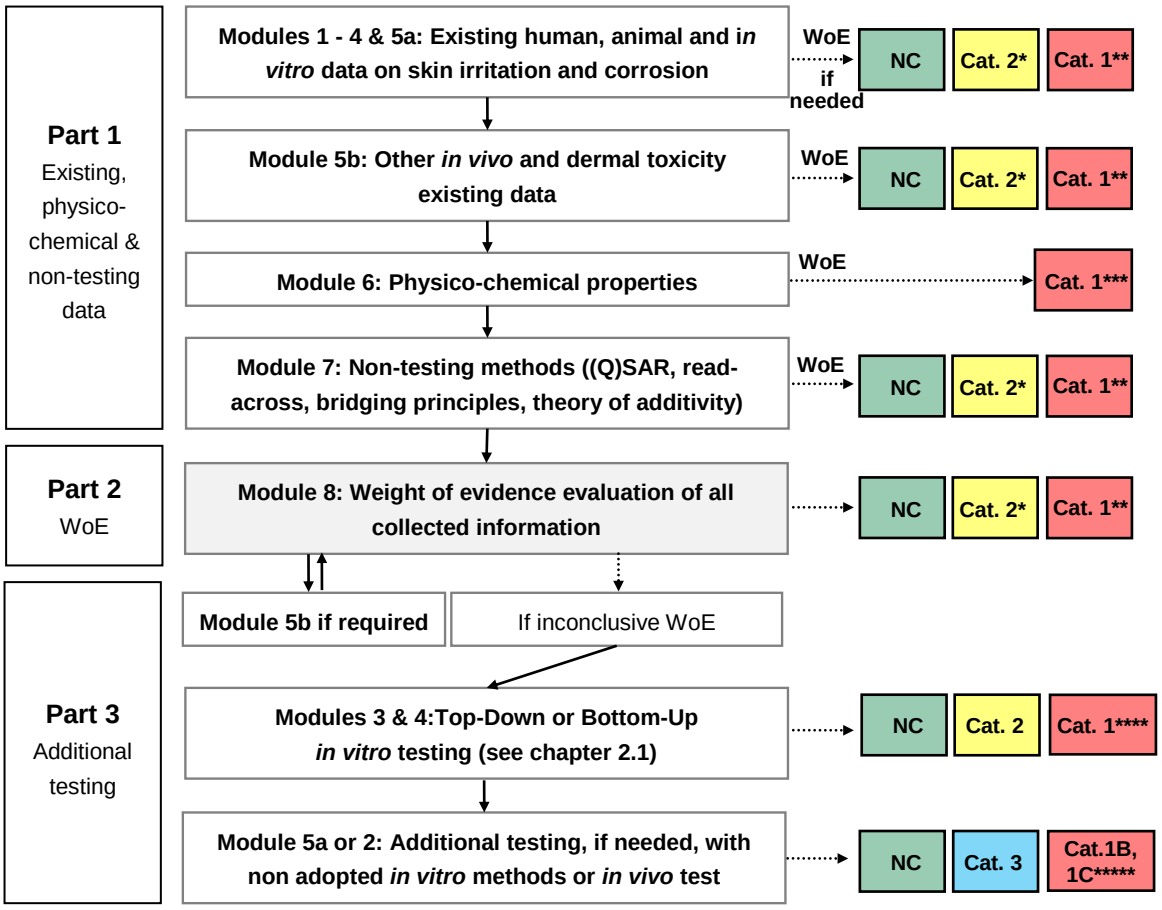

Figure 2.1. Schematic overview of the IATA for skin irritation and corrosion based on the recommendations from the OECD GD 203 (2014). Cat. 1: Corrosive to skin; Cat. 2: Irritating to skin; NC: No Category.

* Including optional Cat. 3, as applicable.

** Including corrosive Sub-categories 1A, 1B and 1C, as applicable.

*** If corrosive sub-categorisation is required an appropriate *in vitro* skin corrosion test needs to be conducted.

**** Possibilities to sub-categorise depends on the specific test method used:
OECD TG 435 allows for the discrimination between Sub-cat. 1A, 1B and 1C but with a limited applicability domain;
OECD TG 431 allows for the discrimination between Sub-cat. 1A and other corrosives – with a variable rate of over-classification into Sub-cat.1A depending on the test methods- but does not permit the sub-categorisation of the latter into Sub-cat. 1B and Sub-cat. 1C.
OECD TG 430 only allows the identification of corrosives into a single category without sub-categorisation, i.e., Cat. 1.

***** If outside of the applicability domain of OECD TG 435.

Under **Part 1** of the IATA **(existing, physico-chemical & non-testing data)**, existing and available information is retrieved from literature and databases and other reliable sources for Modules 1 to 5, while under Module 6 on physico-chemical properties, primarily the pH and the acidic/alkaline reserve are considered, and under Module 7 non-testing methods are considered. Whilst the retrieval of

existing information for Modules 1 to 5a directly relate to skin corrosion and irritation, Module 5b requires a different search for other *in vitro* and *in vivo* dermal toxicity studies.

The collected information from Part 1 is then evaluated in a weight of evidence (WoE) approach in **Part 2 (WoE)**. While a WoE approach implies the weighing of each available piece of information on a case by case basis, the modules included in this IATA differ a priori with respect to their intrinsic weight e.g. based on considerations of relevance relating to the species of interest or biological and mechanistic aspects. Typically, the relative a priori weights of the modules can be expected to be as follows, based on regulatory acceptance of data when it is of equal quality (note that the following relative a priori weights are indicative only and will depend on the quality of the individual data in each specific case):

- Reliable existing human data (in particular HPT data - Module 1b) would be expected to carry the highest weight;
- Followed by, with equal weights, *in vivo* rabbit skin corrosion/irritation data (Module 2) and *in vitro* skin corrosion or irritation data (Modules 3 & 4);
- Non-testing methods (Module 7), non-standard *in vivo* or *in vitro* and other dermal toxicity data (Module 5) and physico-chemical information (Module 6) would typically carry less intrinsic weight.

If the WoE is conclusive, decision for C&L can be conducted accordingly. However, if the WoE evaluation is inconclusive regarding the skin irritation and corrosion potential, other *in vivo* or *in vitro* dermal toxicity tests (Module 5b) for which data are still not available but may be needed to satisfy other regulatory requirements, shall be conducted. Once available, these additional test results should be incorporated into a new WoE analysis. If the WoE is still inconclusive or no other *in vivo* or *in vitro* dermal toxicity tests need to be conducted, all available information from the WoE should then be considered to formulate a hypothesis of the most likely skin corrosion or skin irritation potential of the test chemical.

This hypothesis will then guide the sequence of *in vitro* prospective testing of **Part 3 (additional testing)** in either a top-down or a bottom-up approach (see chapter 3.1 for details). The top-down approach should be used when available information suggests that the substance has a high likelihood of being irritant or corrosive to the skin, starting with an *in vitro* method for identification of skin corrosion (Module 3) eventually followed by an *in vitro* method for identification of skin irritation (Module 4). On the other hand, the bottom-up approach should be followed only when available information suggests that the substance has a high likelihood to not be irritant to the skin, starting with an *in vitro* method for identification of skin irritation (Module 4), followed eventually by an *in vitro* method for identification of skin corrosion (Module 3). Details on the top-down and bottom-up approaches, and on the purposes, applicability and limitations of the *in vitro* tests are described in chapter 3.

If additional testing is still required to satisfy specific requirements, the Guidance Document suggests that other *in vitro* skin irritation or corrosion test methods not yet adopted by the OECD are used that may resolve specific optional- or sub- categorisation issues (e.g., Cat. 3 for mild irritancy or resolving between Sub-categories 1B and 1C in case the test chemical is outside of the applicability domain of OECD TG 435). Animal testing should be used only as a last resort when i) discrimination between optional Sub-categories 1B and 1C for chemicals outside of the applicability domain of OECD TG 435 is required, (ii) discrimination of optional Cat. 3 from No Cat. is required, or (iii) the test chemical cannot be tested with the *in vitro* test methods currently adopted by the OECD due to limitations or non-applicability.

The IATA is considered applicable to both substances and mixtures, although it is acknowledged that there is a different amount of information available on the applicability of the modules of this IATA to mixtures (see chapter 3.6) and that such applicability may depend on the information available in each specific case to be assessed. Indeed, with the exception of OECD TG 435, for which a number of mixtures (n=152) were reported to be tested (NIH, 1999), only limited information is available in the public domain on the testing of mixtures with test methods falling under OECD TGs 430, 431 or 439

(OECD, 2014). Despite the limited information available on mixtures, the test methods falling within OECD TGs 430, 431 or 439 are currently considered to be applicable to the testing of mixtures as an extension of their applicability to substances. However, if additional information is available, this should be taken into account, in combination with the existing evidence, to evaluate the usefulness of a test method to assess mixtures. In cases where evidence can be demonstrated on the non-applicability of the Test Guideline to a specific category of mixtures, the Test Guideline should not be used for that specific category of mixtures. Similar care should be taken in case specific chemical classes or physico-chemical properties are found not to be applicable to the current Test Guidelines (e.g., gases, aerosols, specific pH ranges, etc.).

2.2. The UN GHS tiered evaluation approach

The UN GHS (UN, 2015a) recommends to consider a tiered approach for the evaluation of initial information where applicable, recognizing that not all elements may be relevant (see figure 2.2). The tiered approach provides guidance on how to organize existing information on a substance and to make a weight of evidence decision about hazard assessment and hazard classification (ideally without conducting new animal tests). Although information might be gained from the evaluation of single parameters within a tier, consideration should be given to the totality of the existing information and making an overall weight of evidence determination. This is especially true when there is conflict in information available on some parameters.

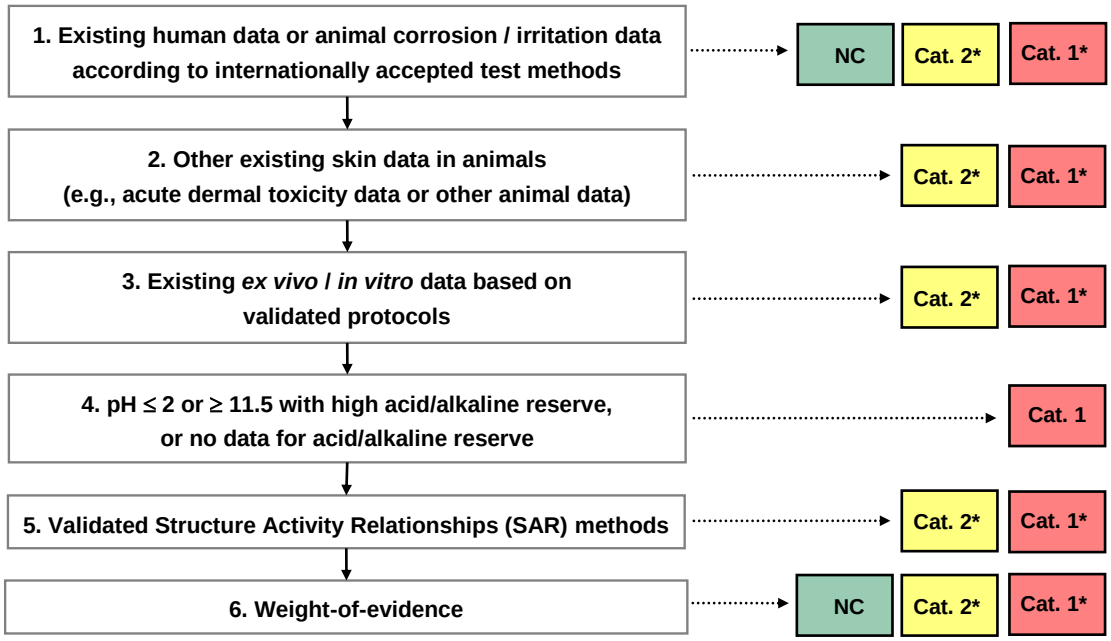

Figure 2.2. Overview of the steps of the tiered approach recommended by the UN GHS (2015a). Cat. 1: Corrosive to skin; Cat. 2: Irritating to skin; NC: No Category. * including Sub-categories, as applicable.

2.3. The EU testing and assessment strategy

In the European Union, the Endpoint Specific Guidance on the REACH Information Requirements and Chemical Safety Assessment proposes the use of a testing and assessment strategy for skin irritation and/or corrosion as summarized in figure 2.3 (ECHA, 2015a). Similar to the IATA proposed within the OECD GD 203 (OECD, 2014), this strategy comprises three parts: Part 1 on retrieval of existing information, Part 2 on a weight of evidence analysis and expert judgement and Part 3 on the generation of new information by testing. While the OECD IATA Guidance Document 203 (OECD, 2014) provides slightly more detailed guidance than the testing and assessment strategy

recommended by ECHA, there is no conceptual difference between the two strategy approaches (ECHA 2015a).

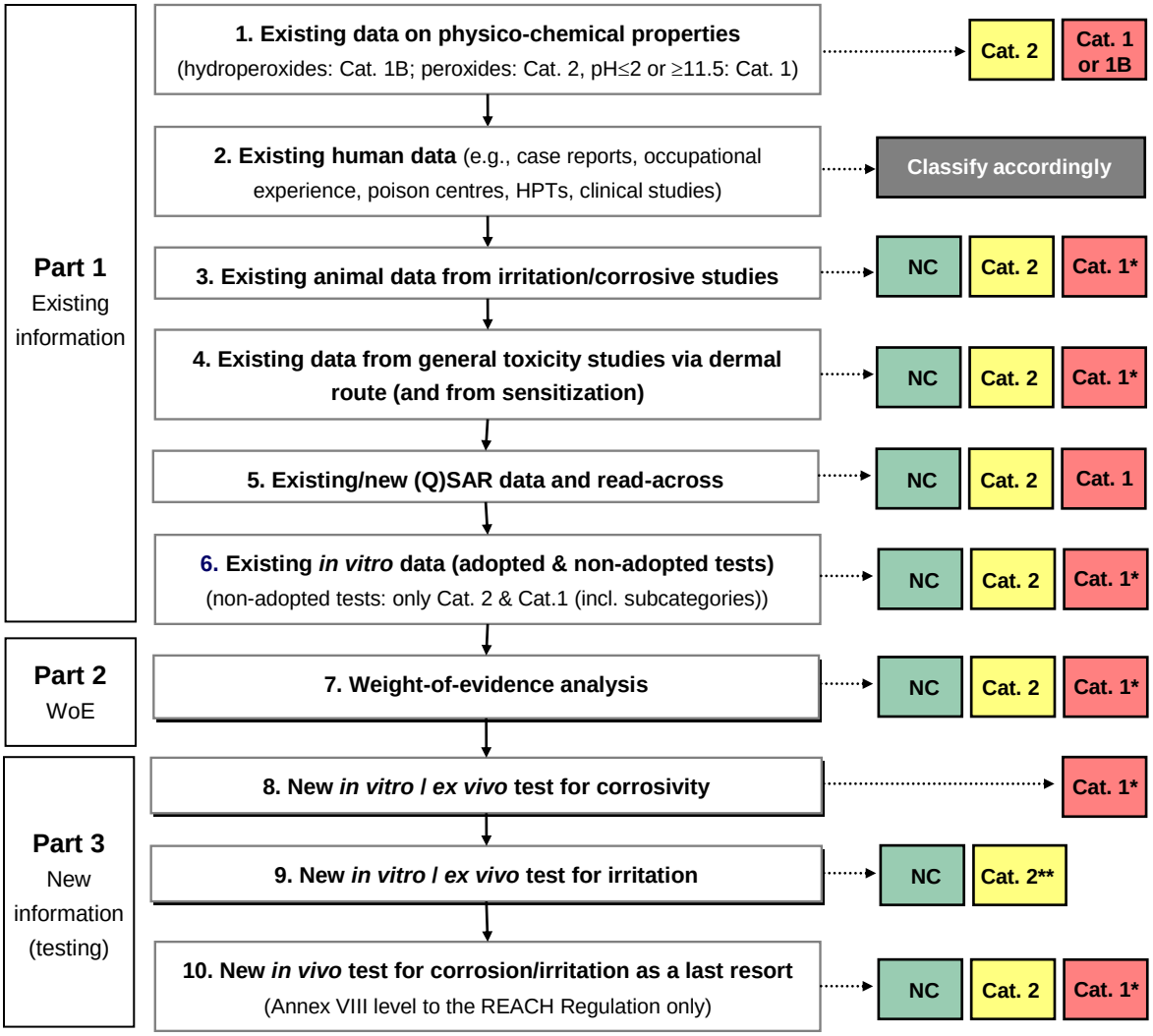

Figure 2.3. Overview of the steps of the testing and assessment strategy recommended in Endpoint Specific Guidance to the REACH Regulation (ECHA, 2015a). Cat. 1: Corrosive to skin; Cat. 2: Irritating to skin; HPT: Human Patch Test; NC: No Category. *Including corrosive Sub-categories 1A, 1B and 1C, as applicable. ** Only if there is existing data from *in vitro* skin corrosion test that showing a negative result.

In Part 1 (elements 1 to 6), existing and available information from the literature and databases is gathered and considered. The order of the elements 1 to 6, is only indicative and they may be arranged as appropriate. This may be especially helpful in cases where a reliable conclusion can be drawn from certain elements without having to consider all of them. For instance, if the substance has an extreme pH (≤ 2.0 or ≥ 11.5) skin corrosivity shall be implicitly considered[1] (element 1) and the substance classified as skin corrosive (Category 1) according to CLP so that further testing is not required.

At the end of Part 1, and if no final conclusion can be derived directly from one or several of the available pieces of information, all the information collected should be analysed using a weight of evidence approach (Part 2, element 7). In case no conclusion can be drawn from Parts 1 and 2, new data should be generated in Part 3. Indeed, even though not all elements of Part 1 might necessarily

[1] Depending on the buffering capacity of the test chemical (see chapter 4.4 for details).

be accompanied by data, it is recommended that all potential data sources are explored prior to starting the weight of evidence analysis in Part 2. While it is recommended that this approach be followed, other approaches may be more appropriate and efficient on a case-by-case basis. For example, in case there is no existing data and it is anticipated that compilation of data at elements 1-7 would be non-conclusive, it may be appropriate to directly proceed to the information generation part in Part 3.

In Part 3 (elements 8 to 10), new information on the corrosion/irritation potential of test chemicals is produced by means of *in vitro* testing (elements 8 and 9). An *in vivo* study for skin corrosion/irritation shall be considered only if the *in vitro* studies are not applicable, or the results of these studies are not adequate for classification and risk assessment. Furthermore, prior to performing any new *in vivo* test, the specific rules for adaptation from standard information requirements, as described in column 2 of Annexes VII-X to the REACH Regulation should be considered, as well as the general rules of Annex XI for adaptation of the standard testing regime (EC, 2006). Therefore, a new *in vivo* test should not be conducted before concluding the weight of evidence analysis in element 7 and performing *in vitro* testing as described in elements 8 and 9.

2.4. Applicability of testing approaches

The applicability of testing approaches as described above has been discussed and/or evaluated by a number of authors. Louekari and co-authors (2006) showed, based on four test substances as case studies, the relevance of incorporating weigh-of-evidence approaches as part of sequential testing strategies. A feasibility study conducted by Hoffmann and co-authors (2008) based on integrated testing strategies (ITS) including a combination of *in silico*, *in vitro* and *in vivo* information, showed that the best performing animal-free test strategy was a combination of TOxicity Prediction by Komputer Assisted Technology (TOPKAT), BfR-Decision Support System and the EPISKIN™ *in vitro* model. However such combination resulted in predictive capacity values almost identical as the EPISKIN™ *in vitro* model as a stand-alone test. The differences in costs was also considered marginal by the authors since the number of chemicals to be tested in EPISKIN™ was reduced only by eight (out of 100) when taking into account the expert system information. The authors also discussed the complexity of systematic construction of ITS, and recommended that further investigation is conducted to explore optimal combinations of methods within ITS (Hoffmann *et al.*, 2008). Hulzebos and Gerner (2010) further reported, based on the evaluation of five substances, how the use of weight factors and integrated assessment schemes based on (Q)SARs, read-across and physico-chemical properties can help to derive skin irritation prediction for classification and labelling.

A question often raised regarding testing approaches is how to evaluate the validity of such approaches for regulatory purposes. The outcome of a workshop organized in 2008 by EURL-ECVAM and the EPAA concluded that integrated testing strategies allowing for flexible and ad hoc approaches cannot be validated, whereas the validation of clearly defined ITS would be feasible, although practically quite difficult (Kinsner-Ovaskainen *et al*., 2012). A further workshop organized by the Swiss Federal Office of Public Health together with EURL-ECVAM and the BfR in 2010, noted that in particular the weight of evidence step in the ITS is considered a difficult step to validate. Furthermore it is believed that there might be no one-size-fits-all strategy for each endpoint but rather different testing strategies designed for e.g., different purposes, product and/or chemical classes (Eskes *et al*. 2012).

Indeed, specific test strategies have been proposed for the identification of skin corrosion and irritation hazard and the risk assessment of cosmetic ingredients as described by Macfarlane and co-authors (2009). Here again the use of weight of evidence analysis is proposed to evaluate all available data such as physicochemical properties, literature, animal, *in vitro*, human, read-across, SAR. Such evaluation is then followed by an *in vitro* test for skin corrosion and an *in vitro* test for skin irritation. No *in vivo* and human testing are used for hazard assessment. However, human testing is proposed for risk assessment. Furthermore, Robinson and co-authors (2002) described the general testing strategies implemented within industry to assess skin corrosion and irritation of ingredients and finished products without the need to test in animals. Scheel and co-authors (2011) further reported

the practical application of tiered testing strategies based on *in vitro* methods and weight of evidence approaches to determine the corrosive and irritation potential of 20 industrial products having extreme pH. Finally, Grandon and co-authors (2008) proposed the use of integrated testing strategy for skin corrosion and irritation in the framework of REACH, in which the authors exclude the *in vivo* testing as the last step, and make use of *in vitro* test methods instead as the last step of the integrated testing strategy.

3 *In vitro* prospective testing

An overview of the validated *in vitro* test methods adopted for skin corrosion and skin irritation regulatory testing are given in table 3.1. These methods have been internationally adopted since 2000 (and 2004) for skin corrosion and since 2009 (and 2010) for skin irritation by the EU (and by the OECD respectively). Although no single *in vitro* test method can cover across the full range of skin corrosion and irritation responses from the traditional Draize *in vivo* regulatory test (OECD TG 404, 2015a; EU B.4, 2008b), the validated and adopted *in vitro* methods shown in table 3.1 can replace the Draize *in vivo* test depending on the outcome of the testing or when combined within a tiered testing strategy as shown in chapter 3.1.

Table 3.1. Overview of the validated and adopted *in vitro* methods available for skin corrosion and skin irritation regulatory testing, their purposes, application and limitations.

Purpose	Test Method	Application and Limitations
Identification of skin corrosives *Positive results lead to skin corrosion classification.* *Negative results lead to no classification as corrosive.*	**OECD TG 431 / EU B.40bis** Reconstructed human epidermis (RhE) test method - EPISKIN™ Standard Model (SM) - EpiDerm™ Skin Corrosion Test (SCT) - SkinEthic™ RHE - epiCS® (previously named EST-1000)	Applicable to substances and mixtures*. Allows identification of corrosives (GHS Cat. 1), and discrimination between Sub-cat. 1A from Sub-cat. 1B-and-1C. The test guideline does not allow discrimination between skin corrosive Sub-cat. 1B and Sub-cat. 1C. However, an EPISKIN™ prediction model exists to distinguish GHS Sub-cat. 1B from 1C but its validity could not be evaluated due to the limited set of well-known *in vivo* corrosive Sub-cat. 1C chemicals. It is not designed to provide information on skin irritation, and it is not applicable to gases and aerosols. Results obtained with test chemicals presenting non-specific interactions with MTT ≥ 50% should be taken with caution when OD is used as measurement for cell viability. This may be circumvented for colour-interfering chemicals using the HPLC/UPLC as alternative measurement.
	OECD TG 430 / EU B.40 Transcutaneous Electrical resistance (TER) test method	Applicable to substances and mixtures*. Allows identification of corrosives (GHS Cat. 1). The test guideline is not able to distinguish the three GHS Sub-categories (1A, 1B and 1C). It is not designed to provide information on skin irritation, and is not applicable to gases and aerosols. Finally, the TER test method may be considered as an animal test in some countries.
	OECD TG 435 Membrane barrier test - Corrositex®	Applicable to substances and mixtures*. Allows identification of corrosives (GHS Cat. 1) and sub-categorisation into the three GHS Sub-cat. (1A, 1B and 1C). In EU, method not adopted in legislation as considered valid for the limited applicability domain of acids, bases and their derivatives. The test method is designed to provide information on skin irritation, and is not applicable to gases and aerosols. Test chemicals not causing detectable changes in the chemical detection system cannot be tested.
Identification of skin irritants *Negative results lead to no classification** .* *Positive results lead to skin irritation classification Cat. 2 if negative result with the skin corrosion test.*	**OECD TG 439 / EU B.46** Reconstructed human epidermis (RhE) test method - EPISKIN™ Skin Irritation Test (SIT) - EpiDerm™ SIT - SkinEthic™ SIT[42bis] - LabCyte EPI-MODEL24 SIT	Applicable to substances and mixtures*. Allows identification of skin irritants according to GHS Cat. 2, in case the test chemical is found to be non-corrosive. Furthermore, for countries not adopting the optional GHS Cat. 3 such as in the EU, the method also allows identification of non-classified substances. The test method is not designed to provide information on skin corrosion nor on mild irritants (optional GHS Cat. 3), and is not applicable to gases and aerosols. Results obtained with test chemicals presenting non-specific interactions with MTT ≥ 50% should be taken with caution when OD is used as measurement for cell viability. This may be circumvented for colour-interfering chemicals using the HPLC/UPLC as alternative measurement.

* Before use of the test method on a mixture for generating data for intended regulatory purposes, 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.

** Classification according to the EU CLP.

3.1. The top-down and bottom-up *in vitro* testing strategies

If generation of new information (Part 3 of the OECD IATA and of the EU testing and assessment strategy) is required (see chapter 2), the WoE assessment of all available information should be used to formulate a hypothesis of the most likely skin irritation/corrosion potential of the chemical. This hypothesis and the regulatory context under which a decision must be taken should then guide the choice of test methods to be used and the sequence of the prospective *in vitro* testing in either a top-down or a bottom-up approach. Figure 3.1 provides a schematic overview of the construction of the top-down and bottom up *in vitro* testing strategies as recommended by the OECD GD 203 and by ECHA (OECD, 2014; ECHA, 2015a).

When all available collected information and the WoE assessment result in a high a-priori probability of the test chemical to be an irritant or a corrosive, the ***top-down approach*** should be used, starting with an *in vitro* method for the identification of skin corrosion hazard followed, in case the test chemical is identified as not being corrosive, by an *in vitro* method for the identification of skin irritation hazard. Conversely, when all available collected information and the WoE assessment result in a high a-priori probability of the test chemical not being an irritant to skin, ***the bottom-up approach*** should be used, starting with an *in vitro* method for identification of skin irritation followed, in case the test chemical is identified as being irritant, by an *in vitro* method for identification of skin corrosion. An example of the applicability of such approaches has been described using the SkinEthic™ RHE model, in which high accuracy values were reported using either approach (Alépée *et al.*, 2015a).

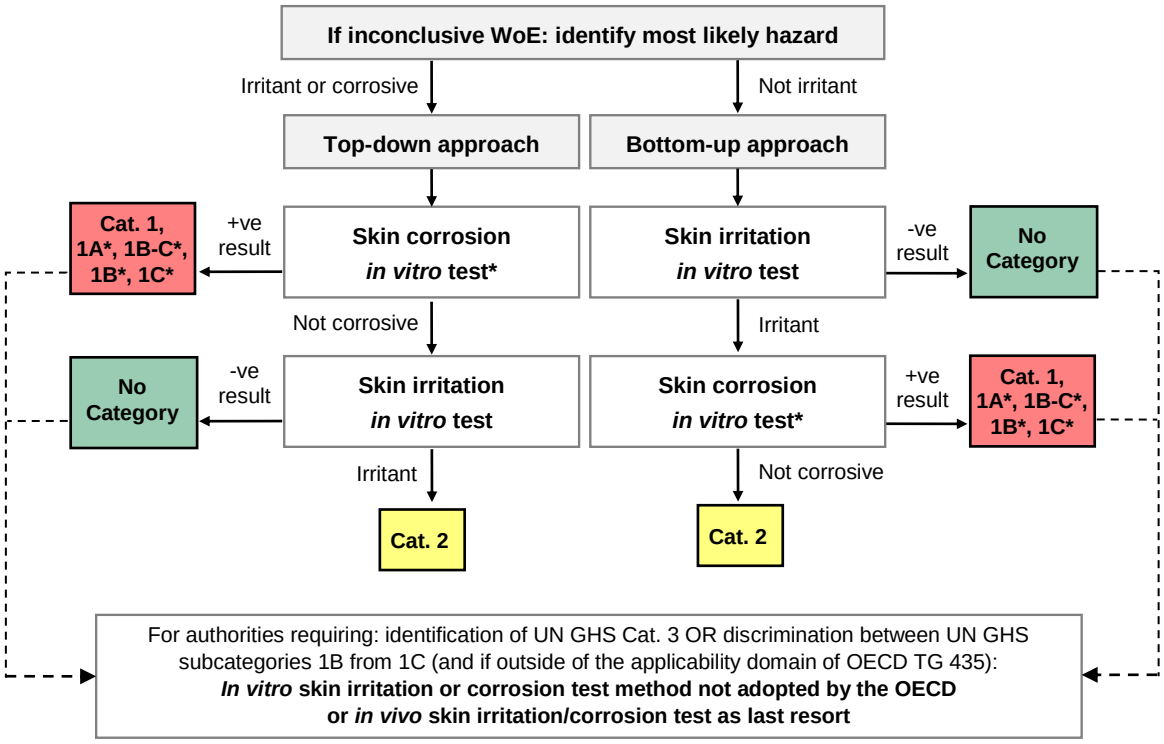

Figure 3.1. Schematic overview of the top-down and bottom-up *in vitro* testing strategies (OECD GD 203, 2014; ECHA, 2015a). Cat. 1: Corrosive to skin; Cat. 2: Irritating to skin.* Corrosive Sub-categories applicable as follows:

- **OECD TG 435** allows for the discrimination between Sub-cat. 1A, 1B and 1C but with a limited applicability domain;
- **OECD TG 431** allows for the discrimination between Sub-cat. 1A from Sub-cat. 1B-and-1C but does not permit the discrimination between Sub-categories 1B and 1C.
- **OECD TG 430** only allows the identification of corrosives into a single category without sub-categorisation, i.e., Cat. 1.

When limitations and domain of the validated and adopted *in vitro* tests are adequately considered, these tests can provide sufficient information for the decision on potential of the substance to cause

skin irritation and/or corrosion. In case of *in vitro* skin corrosion testing, the most appropriate OECD TG for the test chemical and the specific purpose should be chosen. In particular, the applicability domain and the ability of the test methods to provide information on sub-categorisation may play an important role in the choice of the test method to be used.

In the EU, only *in vitro* testing should be conducted for substances manufactured or imported in quantities between one tonne and 10 tonnes per year. In contrast, for substances manufactured or imported in quantities of ≥ 10 tonnes per year, "a*n* in vivo *study for skin corrosion/irritation shall be considered only if the* in vitro *studies (…) are not applicable, or the results of these studies are not adequate for classification and risk assessment"* (EC, 2016b). As a consequence, no *in vivo* testing should be conducted in cases where the substance falls under the scope of the specific *in vitro* tests performed and there are no substance-specific limitations to using those tests (ECHA, 2015a). Furthermore, the *in vivo* testing may be waived if an adaptation is formulated according to Annex XI to the REACH Regulation (ECHA, 2015a).

Instructions on how to submit *in vitro* information in the EU instead of *in vivo* can be found e.g. in Section 3.7 of Practical Guide 1: How to report *in vitro* data[2]. Finally, in the EU, it is considered the responsibility of the registrant to ensure that the chosen test method is suitable to obtain adequate information for classification and labelling from the *in vitro* studies. For most substances, the use of adopted EU or OECD TGs for skin corrosion/irritation purposes will provide results that will have regulatory acceptance under REACH (ECHA, 2015a).

***3.2. In vitro* alternative methods for skin corrosion**

3.2.1. Validated test methods currently available

Several *in vitro* assays for skin corrosion have undergone prevalidation (Botham *et al*, 1995) and validation studies in the '90s (Fentem *et al.*, 1998; Liebsch *et al.*, 2000). Such efforts led to the formal endorsement of the scientific validity of three *in vitro* alternatives which were adopted and included in the EU test guidelines in 2000 and in the OECD testing guidelines in 2004 and 2006 (EC, 2000, 2008b; OECD, 2015b, 2015c, 2016). These assays are:

- The reconstructed human epidermal (RhE) models (EU B.40bis and OECD TG 431), including:
  - The EPISKIN™ Standard Model (SM) validated in 1998 following a formal prospective validation studies (ESAC, 1998a),
  - The EpiDerm™ Skin Corrosion Test (SCT) validated in 2000 following a formal prevalidation and catch-up validation studies (ESAC, 2000),
  - The SkinEthic™ Reconstituted Human Epidermis (RhE) validated in 2006 for having met the Performance Standards as required in the OECD TG 431 (Kandárová *et al.*, 2006; ESAC, 2006),
  - The epiCS® (previously named EST-1000) validated in 2009 for having met the Performance Standards as required in the OECD TG 431 (ESAC, 2009a).

  A follow-up study was further undertaken in the framework of the OECD to investigate the capability of these four RhE models to correctly identify the UN GHS corrosive Sub-categories 1A, 1B and 1C, and the TG 431 has been updated accordingly in 2013 (OECD, 2013, 2016).
- The *in vitro* skin corrosion rat skin transcutaneous electrical resistance (TER) test (ESAC, 1998b), which uses excised rat skin as a test system and its electrical resistance as an endpoint (EU B.40 and OECD TG 430).
- The Corrositex® test (ESAC, 2001; NIH, 1999), which uses penetration of test chemicals through a hydrogenated collagen matrix (biobarrier) and supporting filter membrane, which was considered to be useful for acids, bases and their derivates (OECD TG 435).

[2] Available at http://echa.europa.eu/practical-guides accessed on 19 Nov. 2016.

For a full evaluation of local skin effects after a single dermal exposure, it is recommended that these assays are used within testing approaches such as the IATA recommended in the OECD GD 203 (2014) or the testing and assessment strategy recommended by ECHA (2015a).

3.2.2. Reconstructed human Epidermis (RhE) test methods

The most updated version of the assay is described in details in the OECD TG 431 (2016). The principle of the test method, its known applicability and limitations and a summary of the test method procedure are described below.

3.2.2.1. Principles of the test

The principles of 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. Corrosive test chemicals are identified by their ability to produce a decrease in cell viability below defined threshold levels (see section 3.2.2.4). The test chemical is applied topically to Reconstructed human Epidermis. Cell viability is measured by dehydrogenase conversion of the vital dye MTT (3-(4,5-Dimethyl-2-thiazol-2-yl)-2,5-diphenyltetrazolium bromide), into a blue formazan salt that is quantitatively measured after extraction from tissues.

3.2.2.2. Applicability, limitations and role within the IATA

The test methods described in the OECD TG 431 (OECD, 2016) allow the identification of corrosive and non-corrosive substances and mixtures, and is based on RhE-models using human keratinocytes, which therefore represent *in vitro* the target organ of the species of interest. Based on the overall dataset available (mainly composed of individual substances) that supported the test methods inclusion in the OECD TG 431 (Fentem *et al.*, 1998; Liebsch *et al.*, 2000; OECD, 2013), the Test Guideline is considered applicable to a wide range of chemical classes and physical states including liquids (aqueous or non-aqueous), semi-solids, solids (soluble or insoluble in water) and waxes. In addition, OECD TG 431 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 OECD TG 431 to a specific category of mixtures (*e.g.* following a strategy as proposed by Eskes *et al.*, 2012), the TG should not be used for that specific category of mixtures. Similarly, in cases where evidence can be demonstrated on the non-applicability of test methods included in OECD TG 431 to a specific category of test chemicals, these test methods should not be used for that specific category of test chemicals. Furthermore, before using the test methods falling within the OECD TG 431 on a mixture for generating data for intended regulatory purposes, 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 the mixture. Finally, the OECD TG 431 does not allow testing of gases and aerosols.

The OECD TG 431 (2016) further supports the sub-categorization of corrosive substances and mixtures into the UN GHS Sub-category 1A as well as into a combination of Sub-categories 1B and 1C (OECD, 2013; Alépée *et al.*, 2014a, 2014b). All four test methods falling within OECD TG 431 are able to sub-categorize 1A versus 1B-and-1C versus No Category (OECD, 2016; Desprez *et al.*, 2015).

The OECD TG 431 does however not allow discriminating skin corrosive Sub-category 1B from Sub-category 1C, despite of the fact that the EpiSkin™ proposes a prediction model able to distinguish between these two corrosive Sub-categories (EPISKIN™ SM, 2011), due to the limited set of well-known *in vivo* corrosive Sub-category 1C chemicals. Furthermore, while the OECD TG 431 does not provide adequate information on skin irritation, it should be noted that OECD TG 439 (2015d) specifically addresses the health effect skin irritation *in vitro* and is based on the same RhE test system, though using another protocol. For a full evaluation of local skin effects after a single dermal

exposure, the OECD Guidance Document No. 203 on an IATA for skin corrosion and irritation should be consulted (OECD, 2014). The IATA include the conduct of *in vitro* tests for skin corrosion and skin irritation before considering testing in living animals (see chapter 2.1).

Finally, 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 require 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 section 3.2.2.3-f below). Results for test chemicals producing non specific interactions with MTT ≥ 50% of the negative control should be taken with caution when OD is used as means of measurement. However, the use of HPLC/UPLC spectrophotometry as an alternative means of measuring the MTT formazan offers the possibility of evaluating the skin irritation potential of strongly coloured test chemicals that could interfere with the standard OD measurements (Alépée *et al.*, 2015b; 2016).

3.2.2.3. Method description according to OECD TG 431

The following is a generic description of the main components and procedures of the RhE test methods for skin corrosion assessment as required by the OECD TG 431 (2016).

a) RhE models available, functional conditions and demonstration of proficiency
The three-dimensional RhE models are comprised of normal, human-derived epidermal keratinocytes, which have been cultured to form a multilayered, highly differentiated model of the human epidermis. It consists of organized basal, spinous and granular layers, and a multilayered *stratum corneum* containing intercellular lamellar lipid layers arranged in patterns representing main lipid classes analogous to those found *in vivo*.

Four commercially available models based on reconstructed human epidermis have been endorsed as scientific valid to be used for regulatory purposes. Standard Operating Procedures (SOPs) for the four RhE models are available as described below and should be consulted when implementing and using one of these four models in a test laboratory.

- EPISKIN™ SM (2011),
- EpiDerm™ SCT (2012),
- SkinEthic™ RHE (2012),
- epiCS® (2012).

The test method supplier should demonstrate that each batch of the RhE model used meets defined production release criteria including:

- acceptable viability (OD values) of negative controls,
- acceptable barrier function based on the penetration of benchmark chemicals as estimated by IC_{50} or ET_{50};
- appropriate morphology of the tissues;
- and reproducibility of the model over time with corrosive and non-corrosive chemicals.

The test user on the other hand should demonstrate acceptable viability (OD values) for negative controls, and reproducibility of the test methods over time with the positive and negative controls (see positive and negative control below). Furthermore, prior to the routine use of any of the four validated RhE models that adhere to the TG 431, the test users should demonstrate technical proficiency by correctly classifying the 12 proficiency substances recommended within OECD TG 431 (2016). In case of the use of a method for sub-classification, also the correct sub-categorization should be demonstrated. Finally, as part of the proficiency exercise, it is recommended that the test 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.

b) Number of replicates
A single testing run composed of at least two tissue replicates for each exposure time, each test chemical and each control should be sufficient when the resulting classification for the test chemical 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.

c) Application and exposure time of the test chemical
A 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 (see table 3.4 for details). Depending on the method, 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. Whenever possible, solids should be tested as a fine powder.

The exposure time can vary depending on each RhE model protocols used (i.e., 3 min and 1 h for all models and for EpiSkin™ an additional exposure time of 4h). Depending on the RhE test method used and the exposure period assessed, the incubation temperature during exposure may vary between room temperature and 37ºC (i.e., EpiSkin™: room temperature (RT) for all exposure times; all other models: RT for the 3 min exposure and 37ºC for the 60 min exposure).

At end of exposure time the test chemical must be carefully washed from the epidermis with an aqueous buffer or 0.9% NaCl (usually PBS is recommended for all models).

d) Negative and positive controls
Concurrent negative and positive controls should be used in each run to demonstrate that viability (with negative controls), barrier function and tissue sensitivity (with the positive control) of the tissues are within defined historical acceptance range.

- Suggested negative controls: 0.9% (w/v) NaCl (EpiSkin™) or water (all other models)
- Suggested positive control: glacial acetic acid (EpiSkin™) 8N KOH (which is a direct MTT reducer and therefore requires adapted controls, for all other models).

e) Cell viability measurement
MTT is applied at an appropriate concentration depending on the model (i.e., 2 mL of 0.3 mg/ml for EpiSkin™ or 300 µL of 1 mg/ml for all other models) for 3 hours at 37ºC, 5% CO_2 and 95% relative humidity.

The blue formazan produced is then extracted using a solvent (e.g., acidic isopropanol for EpiSkin™ or isopropanol for all other models), and its concentration determined by measuring the Optical Density (OD) at 570 nm using a filter band pass of maximum ± 30 nm. Alternatively, the MTT concentration may be determined using an HPLC/UPLC spectrophotometry procedure which appears particularly useful for e.g. coloured test chemicals not compatible with the standard OD measurement due to too strong interference with the MTT assay (Alépée *et al.*, 2015b; 2016).

f) Non-specific interaction with MTT
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. Two types of interference may occur requiring adapted controls as follows:

1. Test chemicals that may directly reduce the MTT into blue formazan. To identify these test chemicals, 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 controls using non-viable tissues should be performed (both using the standard OD measurement or the HPLC/UPLC-spectrophotometry procedure). 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 chemical being corrected (%NSMTT).

2. Interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol. To identify such interferences, a spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, further colorant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required. When performing the standard absorbance (OD) measurement, each interfering coloured test chemical 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 true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the test chemical and incubated with MTT solution minus the percent non-specific colour obtained with living tissues exposed to the test chemical (run concurrently) but incubated with medium without MTT ($\%NSC_{living}$).

For test chemicals that produce both direct MTT reduction and colour interference (e.g. blue, purple, black chemicals) a third set of controls is required in addition to NSMTT and NSC_{living} when using the OD measurements (but not if the HPLC/UPLC-spectrophotometry procedure is used). 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 and is calculated relative to the negative control run concurrently to the test being corrected ($\%NSC_{killed}$). 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 $\%NSC_{killed}$.

Finally, it is important to determine the linearity range of the spectrophotometer with the MTT formazan before initiating testing. In particular, the OD measurements are appropriate to assess MTT interfering test chemicals when: the OD values obtained from the treated tissue extracts 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 for the test chemical already defines it as a corrosive. In any case, results for test chemicals producing %NSMTT and/or $\%NSC_{living}$ ≥ 50% of the negative control should be taken with caution when OD is used as means of measurement.

g) Acceptability criteria

For each test method using valid RhE models the following acceptability should be met to qualify the obtained results:

1. Tissues treated with the negative control should exhibit an acceptable OD reflecting their quality as required in OECD TG 431 (i.e., 0.6 ≤ OD (EpiSkin™) ≤ 1.5; 0.8 ≤ OD (EpiDerm™ & epiCS®) ≤ 2.8; 0.8 ≤ OD (SkinEthic™) ≤ 3.0) and should not be below historically established boundaries.
2. Tissues treated with the positive controls should reflect the ability of the tissues to respond to a corrosive chemical following the requirements of the test model (i.e., EpiSkin™: viability of 4h treatment with glacial acetic acid ≤ 20%; EpiDerm™ & SkinEthic™: viability of 8N KOH < 15%; epiCS®: viability of 8N KOH < 20%);
3. The variability between tissue replicates of test chemical and/or control substances should fall within the accepted limits for each valid RhE model requirements (i.e. in the range of 20-100% viability and ODs ≥ 0.3, the difference of viability between the two tissue replicates should not exceed 30%).

If either the negative control or positive control included in a run fall out of the accepted ranges, the run is considered as not qualified and should be repeated. Similarly, if the variability of test chemicals falls outside of the defined range, its testing should be repeated.

3.2.2.4. Interpretation of results and prediction models used for classification

OD values obtained for each test sample are used to calculate the percentage of viability relative to the negative control, which is set at 100%. In case HPLC/UPLC-spectrophotometry is used, the MTT formazan peak area obtained with living tissues exposed to the test chemical relative to negative control is used to determine the percent tissue viability. The Prediction Model used to identify the test chemical as corrosive or non-corrosive are shown in table 3.2 for the EpiSkin™ RhE test method and in table 3.3 for the EpiDerm™ SCT, SkinEthic™ RH and epiCS® test methods (OECD, 2016).

Table 3.2. Prediction model of the EpiSkin™ RhE test method

Viability measured after exposure time points (t=3, 60 and 240 minutes)	**UN GHS prediction**
< 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 models for supporting corrosive sub-categorisation, it was shown that around 22% of the Sub-category 1A *in vitro* results of the EpiSkin™ test method may actually constitute *in vivo* Sub-category 1B or Sub-category 1C substances/mixtures (i.e. over classifications).

** Prediction model proposed for the identification of the optional UN GHS Sub-cat. 1B in the original protocol (EpiSkin™, 2011)

*** Prediction model proposed for the identification of the optional UN GHS Sub-cat. 1C in the original protocol (EpiSkin™ SM, 2011), but not adopted as such within the OECD TG 431 due to the limited set of well-known *in vivo* corrosive Sub-cat. 1C chemicals available to assess the predictivity of the model for this Sub-category.

Table 3.3. Prediction model of the EpiDerm™ SCT, SkinEthic™ RHE and epiCS® test methods

Viability measured after exposure time points (t=3 and 60 minutes)	**UN GHS prediction**
STEP 1 for EpiDerm™ SCT, 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 indentified 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 indentified 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 indentified 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 models for supporting corrosive sub-categorisation, it was shown that the following amount of Sub-category 1A *in vitro* results may actually constitute *in vivo* Sub-category 1B or Sub-category 1C substances/mixtures (i.e. over classifications): around 29% for EpiDerm™ SCT; 31% for SkinEthic™ RHE and 33% for epiCS®.

In addition, an extended exposure period of 4h is recommended for the EpiDerm™ SCT for testing of fatty amine derivatives (characterized as cationic surfactants) which have been shown to have a tendency to be under-predicted with the test methods falling within the OECD TG 431 (Houthoff *et al.*, 2015; Kandárová and Liebsch, in press). Finally, the original prediction model from the SkinEthic™ RHE (2012), suggested the use of a separate prediction model for inorganic acids that involve the use of a 4 h exposure period, but this prediction model was no longer considered relevant following the OECD assessment of the usefulness of the RhE models for supporting corrosive sub-categorisation.

3.2.2.5. Comparison of validated RhE protocols for skin corrosion

The details of the principal protocol components of the four validated RhE models (EPISKIN™ SM, 2011; EpiDerm™ SCT, 2012; SkinEthic™ RHE, 2012; epiCS®, 2012) to be used within OECD TG 431 are shown in table 3.4.

3.2.2.6. Performance standards

In case a similar or modified *in vitro* RhE test method is proposed for regulatory purposes to be used within the OECD TG 431 for skin corrosion testing, the reliability, relevance (predictive capacity), and limitations for its proposed use should be determined to ensure its similarity to the Validated Reference Methods (VRMs), i.e., the EpiSkin™ SM and the EpiDerm™ SCT RhE test methods, in accordance with the requirements of the OECD Guidance Document No. 219 on Performance Standards (PS) to the TG 431 (OECD, 2015f). The OECD 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 the OECD TG 431.

The 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.

Similar (me-too) or modified test methods proposed for use under the OECD TG 431 (OECD, 2016) should be evaluated to determine their reliability and predictive capacity using the set of recommended Reference Chemicals within GD 219 (OECD, 2015f), which represent the full range of the TG 404 *in vivo* corrosivity scores. This should be conducted prior to their use with 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 the combined of Sub-categories 1B and 1C corrosive chemicals (OECD, 2015a; UN, 2015a). 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 VRMs and as described in the Guidance Document N. 219 (OECD, 2015f).

Table 3.4. Main test method components of the RhE test methods validated for skin corrosion testing.

Test Method Components	EpiSkin™ SM	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Model surface	0.38 cm^2	0.63 cm^2	0.5 cm^2	0.6 cm^2
Pre-check for direct MTT reduction	50 µL (liquid) or 20 mg (solid) + 2 mL MTT (0.3 mg/mL) for 180±5 min at 37°C, 5% CO_2, 95% RH, protected from light → 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) for 60 min in incubator 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) for 180± 15 min at 37°C, 5% CO_2, 95% RH, protected from light → 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) for 60 min in incubator at 37°C, 5% CO_2, 95% RH → if solution turns blue/purple, freeze-killed adapted controls should be performed
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
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±3 µL (131.6 µL/cm2) Solids: 20±2 mg (52.6 mg/cm2) + 100±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 mg (39.7 mg/cm^2) + 25 µL H_2O (or more if necessary) Waxes: flat "disc like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15 µL H_2O.	Liquids and viscous: 40±3µl (80 µL/cm^2) using nylon mesh. *Pre-test compatibility of test chemical with nylon mesh* Solids: 20±3 mg (40 mg/cm^2) + 20±2µl H_2O 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
Exposure time and temperature	3 min, 60±5 min, and 240±10 min in ventilated cabinet at 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

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

3.2.3. Transcutaneous Electrical Resistance test (TER)

The most updated version of the transcutaneous electrical resistance test is described in details in the OECD TG 430 (2015b). The principle of the test method, its known applicability and limitations and a summary of the test method procedure are described hereafter.

3.2.3.1. Principles of the test

The Transcutaneous Electrical Resistance is a measure of the electrical impedance of the skin, as a resistance value in kilo Ohms. The test chemical is applied for up to 24 hours to the epidermal surfaces of skin discs in a two-compartment test system in which the skin discs function as the separation between the compartments. The skin discs are taken from humanely killed rats aged 28-30 days.

Corrosive chemicals are identified by their ability to produce a loss of normal *stratum corneum* integrity and barrier function, which is measured as a reduction in the TER below a threshold level. For rat TER, a cut-off value of 5kΩ has been selected based on extensive data for a wide range of substances where the vast majority of values were either clearly well above (often > 10 kΩ), or well below (often < 3 kΩ) this value. Generally, test chemicals that are non-corrosive in animals but are irritating or non-irritating do not reduce the TER below this cut-off value.

Furthermore, a dye-binding step is incorporated into the test procedure for confirmation testing of positive results in the TER including values around 5 kΩ. The dye-binding step determines if the increase in ionic permeability is due to physical destruction of the *stratum corneum*. Indeed, exposure of certain non-corrosive test chemicals can result in a reduction of resistance below the cut-off of 5 kΩ allowing the passage of ions through the *stratum corneum*, thereby reducing the electrical resistance. For example, neutral organics and chemicals that have surface-active properties (including detergents, emulsifiers and other surfactants) can remove skin lipids making the barrier more permeable to ions. In case of skin corrosive effects where the *stratum corneum* is disrupted, the dye sulforhodamine B, when applied to the skin surface rapidly penetrates and stains the underlying tissue. This particular dye is stable to a wide range of substance and is not affected by the extraction procedure. As a consequence, obtaining high dye contents may indicate a corrosive effect (see details below).

3.2.3.2. Applicability, limitations and role within the IATA

The test method described in the OECD TG 430 (OECD, 2015b) allows the identification of corrosive chemical and non-corrosive substances and mixtures. Based on the dataset used in the validation (mainly substances) underlying the OECD TG 430 (Fentem *et al.*, 1998), the TG is considered applicable to a wide range of chemical classes and physical states including liquids (aqueous or non-aqueous), semi-solids, solids (soluble or insoluble in water) and waxes. However, since for specific physical states test items with suitable reference data are not readily available, a comparably small number of waxes and corrosive solids were assessed during validation. The OECD TG 430 does not allow testing of gases and aerosols. Furthermore, in cases where evidence can be demonstrated on the non-applicability of the OECD TG 430 to a specific category of substances, the TG should not be used for that specific category of substances.

In addition, the TG 430 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 OECD TG 430 to a specific category of mixtures (*e.g.* following a strategy as proposed by Eskes *et al.*, 2012), the TG should not be used for that specific category of mixtures. Furthermore, before use of the test method on a mixture for generating data for intended regulatory purposes, 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.

A limitation of the OECD TG 430, as demonstrated by the validation studies, is that it does not allow the sub-categorization of corrosive substances and mixtures in accordance with the UN GHS Sub-categories. Furthermore, while the OECD TG 430 does not provide adequate information on skin irritation, it should be noted that OECD TG 439 (2015d) specifically addresses the health effect skin irritation *in vitro*. For a full evaluation of local skin effects after a single dermal exposure, the OECD Guidance Document No. 203 on an IATA for skin corrosion and irritation should be consulted (OECD, 2014). The IATA include the conduct of *in vitro* tests for skin corrosion and skin irritation before considering testing in living animals (see chapter 2.1).

3.2.3.3. Method description according to OECD TG 430

The following is a generic description of the main components and procedures of the TER test methods for skin corrosion assessment as required by the OECD TG 430 (2015b). A Standard Operating Procedure (SOP) for the TER test method is available and should be consulted when implementing and using this test method in a test laboratory (TER, 2008).

Prior to routine use of the rat skin TER test method that adheres to OECD TG 431, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances recommended within OECD TG 431.

a) Preparation and quality control of skin discs

Skin discs are prepared from young rats as described in the test guideline. Around 10-15 skin discs, with a diameter of approximately 20-mm each, can be obtained from a single rat skin. The skin may be stored before disks are used where it is shown that positive and negative control data are equivalent to that obtained with fresh skin. Each skin disc is placed over one of the ends of a polytetrafluoroethylene (PTFE) tube, ensuring that the epidermal surface is in contact with the tube. A rubber 'O' ring is press-fitted over the end of the tube to hold the skin in place. The tube is supported by a spring clip inside a receptor chamber containing $MgSO_4$ solution (154 mM). The skin disc should be fully submerged in the $MgSO_4$ solution.

Before testing begins, the TER of two skin discs are measured as a quality control procedure for each animal skin. Both discs should give resistance values greater than 10 kΩ for the remainder of the discs to be used for the test. If the resistance value is less than 10 kΩ, the remaining discs from that skin should be discarded.

b) Number of replicates

An experiment (testing run) composed of at least three replicate skin discs should be sufficient for a test chemical when the classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean TER equal to 5 ± 0.5 kΩ, a second independent experiment (testing run) should be considered, as well as a third one in case of discordant results between the first two experiments (testing runs).

c) Application and exposure time to test chemicals

Liquids: 150 μl applied uniformly to the epidermal surface of the skin discs inside the tube.

Solids: a sufficient amount is applied to ensure that the whole surface of the epidermis is covered. 150 μl of deionised water is added on top of the solid. To achieve maximum contact with the skin, solids may need to be warmed to 30°C to melt or soften, or ground to produce a granular material or powder.

Test chemicals are applied for 24 hours at 20-23°C. At end of exposure time, the test chemical is removed by washing with a jet of tap water at up to room temperature until no further test chemical can be removed.

d) Positive and negative controls

Concurrent positive and negative controls should be used for each experiment (run), where skin discs from a single animal should be used.

Suggested negative controls: distilled water

Suggested positive control: 10M hydrochloric acid

e) TER measurements

The skin impedance is measured as TER by using a low-voltage, alternating current Wheatstone bridge. The assay measurements are recorded in resistance, at a frequency of 100 Hz and using series values. The distance between the spring clip and the bottom of the PTFE tube is maintained as a constant because this distance affects the resistance value obtained (for more details, see OECD TG 430, 2015b)

The properties and dimensions of the test apparatus and the experimental procedure used may influence the TER values obtained. The 5 kΩ corrosive threshold was developed from data obtained with the specific apparatus and procedure described in this Test Guideline. Different threshold and control values may apply if the test conditions are altered or a different apparatus is used. Therefore, it is necessary to calibrate the methodology and resistance threshold values by testing a series of Proficiency Substances from the substances used in the validation study, or from similar chemical classes to the substances being investigated. A set of suitable Proficiency Substances is provided in the OECD TG 430.

f) Dye binding method

If the TER values of test chemicals are less than or equal to 5 kΩ in the absence of visual damage, an assessment of dye penetration should be carried out on the control and treated tissues to determine if the TER values obtained were the result of increased skin permeability, or due to skin corrosion. In case of the latter where the *stratum corneum* is disrupted, the dye sulforhodamine B, when applied to the skin surface rapidly penetrates and stains the underlying tissue.

For evaluating the dye penetration, following TER assessment the magnesium sulphate is discarded from the tube and the skin is carefully examined for obvious damage. If there is no obvious major damage, 150μL of a 10% (w/v) dilution of sulforhodamine B dye in distilled water, is applied to the epidermal surface of each skin disc for 2 hours. These skin discs are then washed with tap water at up to room temperature for approximately 10 seconds to remove any excess/unbound dye. Each skin disc is carefully removed from the PTFE tube and placed in a vial containing deionised water. The vials are agitated gently for 5 minutes to remove any additional unbound dye. This rinsing procedure is then repeated, after which the skin discs are removed and placed into vials containing 5ml of 30% (w/v) sodium dodecyl sulphate (SDS) in distilled water and are incubated overnight at 60° C. After incubation, each skin disc is removed and discarded and the remaining solution is centrifuged. A 1ml sample of the supernatant is diluted 1 in 5 (v/v) with 30% (w/v) SDS in distilled water. The OD of the solution is measured at 565nm. The sulforhodamine B dye content per disc is calculated from the OD values by using the appropriate calibration curve, and the mean dye content is then calculated for the replicates.

g) Acceptability criteria

The mean TER and dye binding results are accepted if the concurrent positive and negative control values fall within the acceptable ranges for the method in the testing laboratory. The acceptable ranges for the methodology and apparatus described in OECD TG 430 are given in table 3.5.

Table 3.5. Acceptable TER and dye binding results for the positive and negative controls in the TER test method.

Controls	**Resistance (TER) range (kΩ)**	**Dye content range (μg/disc)**
Positive: 10 M Hydrochloric acid	0.5 - 1.0	40 - 100
Negative: Distilled water	10 - 25	15 – 35

3.2.3.4. Interpretation of results and prediction model used for classification

Table 3.6 illustrates the prediction model used by the TER to identify corrosive and non-corrosive test chemicals based on the UN GHS classification system.

Table 3.6. Prediction model of the TER test method

Observed results	UN GHS Prediction
TER $>$ 5 kΩ	Non corrosive
TER $\leq$ 5 kΩ, AND no obvious damage to skin disc, AND the mean disc dye content is < the values obtained with the concurrent positive control	
TER $\leq$ 5 kΩ AND obvious damage to skin disc	Corrosive
TER $\leq$ 5 kΩ AND no obvious damage to skin disc, AND the mean disc dye content is $\geq$ to the values obtained with the concurrent positive control	

3.2.3.5. Performance standards

In case a similar or modified *in vitro* RhE test method is proposed for regulatory purposes to be used within the OECD TG 430 for skin corrosion testing, the reliability, relevance (predictive capacity), and limitations for its proposed use should be determined to ensure its similarity to the VRM, i.e., the TER as described above, in accordance with the requirements of the OECD Guidance Document No. 218 on Performance Standards (PS) to the TG 430 (OECD, 2015e). The OECD 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 the OECD TG 430.

The 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 24 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.

Similar (me-too) or modified test methods proposed for use under the OECD TG 430 (OECD, 2015b) should be evaluated to determine their reliability and predictive capacity using the set of Reference Chemicals recommended within GD 218 (OECD, 2015e), which represent the full range of the TG 404 *in vivo* corrosivity scores. This should be conducted prior to their use with new test chemicals, in order to ensure that these methods are able to identify correctly UN GHS Cat. 1 corrosive chemicals and non-corrosive chemicals (OECD, 2015a; UN, 2015a). 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 VRM and as described in the Guidance Document N. 218 (OECD, 2015e).

3.2.4. In vitro membrane barrier test method for skin corrosion (Corrositex®)

The *in vitro* membrane barrier test method for skin corrosion is described in details in the OECD TG 435 (2015c). It is also recommended by the UN model regulations for the transport of dangerous goods (UN, 2015b). The only *in vitro* membrane barrier method currently endorsed as valid falling within the OECD TG 435 is the commercially available Corrositex® test method. In Europe, although it was endorsed by the ECVAM's Scientific Advisory Committee (ESAC) as being scientific valid, it has not been taken up in the EU legislation due to the fact that the Corrositex® test method was considered valid only for limited applicability domain of acids, bases and their derivates (ESAC, 2001; NIH, 1999). The principle of the test method, its known applicability and limitations and a summary of its procedures are described hereafter.

3.2.4.1. Principles of the test

The *in vitro* membrane barrier test method utilizes an artificial membrane designed to respond to corrosive chemicals in a manner similar to animal skin *in situ.* The system is comprised of two components: a synthetic macromolecular bio-barrier and a chemical detection system (CDS) which allows detecting the membrane barrier damage caused by corrosive test chemicals after the application of the test chemical to the surface of the membrane barrier.

The classification assigned is based on the time (in minutes) it takes for a test chemical to penetrate through the membrane barrier and its supporting filter to the indicator solution. Penetration of the membrane barrier (or breakthrough) is measured by a change in the colour of a pH indicator dye or changes in other properties of the indicator solution such as physical appearance (flaking, precipitation, etc.).

The time required for this change to occur (the breakthrough time) is reported to be inversely proportional to the degree of corrosivity of the test chemical, i.e., the longer it takes to detect a change, the less corrosive is the substance.

3.2.4.2. Applicability, limitations and role within the IATA

The OECD TG 435 allows the identification of corrosive test chemicals as well as the sub-categorisation of corrosive test chemicals according to the UN GHS Sub-categories of corrosivity (1A, 1B and 1C) and to the UN Transport Packing Groups I, II and III for corrosivity hazard (UN, 2015b). In addition, the test method may be used to make decisions on the corrosivity and non-corrosivity of specific classes of chemicals including organic and inorganic acids, acid derivatives[3], and bases (NIH, 1999; ESAC, 2001).

In contrast to the OECD TG 430 and 431 which were validated mainly using individual substances, the validation dataset of the Corrositex® test method falling within OECD TG 435 comprised both substances and mixtures. As a consequence the method is considered applicable to both substances and mixtures. Furthermore, the *in vitro* membrane barrier test method may be used to test solids (soluble or insoluble in water), liquids (aqueous or non-aqueous), and emulsions.

A limitation of the Corrositex® test method is that many non-corrosive and some corrosive test chemicals may not qualify for testing, based on the compatibility test. Indeed, test chemicals that do not cause a detectable change in the compatibility test (i.e., colour change of the CDS) cannot be tested with the membrane barrier test method and should be tested using other test methods. For instance, aqueous test chemicals with a pH in the range of 4.5 to 8.5 often do not qualify for testing, even though 85% of chemicals tested in this pH range were found to be non-corrosive in animal tests (NIH, 1999).

Finally, while the OECD TG 435 does not provide adequate information on skin irritation, it should be noted that OECD TG 439 (2015d) specifically addresses the health effect skin irritation *in vitro*. For a full evaluation of local skin effects after a single dermal exposure, the OECD Guidance Document No. 203 on an IATA for skin corrosion and irritation should be consulted (OECD, 2014). The IATA include the conduct of *in vitro* tests for skin corrosion and skin irritation before considering testing in living animals (see chapter 2.1).

3.2.4.3. Method description according to OECD TG 435

The following is a generic description of the main components and procedures based on the validated Corrositex® test method falling within the OECD TG 435 (2015c). Standard Operating Procedures (SOPs) for the TER test method is available and should be consulted when implementing and using this test method in a test laboratory (Corrositex, 2008).

[3] "Acid derivative" is a non-specific class designation and is broadly defined as an acid produced from a chemical either directly or by modification or partial substitution. This class includes anhydrides, halo acids, salts, and other types of chemicals.

The test system is composed of two components:
- a synthetic macromolecular bio-barrier consisting of 1) a proteinaceous macromolecular aqueous gel serving as the target for the test chemical, and 2) a permeable supporting membrane.
- the CDS which is an indicator solution that responds to the presence of a test chemical with the help of a pH indicator dye or a combination of dyes, e.g., cresol red and methyl orange, or other types of chemical or electrochemical reactions.

Prior to routine use of the *in vitro* membrane barrier test method adhering to OECD TG 435, laboratories should demonstrate technical proficiency by correctly identifying the corrosivity and UN GHS corrosive Sub-categories of the twelve Proficiency Substances recommended in OECD TG 435.

a) Test chemical compatibility test
Prior to performing the membrane barrier test, a compatibility test is conducted to determine if the test chemical is detectable by the CDS. If the CDS does not detect the test chemical, the membrane barrier test method is not suitable for evaluating the potential corrosivity of that particular test chemical and a different test method should be used. The CDS and the exposure conditions used for the compatibility test should reflect the exposure in the subsequent membrane barrier test.

In the case of Corrositex®, 150 μl or 100 mg of test chemical is added to the 'Qualify' test tube. If the test chemical fails to produce a colour or physical change, it cannot be analysed with Corrositex®.

b) Test chemical timescale category test
A test chemical that has been qualified by the compatibility test should then undergo a timescale category test, *i.e.,* a screening test to distinguish between weak and strong acids or bases.

In the case of Corrositex®, a total of 150 μl or 100 mg of test chemical is added in the "tube A" and "tube B" provided in the kit. Tubes are mixed and the resulting colours are compared to a colour chart provided to determine the category. If no colour change is observed in either tube, two drops of the 'confirm' reagent are added to tube B, which is mixed and the resulting colour used to determine the category.
- Category 1 represent test chemicals having high acid/alkaline reserve.
- Category 2 represent test chemicals having low acid/alkaline reserve.

Such timescale categorisation is then used to indicate which of the two timescales should be used in the prediction model for determining corrosivity Sub-categories (see table 3.7).

c) Number of replicates
The number of replicates should be appropriate, e.g., four for each test chemical in the case of Corrositex® (e.g., two repeats in two different batches).

d) Application of the test chemical, exposure time and barrier penetration
A suitable amount of test chemical, e.g., for Corrositex® 500 μl of a liquid or 500 mg of finely powdered solid is carefully layered and evenly distributed (at RT, 17 – 25°C) onto the upper surface of the membrane barrier placed on the top of a vial containing CDS.

The time of applying the test chemical to the membrane barrier is recorded. To ensure that short corrosion times are accurately recorded, the application times of the test chemical to the replicate vials are staggered. Each vial is then appropriately monitored and the time of the first change in the CDS indicating a barrier penetration, is recorded (minutes), so that the time elapsed between test chemical application and penetration of the membrane barrier is determined.

The measurement system can be visual or electronic. In the case of Corrositex®, the changes are observed visually and can either represent a change in colour or in physical appearance of the CDS such as flaking or precipitation in the CDS compared to the blank control.

e) Positive, negative, vehicle and blank controls

A positive (corrosive) control with intermediate corrosivity activity, *e.g.,* 110 ± 15 mg sodium hydroxide (UN GHS Sub-cat. 1B), should be tested concurrently with the test chemical. It allows assessing whether the test system is performing in an acceptable manner, or detecting if changes in the penetration time may be unacceptably longer or shorter than the established reference value indicating that the test system is not functioning properly. A second positive control that is of the same chemical class as the chemical being tested may be useful for evaluating the relative corrosivity potential of a corrosive test chemical. Finally, a weak corrosive (UN GHS Sub-cat. 1C) might also be employed as a positive control to measure the ability of the test method to consistently distinguish between weakly corrosives and non-corrosive substances. Regardless of the approach used, an acceptable positive control response range should be developed based on historical range, such as the mean ± 2 to 3 standard deviations, so that deviations outside of the acceptable range can be detected.

A negative control that is non-corrosive, e.g., 10% citric acid or 6% propionic acid, should be tested concurrently with the test chemical to demonstrate functional integrity of the membrane barrier.

If the test involves the use of a vehicle or a solvent with the test chemical, users should confirm that the vehicle or solvent is compatible with the membrane barrier system, *i.e.,* that it does not alter the integrity of the membrane barrier system, nor the corrosivity of the test chemical. In such cases, a solvent (or vehicle) control should be tested concurrently with the test chemical to demonstrate the compatibility of the solvent with the membrane barrier system.

f) Acceptability criteria

For a study to be considered acceptable, the following conditions should be met:

- The concurrently tested positive control should give the expected penetration response time (e.g. 8-16 min breakthrough time if sodium hydroxide is used as the positive control);
- The concurrently tested negative control should not be corrosive; and
- When included, the concurrently tested solvent control should neither be corrosive nor should it alter the corrosivity potential of the test chemical.

3.2.4.4. Interpretation of results and prediction model used for classification

The average time (in minutes) of the four replicates, elapsed between application of the test chemical to the membrane and its barrier penetration is used to classify the test chemical in terms of UN GHS corrosive Sub-categories (UN, 2015a) and, if applicable, UN Packing Group (UN, 2015b). In the case of Corrositex® the prediction model used is shown in table 3.7, in which the UN GHS predictions depend on the high or low acid/alkaline reserve of the test chemical as determined by the timescale category test.

Table 3.7. Corrositex® prediction model

Mean breakthrough time (min.)		**UN GHS prediction*** **
Category 1 test chemicals* (determined by the categorization test)	**Category 2 test chemicals**** (determined by the categorization test)	
0-3 min.	0-3 min.	**Corrosive optional Sub-category 1A**
> 3 to 60 min.	> 3 to 30 min.	**Corrosive optional Sub-category 1B**
> 60 to 240 min.	> 30 to 60 min.	**Corrosive optional Sub-category 1C**
> 240 min.	> 60 min.	**Non-corrosive**

* Test chemicals with high acid/alkaline reserve (Fentem *et al.*, 1998)

** Test chemicals with low acid/alkaline reserve (Fentem *et al.*, 1998)

*** UN GHS Sub-categories 1A, 1B and 1C relate to the prediction of UN packaging groups I, II and III respectively (UN, 2015a,b).

3.2.4.5. Performance standards

In case a similar or modified *in vitro* membrane barrier test method is proposed for regulatory purposes to be used within the OECD TG 435 for skin corrosion testing, the reliability, relevance (predictive capacity), and limitations for its proposed use should be determined to ensure that it is similar to that of the validated reference method, i.e. the Corrositex® test method, in accordance with the pre-defined performance standards (OECD, 2015g). The OECD 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 PS include the following sets of information (OECD, 2015g):

(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 recommended 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.

Similar (me-too) or modified test methods proposed for use under the OECD TG 435 (OECD, 2015c) should be evaluated to determine their reliability and predictive capacity using the set of recommended Reference Chemicals, which represent the full range of the TG 404 *in vivo* corrosivity scores. This should be conducted prior to their use with new test chemicals, in order to ensure that these methods are able to correctly identify UN GHS Cat. 1 corrosive and non-corrosive chemicals as well as the UN GHS skin corrosive Sub-categories 1A, 1B and 1C. 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 VRM as described within the OECD Performance Standards document currently under revision (OECD, 2015g).

3.2.5. Comparison to the in vivo *test method*

A summary of the major components of the regulatory *in vivo* and *in vitro* tests for skin corrosion is shown in table 3.8.

Morphologically, the adopted *in vitro* reconstructed human epidermis methods are closer to the human epidermis as compared to the rabbit skin. Although these models do not present all functional complexity that exist *in vivo*, i.e., the dermis and its features such as hair follicules, subaceous glands, nerve and immune cells, such features seem to play a less important role in the mechanisms of skin corrosion than in the inflammatory reactions that could lead to skin irritation. On the other hand, the adopted *in vitro* TER method makes use of excised rat skin which does include the dermis, but no blood circulation. Finally, the adopted *in vitro* membrane barrier assay does only mimic the morphological features of the *in vivo* skin.

The various adopted *in vitro* models for regulatory purposes also mimic the mechanisms of skin corrosion occurring in the *in vivo* test. These encompass:

- Cell viability (reconstructed human epidermis models) based on the principle that corrosive chemicals are able to penetrate the *stratum corneum* and are cytotoxic to the underlying layers.
- Loss of barrier function and integrity (TER assay), based on the principle that corrosive test chemicals can produce loss of *stratum corneum* integrity and barrier function.
- Membrane barrier damage (membrane barrier test) presumably by the same mechanism(s) of corrosion that operate on living skin.

With the exception of TER, the exposure times used with the adopted *in vitro* assays are comparable to those used *in vivo* (3 min, 1 h and 4 hours for the RhE test methods, and cut-offs of 3 min, 1 h and 4 h for the *in vitro* membrane barrier test), and the doses applied *in vitro* are similar or greater than those applied *in vivo* (for details see table 3.8).

Unlike the *in vivo* test, the *in vitro* test methods make systematically use of positive and negative controls to check for the functionality of the test method. In addition, the *in vitro* test methods require ensuring technical proficiency by the laboratory prior to the routine use of the *in vitro* methods, by testing a list of recommended proficiency chemicals.

Overall, all adopted *in vitro* assays for skin corrosion are all able to distinguish between corrosives (UN GHS Cat. 1) and non corrosives test chemicals according to the GHS classification system. However, regarding the possibility of the assays to identify the UN GHS corrosive Sub-categories, the following currently applies:

- The membrane barrier test falling within the OECD TG 435, is considered valid to distinguish the three UN GHS Sub-categories 1A, 1B and 1C (as well as the three UN packaging groups I, II and III), even though its applicability is limited to test chemicals that are compatible with the chemical detection system of the assay and, in the EU, to acids, bases and their derivates.
- The RhE test methods falling within the OECD TG 431 are accepted to distinguish between the UN GHS Sub-cat. 1A from a combination of Sub-categories 1B-and-1C; but not to distinguish between the UN GHS Sub-cat. 1B from the Sub-cat. 1C due to the limited set of well-known *in vivo* corrosive Sub-cat. 1C chemicals.
- The TER assay falling within the OECD TG 430 does not allow the sub-categorization of corrosive test chemicals in accordance with the UN GHS Sub-categories.

Finally, none of the *in vitro* assays adopted for skin corrosion testing provide with adequate information on skin irritation. For that purpose, the *in vitro* method falling within the OECD TG 439 should be used as it specifically addresses the health effect skin irritation *in vitro*. For a full evaluation of local skin effects after a single dermal exposure, it is recommended that the OECD Guidance Document No. 203 on an IATA for skin corrosion and irritation should be consulted (OECD, 2014). The IATA include the conduct of *in vitro* tests for skin corrosion and skin irritation before considering testing in living animals (see chapter 2.1).

Table 3.8. Comparison of the principal method components of the regulatory accepted *in vivo and in vitro* tests for skin corrosion.

	***In vivo* test for skin corrosion (OECD TG 404)**	***In vitro* human skin model (OECD TG 431)**	***In vitro* TER (OECD TG 430)**	***In vitro* membrane barrier test - Corrositex® (OECD TG 435)**
Model used	Albino rabbit	Three-dimensional reconstructed human epidermis, consisting of organized basal, spinous and granular layers, and a multilayered *stratum corneum* (0.38 to 0.63 cm^2 surface depending on the model).	Skin disks prepared from young rats, where 10-15 skin discs can be obtained per rat skin (0.79 cm^2 surface)	A synthetic macromolecular bio-barrier and a chemical detection system (CDS) which indicates the presence of a test chemical.
Number of replicates	1 to 3 animals based on severity of effects	At least 2 replicates for each exposure time	At least 3 skin disks	2 repeats in 2 batches
Dose and application of test chemical	0.5 ml (liquids) or 0.5 g (solids) applied to ~ 6 cm^2 of skin and covered with a gauze patch (~ 83.3 µl or mg / cm^2). Solids might be moisten to ensure good skin contact.	Liquids: 40-50 µl (79.4 to 131.6 µl/cm^2 depending on model). Solids: 20-25 mg (39.7 to 52.6 mg / cm^2 depending on model).	Liquids: 150 µl (~ 189.9 µl/cm^2). Solids: sufficient amount to cover surface, and 150 µl of deionised water added on top of the solid.	0.5 ml (liquids) or 0.5 g (solids) applied on membrane.
Controls	The potential influence of the vehicle on irritation of the skin by the test chemical should be minimal, if any.	Negative control: 0.9% NaCl or water Positive control: 8N KOH or glacial acetic acid	Negative control: distilled water Positive control: 10M hydrochloric acid	- Negative control: e.g., 10% citric acid or 6% propionic acid - Positive control: e.g., sodium hydroxide - Vehicles or solvents should not alter integrity of the membrane barrier system, and should not alter the corrosivity of the test chemical.
Exposure time	3 min, 1 hour, 4 hours applied in a sequential way, so that if corrosive effects are observed the test is terminated. If no corrosive effects seen after 4 h exposure, the animal is observed up to 14 days.	3 min at RT 1 hour at RT or at 37°C depending on the model. In the Episkin™ model, also 4 hours at RT	24 hours at RT	The time needed for a test chemical to penetrate the membrane barrier is used to predict corrosivity. It is reported to be inversely proportional to the degree of corrosivity, i.e., the longer it takes to penetrate, the less corrosive is the substance.
Washing	At the end of exposure time to remove test chemical	At the end of exposure time to remove test chemical	At the end of exposure time to remove test chemical	Not necessary
Endpoint assessed	- Grading of skin reactions. - Other reactions such as: defatting of skin, clinical signs of toxicity and body weight, persistence of alopecia, hyperkeratosis, hyperplasia and scaling. - Histopathology may be carried out in case of equivocal responses	Cell viability: based on the principle that corrosive chemicals are able to penetrate the *stratum corneum* and are cytotoxic to the underlying layers.	- Transcutaneous Electrical Resistance: based on the principle that corrosive test chemicals can produce loss of *stratum corneum* integrity and barrier function, measured by the TER. - Dye binding: to determine if TER values below the cut-off but in absence of visual damage, are due to increase in permeability or to skin corrosion.	The time it takes a substance to penetrate through the membrane barrier. Penetration is measured by colour or physical change of the Chemical Detection System.
Interpretation of results	Results can be used to classify according to all UN GHS Categories, optional categories and Sub-categories for skin corrosion and irritation.	Able to distinguish between corrosive and non corrosive, and discrimination between UN GHS Sub-cat. 1A from combined Sub-cat. 1B-and-1C. Not able to distinguish GHS Sub-cat. 1B from 1C due to the limited set of well-known *in vivo* corrosive Sub-cat. 1C chemicals.	Able to distinguish between corrosive (UN GHS Cat. 1) and non corrosive.	Allows identification of corrosives (GHS Cat. 1) and sub- categorisation into the three GHS Sub-categories (1A, 1B and 1C).
Limitations	- May over-predict human responses. - May be variable between laboratories. - Does not assess repetitive low-dose exposure. - Has the potential to cause considerable discomfort or pain to laboratory animals	- Does not allow discrimination between skin corrosive Sub-cat. 1B and 1C. - Not designed to provide information on skin irritation. - Not applicable to gases and aerosols. - Results obtained with test chemicals having non-specific interactions with MTT ≥ 50% should be taken with caution when OD is used as measurement for cell viability. This may be circumvented for coloured interfering test chemicals with the use of HPLC/UPLC as an alternative measurement.	- Not able to distinguish the three GHS Sub-categories (1A, 1B and 1C). - Not designed to provide information on skin irritation. - Not applicable to gases and aerosols.	- Not designed to give information on skin irritation. - Not applicable to gases and aerosols. - Test chemicals not causing detectable changes in the chemical detection system cannot be tested. - In EU, considered valid only for acids, bases and their derivatives.

3.3. In vitro Alternative Methods for Skin Irritation

3.3.1. Validated test methods currently available

A number of *in vitro* assays for skin irritation have undergone prevalidation (Fentem *et al.*, 2001) followed by optimization studies (Cotovio *et al.*, 2005; Portes *et al.*, 2002; Zuang *et al.*, 2002) that led to a formal validation study (Spielmann *et al.*, 2007; Eskes *et al.,* 2007) and the endorsement of the scientific validity of the EPISKIN™ Skin Irritation Test (SIT) considered to be "*a reliable and relevant stand-alone test for predicting rabbit skin irritation, when the endpoint is evaluated by MTT reduction, and to be used as a replacement for the Draize Skin Irritation Test (OECD TG 404) for the purposes of distinguishing between skin irritating and no-label (non-skin irritating) test substances"* based on the former EU DSD classification system (ESAC, 2007).

Following the validation study, two RhE models considered to be 'similar' tests to the validated test method EPISKIN™ SIT, i.e. the EpiDerm™ Skin Irritation Test (SIT) modified protocol and the SkinEthic™ RHE test method, were shown to meet the requirements of the performance standards as defined by EURL-ECVAM for *in vitro* skin irritation testing and to have sufficient accuracy and reliability for prediction of skin irritating and non-skin irritating test chemicals compared to the validated EPISKIN™ assay (Alépée *et al.*, 2010; Kandárová *et al.*, 2009a, 2009b; OECD, 2010a; Tornier *et al.*, 2010). Following a review by ESAC, both test methods were endorsed to be scientific valid for having met the criteria outlined in the performance standards, also based on the former EU DSD classification system (ESAC, 2008).

With the implementation in 2008 of the UN GHS Classification system in the EU by means of the EU CLP Regulation (EC, 2008a, see also chapter 1.4), the performances of all three test methods (EPISKIN™ SIT, EpiDerm™ SIT modified protocol and SkinEthic™ RHE) have been re-evaluated to take into account the change in the cut-off value for the classification of skin irritants (shifted from a cut-off of 2 for the EU DSD (EC, 2001) to a cut-off of 2.3 for the EU CLP/UN GHS Cat.2 (EC, 2008a, 2016a), see figure 1.1). Results from the three test methods were considered to be satisfactory so that the statements relating to their scientific validity continued to be accurate and were extended to the EU CLP (GHS) classification system (ESAC, 2009b).

An EU Test Guideline was adopted in 2009 (EU test method B.46; EC, 2009b) and an OECD TG in 2010 (OECD TG 439, 2015d) on "*In vitro* Skin Irritation: Reconstructed Human Epidermis Model". In 2013, the OECD TG 439 has been revised to include a fourth test method, the
Labcyte EPI-MODEL 24SIT, considered to be scientific valid for having met the established performance standard criteria (Katoh *et al.*, 2009, 2011; Kojima *et al.*, 2012, 2014; OECD, 2011).

As a consequence, four commercially available RhE test methods currently comply with the OECD TG 439 (2015d) for *in vitro* skin irritation regulatory testing. These are:

- EPISKIN™ SIT, validated following an ECVAM prospective validation study (ESAC, 2007, 2009b),
- EpiDerm™ SIT modified protocol, validated for having met the established performance standards (ESAC, 2008, 2009b),
- SkinEthic™ RHE SIT[42bis], validated for having met the established performance standards (ESAC, 2008, 2009b), and
- Labcyte EPI-MODEL 24SIT, validated for having met the established performance standards (OECD, 2011).

For a full evaluation of local skin effects after a single dermal exposure, it is recommended that these assays are used within testing approaches such as the IATA recommended in the OECD GD 203 (2014) or the testing and assessment strategy recommended by ECHA (2015a).

3.3.2. Principles of the test

The three-dimensional RhE models are comprised of non-transformed human-derived epidermal keratinocytes cells which have been cultured in an air-liquid interface to form a multilayered, highly differentiated model of the human epidermis. They consist of organised basal, spinous and granular

layers, and a multilayered *stratum corneum* containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*. The *in vitro* RhE models represent therefore the target organ of the species of interest.

Chemical-induced skin irritation, manifested by erythema and oedema, is the result of a cascade of events beginning with penetration of the chemicals through the *stratum corneum* where they may damage the underlying layers of keratinocytes and other skin cells. The damaged cells may either release inflammatory mediators or induce an inflammatory cascade which also acts on the cells in the dermis, particularly the stromal and endothelial cells of the blood vessels. It is the dilation and increased permeability of the endothelial cells that produce the observed erythema and oedema (Wells *et al.*, 2004).

The RhE-based test methods (in the absence of any vascularisation in the *in vitro* test system) measure the initiating events in the cascade of skin irritation, e.g. cell / tissue damage, using cell viability as readout. Test chemicals are applied topically to the RhE models and cell viability is measured by enzymatic conversion of the vital dye MTT into a blue formazan salt that is quantitatively measured after extraction from tissues.

Irritant chemicals are identified by their ability to decrease cell viability below defined threshold levels (i.e. ≤ 50 %, for UN GHS category 2 irritants). For countries or regions that do not adopt the optional UN GHS Category 3 (mild irritants), such as in the EU, test chemicals that produce cell viabilities above the defined threshold level, are identified as not requiring classification (i.e. > 50%, EU CLP No Category). Finally, if the main endpoint considered in the regulatory adopted RhE models is the cell viability assessed by the reduction of MTT, the release of IL-1α in the EPISKIN^{TM} SIT was considered as a useful adjunct to the MTT assay as it has the potential to increase the sensitivity of the test without reducing its specificity. This endpoint could be used to confirm negative results obtained with the MTT endpoint (ESAC, 2007).

3.3.3. Applicability, limitations and role within the IATA

The reconstructed human epidermis tests falling under the OECD TG 439 can be used for the hazard identification of UN GHS Cat. 2 irritant chemicals (substances and mixtures), when test results are supported by a non-corrosive outcome (based on e.g., OECD TG 430, 431 or 435). In member countries or regions that do not adopt the optional UN GHS Category 3 (mild irritants), such as in the EU, the OECD TG 439 can also be used to identify test chemicals not requiring classification according to the UN GHS classification system. Therefore, depending on the regulatory framework and the classification system in use, the OECD TG 439 may be used to determine the skin irritancy of chemicals either as a stand-alone replacement test for *in vivo* skin irritation testing or as a partial replacement test within a testing strategy (OECD, 2014).

A limitation of the OECD TG 439 is that it does not allow the classification of chemicals to the optional UN GHS Category 3 (mild irritants). Furthermore, the OECD TG 439 does not provide adequate information on skin corrosion. For that purpose other *in vitro* methods such as OECD TG 430, 431 or 435 (see chapter 3.2) may be used as they specifically addresses the identification of skin corrosion hazard. For a full evaluation of local skin effects after a single dermal exposure, the Guidance Document No. 203 on an IATA for skin corrosion and irritation should be consulted (OECD, 2014). The IATA include the conduct of *in vitro* tests for skin corrosion and skin irritation before considering testing in living animals (see chapter 2.1).

The OECD TG 439 is applicable to mixtures and substances as well as to liquids (aqueous or non-aqueous), semi-solids, solids (soluble or insoluble in water) and waxes. However, before using the test methods falling within the OECD TG 439 on a mixture for generating data for intended regulatory purposes, 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 the mixture. 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 OECD TG 439 to a specific category of mixtures (*e.g.* following a strategy as proposed by Eskes *et al.*, 2012), the TG should not be used for that specific category of mixtures. Similar care should be taken in case specific chemical classes or physico-chemical properties are found not to be applicable to the current Test Guideline.

Finally, the OECD TG 439 does not allow testing of gases and aerosols. Furthermore, test chemicals absorbing light in the same range as MTT formazan and test chemicals that are able to directly reduce the vital dye MTT (to MTT formazan), may interfere with the tissue viability measurements and require the use of adapted controls for corrections. The type of adapted controls required will vary depending on the type of interference produced by the test chemical and the procedure used to measure MTT formazan (see section 3.3.4-g). However, results for test chemicals producing non specific interactions with MTT ≥ 50% of the negative control should be taken with caution when OD is used as means of measurement. However, the use of HPLC/UPLC spectrophotometry as an alternative means of measuring the MTT formazan offers the possibility of evaluating the skin irritation potential of strongly coloured test chemicals that could interfere with the standard OD measurements (Alépée *et al.*, 2015b; 2016).

3.3.4. Method description according to OECD TG 439

The following is a generic description of the main components and procedures of the RhE test methods for skin irritation assessment as required by the OECD TG 439 (2015d).

a) RhE models available, functional conditions and demonstration of proficiency
The RhE models are based on three-dimensional reconstituted human epidermis and are generated by growing keratinocyte cultures at the air-liquid interface on various substrates, enabling the topical application of either neat or diluted test chemicals (Botham *et al*., 1998; van de Sandt *et al*., 1999). Four commercially available RhE models have been endorsed as scientific valid and are included within the OECD TG 439. Standard Operating Procedures for these four models are available as described below and should be consulted when implementing and using one of these four models for regulatory purposes.

- EPISKIN™ SIT (2009),
- EpiDerm™ EPI-200-SIT (2009),
- SkinEthic™ RHE SIT-42bis (2009),
- Labcyte EPI-MODEL 24SIT (2011).

Test method suppliers should provide to the test user data that demonstrates that each batch of the RhE model provided meets defined production release criteria including:

- acceptable viability (OD values) of negative controls,
- acceptable barrier function based on the penetration of benchmark chemicals as estimated by IC_{50} or ET_{50}; and
- appropriate morphology of the tissues.

The test user on the other hand should demonstrate acceptable viability (OD values) for negative controls, as well as reproducibility of the test method over time with the positive and negative controls (see below).

Furthermore, prior to the routine use of any of the four validated RhE models that adhere to the TG 439, test users should demonstrate technical proficiency by correctly classifying the 10 proficiency substances recommended within OECD TG 439 (2015d). Finally, as part of the proficiency exercise, it is recommended that the test 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 and/or time periods.

b) Number of replicates
A single testing run composed of three replicate tissues should be sufficient for a test chemical when the classification is unequivocal. However, in cases of borderline results, such as non-concordant

replicate measurements and/or mean percent viability equal to 50 ± 5%, a second run should be considered, as well as a third one in case of discordant results between the first two runs.

c) Application and exposure time of the test chemical
A sufficient amount of test chemical should be applied to uniformly cover the epidermis surface while avoiding an infinite dose, i.e. ranging from 26 to 83 μL/cm^2 or mg/cm^2 (see table 3.10 for details). For solid chemicals, the *epidermis* surface should be moistened with deionised or distilled water before application to improve contact between the test chemical and the *epidermis* surface. Whenever possible, solids should be tested as a fine powder. A nylon mesh may be used as a spreading aid in some cases (see table 3.10).

Depending on the RhE test methods used, the exposure period ranges from 15 to 60 minutes (i.e., 15 min for EpiSkin™ and Labcyte, 42 min for SkinEthic™ and 60 min for EpiDerm™), and the incubation temperature from 20 to 37°C (i.e., EpiDerm™: 35 out of the 60 min at 37°C; all other models: RT).

At end of exposure period the test chemical should be carefully washed from the epidermis surface with aqueous buffer or 0.9% NaCl (i.e., EpiDerm™: DPBS; all other models: PBS).

d) Positive and negative controls
Concurrent negative and positive controls should be used in each run to demonstrate that viability (using the negative controls), barrier function and tissue sensitivity (using the positive control) are within defined historical acceptance range.

- The suggested positive control is 5% aqueous SDS (for all models).
- The suggested negative controls are water or phosphate buffered saline (PBS) (i.e., PBS for EpiSkin™ and SkinEthic™; DPBS for EpiDerm™ and distilled water for Labcyte)

e) Post-treatment incubation
It is essential that viability measurements are not performed immediately after exposure to the test chemical, but after a sufficiently long post-treatment incubation period of the rinsed tissue in fresh medium. This period allows both for recovery from weak cytotoxic effects and for appearance of clear cytotoxic effects. A 42 hours post-treatment incubation period is used for all RhE models.

f) Cell viability measurement
Tissue samples are placed in MTT solution of appropriate concentration (e.g., 0.3 to 1 mg/ml) for 3 hours (i.e., 2 mL of 0.3 mg/ml for EpiSkin™; 300 μL of 1 mg/ml for EpiDerm™ and SkinEthic™; and 500 μL of 0.5 mg/ml for Labcyte). The MTT is converted into blue formazan by the viable cells. The precipitated blue formazan product is then extracted using a solvent (e.g., isopropanol, acidic isopropanol) (e.g., acidic isopropanol for EpiSkin™ or isopropanol for all other models), and the concentration of formazan determined by measuring the OD at 570 nm using a filter band pass of maximum ± 30 nm. Alternatively, the MTT concentration may be determined using an HPLC/UPLC spectrophotometry procedure which appears particularly useful for e.g. coloured test chemicals not compatible with the standard OD measurement due to too strong interference with the MTT assay (Alépée *et al.*, 2015b; 2016).

g) Non-specific interaction with MTT
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. Two types of interference may occur requiring adapted controls as follows:

1. Test chemicals that may directly reduce the MTT into blue formazan. To identify these test chemicals, 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 controls using non-viable tissues should be performed (both using the standard OD measurement or the HPLC/UPLC-spectrophotometry procedure). 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 chemical being corrected (%NSMTT).

2. Interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol. To identify such interferences, a spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, further colorant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required. When performing the standard absorbance (OD) measurement, each interfering coloured test chemical 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 true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the test chemical and incubated with MTT solution minus the percent non-specific colour obtained with living tissues exposed to the test chemical (run concurrently) but incubated with medium without MTT ($\%NSC_{living}$).

For test chemicals that produce both direct MTT reduction and colour interference (e.g. blue, purple, black chemicals) a third set of controls is required in addition to NSMTT and NSC_{living} when using the OD measurements (but not if the HPLC/UPLC-spectrophotometry procedure is used). In this additional control (NSC_{killed}), 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 and is calculated relative to the negative control run concurrently to the test being corrected ($\%NSC_{killed}$). 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 $\%NSC_{killed}$.

Finally, it is important to determine the linearity range of the spectrophotometer with the MTT formazan before initiating testing. In particular, the OD measurements are appropriate to assess MTT interfering test chemicals when: the OD values obtained from the treated tissue extracts 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 for the test chemical already defines it as a corrosive. In any case, results for test chemicals producing %NSMTT and/or $\%NSC_{living}$ ≥ 50% of the negative control should be taken with caution as this is the cut-off used to distinguish classified from not classified chemicals when OD is used as means of measurement.

h) Acceptability criteria

For each test method using valid RhE model batches the following acceptability should be met to qualify the obtained results:

1. Tissues treated with the negative control should exhibit an acceptable OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes (i.e., OD ≥ 0.6 for EpiSkin™ ; 1.0 ≤ OD ≤ 2.5 for EpiDerm™; OD ≥ 1.2 for SkinEthic™; 0.7 ≤ OD ≤ 2.5 for LabCyte) and should not be below historically established boundaries.
2. Tissues treated with positive controls (e.g., 5% aqueous SDS) should reflect the ability of the tissues to respond to an irritant chemical under the conditions of the test method (e.g., viability ≤ 20% for EpiDerm™ and ≤ 40% for the other RhE models).
3. Associated and appropriate measures of variability between tissue replicates, i.e., standard deviations (SD) should fall within the established acceptance limits for the test method used (i.e., SD ≤18 for all adopted RhE models).

3.3.5. Interpretation of results and prediction model used for classification

The OD values obtained for each test chemical are used to calculate the 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 MTT formazan peak area obtained with the concurrent negative control. The cut-off value of percentage cell viability distinguishing irritant from non-classified test chemicals are given in table 3.9 below.

Table 3.9. Prediction model of the four *in vitro* RhE test methods falling within OECD TG 439 used to identify skin irritation hazard

***In vitro* result**	**UN GHS prediction**
Mean tissue viability* ≤ 50%	Requires UN GHS classification (UN GHS Cat. 1 or Cat. 2) If test chemical is found to be non-corrosive (e.g., based on TG 430, 431 or 435), it is predicted as: UN GHS Cat. 2 irritant to skin
Mean tissue viability* > 50%	In member countries or regions that do not adopt the optional UN GHS Category 3 (mild irritants), such as in the EU: No Category (i.e., EU CLP No Cat.)

*after exposure and post-treatment incubation

3.3.6. Comparison of validated RhE protocols for skin irritation

The details of the principal protocol components of the four validated RhE models (EPISKIN™ SIT, 2009; EpiDerm™ EPI-200-SIT, 2009; SkinEthic™ RHE SIT-42bis, 2009; Labcyte EPI-MODEL 24SIT, 2011) to be used within OECD TG 439 are shown in table 3.10.

3.3.7. Performance standards

Performance Standards (PS) are available in the OECD Guidance Document No. 220 (OECD, 2015h) to facilitate, in accordance with the principles of the OECD Guidance Document No. 34 (OECD, 2005), the validation and assessment of new similar and modified RhE-based test methods proposed to be used for regulatory purposes within the OECD TG 439 for skin irritation testing. These PS allow determining the validation status (reliability and relevance) of similar and modified skin irritation test methods that are structurally and mechanistically similar to the RhE test methods included within OECD TG 439 (2015d). The Validated Reference Methods used to develop the PS was the EpiSkin™ SIT test method; whereas the EpiDerm™ SIT (EPI-200) test method was used as a VRM together with EpiSkin™ test method to define only the Essential Test Method Components.

The 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 20 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.

Similar (me-too) or modified test methods proposed for use under the OECD TG 439 (OECD, 2015d) should be evaluated to determine their reliability and predictive capacity using the Reference Chemicals recommended within GD 220 (OECD, 2015h), that represent the full range of the TG 404 *in vivo* irritation scores. This should be conducted prior to their use for testing other chemicals, in order to ensure that these methods are able to correctly discriminate No Category chemicals (for countries and regions that do not adopt the optional UN GHS Cat. 3) from Irritant chemicals (UN GHS Cat. 2). 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 VRM and as described in the Guidance Document No. 220 (OECD, 2015h).

Table 3.10: Principal protocol components of the RhE models for skin irritation testing based on the SOPs of the validation and catch-up studies

Test Method Components	EpiSkin™ SIT	EpiDerm™ EPI-200-SIT	SkinEthic™ RHE SIT-42bis	Labcyte EPI-MODEL 24SIT
Model surface	0.38 cm^2	0.63 cm^2	0.5 cm^2	0.3 cm^2
Pre-check for direct MTT reduction	10 µL (liquid) or 10 mg (solid) + 2 mL MTT (0.3 mg/mL) for 180±5 min at 37ºC protected from light → if solution turns blue/purple, water-killed adapted controls should be performed	30 µL (liquid) or 25 mg (solid) + 1 mL MTT (1 mg/mL) for 60 min at 37ºC, 5% CO_2, 95% RH → if solution turns blue/purple, freeze-killed adapted controls should be performed	16 µL (liquid) or 16 mg (solid) + 300 µL MTT (1 mg/mL) for 180± 15 min at 37°C protected from light → if solution turns blue/purple, freeze-killed adapted controls should be performed	25 µL (liquid) or 25 mg (solid) + 0.5 mL MTT (0.5 mg/mL) for 60 min at 37ºC, 5% CO_2, 95% RH → if solution turns blue/purple, freeze-killed adapted controls should be performed
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	30 µL (liquid) or 25 mg (solid) + 300 µL H_2O for 60 min at 37ºC, 5% CO_2, 95% RH → if solution becomes coloured, living adapted controls should be performed	Applicable to colouring test substances or dye test substances able to stain RHE tissues → if test chemical is coloured, living adapted controls should be performed	25 µL (liquid) or 25 mg (solid) + 500 µL H_2O for 15 min at 37ºC, 5% CO_2, 95% RH → if solution becomes coloured, living adapted controls should be performed
Number of tissue replicates	3 tissues (replicates)	3 tissues (replicates)	3 tissues (replicates)	3 tissues (replicates)
Treatment doses and application	Liquids and viscous: 10 µL (26.3 µL/cm2) Solids: 10±2 mg (26.3 mg/cm2) + 5µL distilled water Viscous/sticky: 10±2 mg (26.3 mg/cm2)	Liquids: 30 µL (47.6 µL/cm^2) with a nylon mesh if needed (*pre-test compatibility of test chemical with nylon mesh)* Semisolids: 30 µL (47.6 µL/cm^2) Solids: 25 mg (39.7 mg/cm^2) + 25 µL DPBS Waxes: flat "cookie like" piece of ca. 8 mm diameter placed atop the tissue wetted with sterile DPBS.	Liquids: 16±0.5µl (32 µL/cm^2) using nylon mesh (*pre-test compatibility of test chemical with nylon mesh)* Solids: 16±2 mg (32 mg/cm^2) + 10 µl distilled water Sticky substances: 16±2 mg (32 mg/cm^2) using nylon mesh (*pre-test compatibility of test chemical with nylon mesh)*	Liquids: 25µl (83.3 µL/cm^2) Solids: 25±1 mg (83.3 mg/cm^2) + 25 µl distilled water
Exposure time and temperature	15±0.5 min in ventilated cabinet at RT (19-23ºC)	60 ± 1 min. The first 35 min at 37ºC, 5% CO_2, 95% RH, and the remaining time at RT in sterile hood	42±1 min at RT	15±0 .5 min in the cabinet
Rinsing	25 mL PBS filling and emptying the tissue inserts	15 times with DPBS + submerging inserts 3 times in 150 ml DPBS	25 times with 1 ml PBS	Fill and discard inserts at least 15 times (or more if needed) with PBS

Test Method Components	EpiSkin™ SIT	EpiDerm™ EPI-200-SIT	SkinEthic™ RHE SIT-42bis	Labcyte EPI-MODEL 24SIT
Post-treatment incubation period	42 ± 1 hours at 37°C, 5% CO_2, 95% RH (in culture medium)	42 ± 2 hours at 37°C, 5% CO_2, 95% RH (in culture medium)	42 ± 1 hours at 37°C, 5% CO_2, 95% RH (in culture medium)	42 ± 1 hours at 37°C, 5% CO_2, 95% RH (in culture medium)
Negative control	PBS	Dulbecco's PBS (DPBS)	PBS	Distilled water
Positive control	5% aq. SDS re-spread after 7 min of exposure	5% aq. SDS	5% aq. SDS	5% aq. SDS
MTT solution	2 mL 0.3 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL	500 µL 0.5 mg/mL
MTT incubation time and temperature	180 ± 5 min) at 37°C, 5% CO_2, 95% RH	180 ± 5 min at 37°C, 5% CO_2, 95% RH	180 ± 5 min at 37°C, 5% CO_2, 95% RH	180 ± 5 min at 37°C, 5% CO_2, 95% RH
Extraction solvent	500 µl acidic isopropanol (0.04 N HCl in isopropanol) (tissues fully immersed)	2 mL isopropanol (extraction from top and bottom of insert)	1.5 mL isopropanol (extraction from top and bottom of insert)	300 µL isopropanol (tissues fully immersed)
Extraction time and temperature	4h at RT (18-23°C) with vortex mixing at the middle of the incubation period, or 72h at 4°C protected from light	Overnight without shaking in dark at RT or 120 min with shaking (~120 rpm) at RT	120±5 min with gentle agitation (~150 rpm) at RT	Overnight (> 15h) in a cold place (or refrigerator, in dark
OD reading	570±30 nm without reference filter	570 nm (540-595 nm) without reference filter	570 nm (540 - 600 nm)	$OD_{570nm} – OD_{650\ nm}$
Tissue Quality Control	18 hours treatment with SDS 1.0 mg/mL ≤ IC_{50} ≤ 3.0 mg/mL	Treatment with 1% Triton X-100 4.0 hours ≤ ET_{50} ≤ 8.7 hours	Treatment with 1% Triton X-100 4.0 hours ≤ ET_{50} ≤ 10.0 hours	18 hours treatment with SDS 1.4 mg/mL ≤ IC_{50} ≤ 4.0 mg/mL
Acceptability Criteria	1. Mean OD of 3 tissue replicates treated with the negative control should be ≥ 0.6 and SD≤18 2. Mean viability of positive control should be ≤ 40% of the negative control and SD ≤18 3. SD of test chemical should be ≤18 (re-tested up to 3 batches max)	1. Mean OD of the tissue replicates treated with the negative control should be ≥ 1.0 and ≤ 2.5 2. Mean viability of positive control should be ≤ 20% of the negative control and within the 95±1% confidence interval of historical data 3. SD of 3 tissue replicates treated with test chemical <18	1. Mean OD of 3 tissue replicates treated with the negative control should be ≥ 1.2 and SD ≤18% 2. Mean viability of positive control should be < 40% of the negative control and SD ≤18% 3. SD of test chemical should be ≤18 (re-tested up to 3 batches max)	1. Mean OD of negative control should be ≥ 0.7 and ≤ 2.5 2. Mean viability of positive control should be ≤ 40% 3. SD of replicates should be ≤18%

3.3.8. Comparison to the in vivo test method

The adopted *in vitro* reconstructed human epidermis methods falling within the OECD TG 439 for assessing skin irritation hazard make use of human keratinocytes, representing therefore the target organ of the species of interest. Morphologically, the adopted *in vitro* RhE test methods are closer to the human epidermis as compared to the rabbit skin. Although these models do not present all functional complexity that exist *in vivo* (including the dermis and its components such as hair follicules, subaceous glands, nerve and immune cells, which could play a role in the mechanisms of skin irritation), the *in vitro* reconstructed human epidermis were found to have similar profiles of phase I and II enzymatic activities as compared to the human skin such as the low expression and function levels of phase I enzymes, and measurable activity of some phase II enzymes (Hewitt *et al.*, 2013). Furthermore, *in vitro* reconstructed human epidermis models using multiple endpoint analyses were shown to have good correlation with the results of the human patch test (Welss *et al.,* 2007). In particular, the RhE test methods were found to better predict the effects on humans than the rabbit test (Jírová *et al.*, 2010). Out of 16 chemicals classified as irritants in the rabbit, only five substances were irritating to the human skin. The concordance of the rabbit test with the 4 hour Human Patch Test (HPT) was only of 56%, whereas the concordance obtained between the RhE test methods and the HPT was of 76% for the EpiDerm™ RhE model and of 70% for the EpiSkin™ RhE model (Jírová *et al.*, 2010). Such findings were confirmed by Basketter and co-authors who showed that the rabbit skin irritation test largely over-predicts human responses to chemicals (Basketter *et al.*, 2012). In their study, the authors show that out of 81 substances found to have HPT data, about 50% were classified as irritating based on the rabbit skin test whereas with the 4h HPT test less than 20% were identified as acutely irritant to human skin.

The main endpoint considered in the OECD TG 439 is cell viability, based on the principle that irritant chemicals are able to penetrate the *stratum corneum* by diffusion and are cytotoxic to the cells in the underlying layers. The *in vitro* test methods cover mainly the initial mechanisms of skin irritation occurring in the *in vivo* test (figure 3.2). However, the evaluation of the release of Interleukin 1α, considered to be a useful adjunct to the MTT assay to increase the sensitivity of the EpiSkin™ SIT assay without reducing specificity (ESAC, 2007), could give additional insight on the release of inflammatory mediators that may act in the subsequent mechanistic cascade of events occurring during skin irritation reactions. Furthermore the use of HPLC/UPLC spectrophotometry allows evaluating the skin irritation potential of coloured test chemicals (Alépée *et al.*, 2016), that could have impaired visual observations in the *in vivo* animal testing.

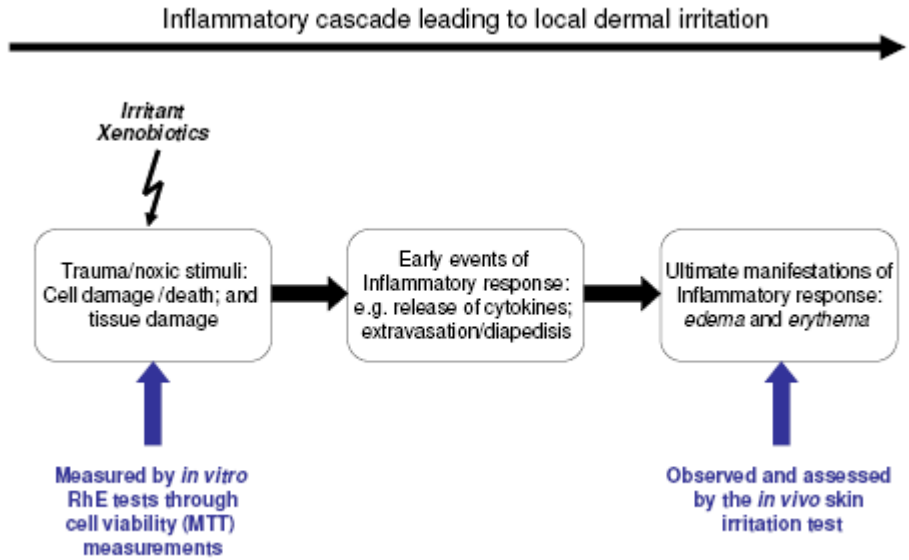

Figure 3.2. Extract from the Explanatory Background Document to the OECD test guideline on *in vitro* skin irritation testing (OECD, 2010a). Schematic representation of the inflammatory cascade leading to local acute dermal irritation.

A summary of the major components of the regulatory *in vivo* and *in vitro* reconstructed human epidermis methods for skin irritation is shown in table 3.11. The exposure times used by the adopted *in vitro* RhE assays are in general shorter as compared to those used *in vivo* (15 - 60 min *in vitro* versus 4 hours *in vivo*). Similarly the post-treatment time is shorter *in vitro* with respect to the *in vivo* test (42 hours versus 14 days). The ability of the RhE test methods to detect skin irritants classified *in vivo* on the basis of persistence only could not be assessed due to the poor availability of such test chemicals in around 5000 screened chemicals from the industrial commerce (Eskes *et al.*, 2007). However, the need to identify such scarce occurring test chemicals classified based on persistence only may be questionable. Furthermore, these differences might be compensated by the more simple structure of the skin components involved in skin irritation reactions present in the *in vitro* models with respect to the *in vivo* situation. Moreover, the doses applied *in vitro* (26 to 47.6 μl or mg/cm^2) are generally smaller with respect to those applied *in vivo* (~83.3 μl or mg / cm^2), with the exception of the LabCyte EPI-MODEL24 SIT, which makes use of similar doses (83.3 μl or mg / cm^2).

Table 3.11. Comparison of the principal method components of the regulatory accepted *in vivo and in vitro* tests for skin irritation

	***In vivo* test for skin irritation (OECD TG 404)**	***In vitro* reconstructed human epidermis (RhE) test methods (OECD TG 439)**
Model used	Albino rabbit.	Three-dimensional reconstructed human epidermis, consisting of organized basal, spinous and granular layers, and a multilayered *stratum corneum.* Surface of tissue models: 0.38 cm^2 for Episkin^{TM}-SIT, 0.63 cm^2 for EpiDerm^{TM} 200-SIT and 0.5 cm^2 for SkinEthic^{TM} SIT.
Number of replicates	2 to 3 animals based on severity of effects.	At least 3 replicates for each test chemical.
Dose and application of test chemical	0.5 ml (liquids) or 0.5 g (solids) applied to ~ 6 cm^2 of skin and covered with a gauze patch (~ 83.3 μl or mg / cm^2). Solids might be moisten to ensure good skin contact.	Liquids: 10 to 30 μl (26 to 83 μl/cm^2 depending on model). Solids: 10 to 25 mg (26 to 83 mg / cm^2 depending on model). Tissues should be moisten prior to solid application to ensure good contact with the RhE.
Controls	Potential influence of the vehicle on irritation of the skin by the test chemical should be minimal, if any.	Negative control: water or PBS Positive control: 5% aqueous SDS
Exposure time	4 hours	15 to 60 min depending on the model (see table 3.10 for details).
Washing	At the end of exposure time to remove test chemical	At the end of exposure time to remove test chemical.
Post-treatment incubation time	If no corrosive effects seen, the animal is observed up to 14 days.	After washing, the exposure time is followed by a post-treatment incubation time of 42 hours to allow for recovery from weak cytotoxic effects as well as for appearance of clear cytotoxic effects.
Endpoint assessed	- Grading of skin reactions. - Other reactions such as: defatting of skin, clinical signs of toxicity and body weight, persistence of alopecia, hyperkeratosis, hyperplasia and scaling. - Histopathology may be carried out in case of equivocal responses.	Cell viability based on the premise that irritant chemicals are able to penetrate the *stratum corneum* by diffusion and are cytotoxic to the cells in the underlying layers. Use of HPLC/UPLC spectrophotometry allows evaluating strongly coloured test chemicals. Inflammatory mediators such as Interleukin 1alpha was considered as a useful adjunct to the MTT assay in order to increase sensitivity of the assay without reducing specificity.
Interpretation of results	Classification systems as shown in section 1.4. In the EU, two categories for skin irritation (irritants and no category). For GHS a third optional category for mild irritants.	Can be used for hazard identification of: - UN GHS Cat. 2 if supported by corrosive negative results, and – No Category in countries not adopting the optional GHS Cat. 3 (mild irritants)
Limitations	- Over-predicts human responses. - Coloured chemicals may interfere with observations. - May be variable between laboratories. - Does not assess repetitive low-dose exposure. - Has the potential to cause discomfort or pain to laboratory animals	- Not designed to distinguish the optional GHS Cat. 3 for mild irritants, corrosive chemicals, gases and aerosols. - Results obtained with test chemicals presenting non-specific interactions with MTT ≥ 50% should be taken with caution when OD is used as measurement for cell viability. This may be circumvented for coloured interfering test chemicals with the use of HPLC/UPLC as an alternative measurement.

Unlike the *in vivo* test, the *in vitro* assays make systematically use of positive and negative controls to check for the functionality of the test method. In addition, the OECD TG 439 recommend also ensuring the technical proficiency of the assays, by the laboratory, prior to the routine use of the *in vitro* assays by testing a list of recommended proficiency chemicals.

In the EU, where the optional UN GHS Category 3 (mild irritants) is not implemented, the adopted *in vitro* assays for skin irritation can be used as a stand-alone assay for identifying test chemicals that do not require classification for skin irritation in case of a negative result (EU CLP No Category). Furthermore, in the case of a positive result, they can be used for the hazard identification of UN GHS Cat. 2 irritant chemicals (substances and mixtures), when test results are supported by a separate non-corrosive outcome (based on e.g., OECD TG 430, 431 or 435). OECD TG 439 does however not allow classifying test chemicals in the optional GHS Cat. 3 as mild irritants, nor does it provide adequate information on skin corrosion.

3.4. Applicability of *in vitro* test methods for the testing of mixtures

3.4.1. Detergent and cleaning products

No publically available literature could be found on the testing of detergents and cleaning products for skin irritation and corrosion having comparative *in vitro* and *in vivo* data. However in the case of skin irritation, a compilation could be made of publically available comparative data obtained with ingredients that could be of interest for detergents and cleaning products. Data on 26 ingredients of potential interest could be found having parallel *in vitro* and *in vivo* data in either the EpiSkin™ SIT, the EpiDerm™ EPI-200-SIT or the SkinEthic™ SIT42bis test method (Alépée *et al.*, 2010; Cotovio *et al.*, 2005; de Brougerolle de Fraissinette *et al.*, 2009; Kandárová *et al.*, 2009b; OECD, 2010a; Tornier *et al.*, 2010). When results of these different models were combined together, an accuracy of 73.3% (44/60), a false positive rate of 36.4% (16/44) and a false negative rate of 0% (0/16) was found for ingredients that could be of interest for detergents and cleaning products. These performances suggest that test methods falling within the OECD TG 439 have rather a tendency to over-predict those types of chemistries as compared to the performances required in the performance standard GD 220 (i.e., false positives $\leq$ 30%), whilst achieving no false negative predictions.

In the case of skin corrosion, data on 149 ingredients could be found having parallel *in vitro* and *in vivo* data in test methods falling within the OECD TG 430, 431 or 435 (Barrat *et al.*, 1998; Fentem *et al.*, 1998; Liebsch *et al.*, 2000; Kandárová *et al.*, 2006, 2011; Hoffmann *et al.*, 2005; Bytheway *et al.*, 2009; NIH, 1999). The obtained performances divided by chemical classes are shown in table 3.12. It is interesting to note that, despite of the small number of tested chemicals, inorganic acids and inorganic bases seem to have a tendency to be over-predicted by all *in vitro* test methods. In addition, Houthoff and co-workers (2015) reported that the specific class of fatty amine derivatives, characterized as cationic surfactants, were found to be under-predicted for skin corrosion testing with test methods falling within the OECD TG 431. A total of 7 out of 8 corrosive fatty amine derivatives were under-predicted by the Rhea models, representing 3 UN GHS Sub-category 1C and 4 UN GHS Sub-category 1B under-predicted as non-corrosives, whereas one UN GHS Sub-category 1B was correctly predicted as a UN GHS Sub-category 1B-and-1C. In particular these substances were reported to induce corrosive effects *in vivo* after some delay. The authors suggest the use of modified protocols to potentially identify if a delayed response may happen *in vitro*. Indeed an extended exposure period of 4h is recommended for this class of chemicals for the EpiDerm™ SCT falling within the OECD TG 431 (Kandárová and Liebsch, *in press*).

Table 3.12. Performance of adopted test methods for *in vitro* skin corrosion testing of ingredients that could be of interest for detergents and cleaning products

	OECD TG 431 - RhE		OECD TG 430 - TER		OECD TG 435 - Corrositex®	
	False +	False -	False +	False -	False +	False -
Acid derivatives	-	0/1	-	0/1	1/4	0/5
Amines	-	-	-	-	5/7	1/7
Electrophiles	0/9	1/5	1/5	2/3	-	-
Inorganic acids	2/2	1/11	1/1	0/6	4/4	0/18
Inorganic bases	1/3	1/6	2/2	0/2	7/8	0/5
Inorganic salts	0/2	0/1	1/2	0/1	1/4	0/3
Neutral organics	1/16	-	0/9	-	-	-
Organic acids	0/9	0/12	0/5	0/7	6/12	4/23
Organic bases	2/8	4/14	1/3	1/7	6/9	2/14
Phenols	1/6	0/5	0/3	0/2	-	-
Surfactants	1/5	-	2/3	-	-	-

3.4.2. Other mixtures

Regarding the applicability of *in vitro* test methods for skin irritation testing of mixtures, Molinari and co-workers (2013) reported that the use of an optimised washing procedure allowed the testing ***sticky and greasy natural botanicals*** with appropriate performance (6/7 correct non-irritant predictions, 3/3 correct irritant predictions and 2/2 correct classified predictions for 2 skin corrosive ingredients). Regarding agrochemical formulations, contradictory and limited information was reported on the applicability of test methods falling within the OECD TG 439, suggesting that further investigations are needed (Eskes *et al.*, 2012; Kolle *et al.*, 2013). Finally, recent studies reported that test methods falling under TG 439 seem suitable to test medical device extracts (Casas *et al.*, 2013) and nanoparticles (Miyani et. 2016; Kim *et al.*, 2016), however no correlation was made between the obtained *in vitro* and *in vivo* effects.

Regarding skin corrosion testing in contrast, agrochemical formulations were reported to have similar predictions as those obtained with general substances when using test methods falling within the TG 431, however only non-corrosive formulations were tested (Kolle *et al.*, 2013).

3.5. Comparison of OECD adopted skin irritation and corrosion test methods

Table 3.13 provides an overview of the currently *in vitro* test methods adopted by the OECD, their regulatory use, applicability and limitations, performance and role within an IATA. Furthermore table 3.14 provides with a comparison of the main protocol characteristics of the existing *in vivo* and *in vitro* OECD adopted test methods for skin irritation and corrosion hazard identification.

3.6. Other *in vitro* information sources

A number of similar RhE models for skin irritation testing to the ones already adopted within the OECD TG 439 have been developed. Among those three models have undergone a catch-up multi-laboratory validation study based on the Performance Standards (PS) as defined in the OECD Guidance Document No. 220 (OECD, 2015h). These are:

- the commercially available epiCS SIT model which underwent a positive independent peer-review (Engelking *et al.*, *in press*),
- an open-source RhE model (OS-Rep), having an openly accessible protocol for tissue production, that underwent a PS-based validation study in which each participating laboratory made use of their in-house generated OS-Rep to assess the set of PS reference chemicals (Mewes *et al.*, 2016; Groeber *et al.*, 2016), and

- the commercially available Skin + ® RhE test system produced by Sterlab and commercialized by ATERA, which also underwent a PS-based validation study.
Other similar models developed for skin irritation testing include the Leiden human epidermal (LHE) model that showed similar skin irritation results with the 20 reference chemicals to those reported for the validated skin models (El Ghalbzouri *et al.*, 2008), a model based on human skin obtained from surgery (Miles *et al.*, 2014), and a viable human full thickness skin model (Lönnqvist *et al.*, 2016). Similar test methods for skin corrosion testing to the ones falling within TG 431 are also being developed in e.g. different geographical regions such as India (Deshmukh *et al.*, 2012).

Assays based on the measurement of parameters other than cell viability are also being developed and show promise to distinguish not only irritants from non irritants but also to determine the skin irritancy potential of chemicals. For example, the IRR-IS assay exploiting the quantitative analysis of expression profiles of relevant genes appears to be a promising methodology to contribute to the discrimination of non-irritants (No Cat.), mild-irritants (Cat. 3) and irritants (Cat. 2) as shown in a study evaluating gene expression changes in the validated EpiSkin^{TM} test system in response to chemical exposure (Groux et al, 2012). Furthermore, use of biomarkers such as IL-1α, IL-1RA, IL-8 and MTT in a reconstructed epidermis model was shown to determine the skin irritant potency of chemicals in addition to distinguishing irritants from non irritants (Spiekstra *et al.*, 2009). Other endpoints investigated include the use of proteomics (Zhang *et al.*, 2011) and toxicogenomics (Niwa *et al.*, 2009; Borlon *et al.*, 2007).

Finally, attempts have also been made to develop an innervated *in vitro* model of human skin including sensory neurons derived from embryonic rat dorsal root ganglion as neural components (Khammo *et al.*, 2007). The aim was to integrate the sensory neuronal components which are usually present in the skin and may play a role *in vivo* in the production of neurogenic inflammation leading to sensory irritation and pain (Garle and Fry, 2003).

Table 3.13. Overview of the regulatory use, applicability, limitations and performance of the OECD adopted *in vitro* test methods for skin irritation and corrosion hazard identification.

	***In vitro* skin corrosion: TER test method (OECD TG 430)**	***In vitro* skin corrosion: RhE test method (OECD TG 431)**	***In vitro* skin corrosion: membrane barrier test - Corrositex® (OECD TG 435)**	***In vitro* skin irritation: RhE test method (OECD TG 439)**
UN GHS Cat. prediction	**UN GHS Cat. 1** vs. non-corrosives	**UN GHS Cat. 1** vs. non-corrosives **UN GHS / EU CLP Sub-cat. 1A** and a combination of **Sub-cat. 1B-and-1C**	**UN GHS Cat. 1** vs. non-corrosives **UN GHS / EU CLP Sub-cat. 1A, Sub-cat. 1B and Sub-cat. 1C**	**UN GHS / EU CLP Cat. 2** when supported by corrosive negative results, and **EU CLP No Category**
Applicability	Substances and mixtures (as an extension to the applicability to substances)	Substances and mixtures (as an extension to the applicability to substances)	Substances and mixtures	Substances and mixtures (as an extension to the applicability to substances)
Accuracy	81% (99/122) (a)	≥ 82.5% (b)	79% (128/163) (c)	≥ 75% (d)
False negatives	6% (3/54) (a)	< 5% (b)	15% (13/89) (c)	< 20% (d)
False positives	29% (20/68) (a)	< 30% (b)	30% (22/74) (c)	< 30% (d)
Reliability	≥ 80% BLR and ≥ 90% WLR concordance of classification	≥ 80% BLR and ≥ 90% WLR concordance of classification	≥ 93% BLR concordance of classification, CV ≤ 30% BL and ≤ 5% WL	≥ 80% BLR and ≥ 90% WLR concordance of classification
Limitations	- Not able to distinguish the three GHS Sub-categories (1A, 1B and 1C) - Not designed to provide information on skin irritation - Not applicable to gases and aerosols - May be considered an animal test in some countries	- Does not allow discrimination between skin corrosive Sub-cat. 1B and 1C - Not designed provide information on skin irritation - Not applicable to gases and aerosols - Results obtained with test chemicals having non-specific interactions with MTT ≥ 50% should be taken with caution when OD is used as measurement for cell viability. This can be circumvented for coloured-interfering chemicals by using HPLC/UPLC.	- Not designed to give information on skin irritation - Not applicable to gases and aerosols - Test chemicals not causing detectable changes in the chemical detection system cannot be tested (e.g. aqueous test chemicals with a pH in the range 4.5-8.5 often do not qualify (NIH, 1999)) - In EU, considered valid only for acids, bases and their derivatives	- Not designed to distinguish the optional GHS Cat. 3 for mild irritants, corrosive chemicals - Not applicable to gases and aerosols - Results obtained with test chemicals presenting non-specific interactions with MTT ≥ 50% should be taken with caution when OD is used as measurement for cell viability. This can be circumvented for coloured-interfering chemicals by using HPLC/UPLC.
Role in the IATA	If non-corrosive outcome, an *in vitro* skin irritation test should be conducted. Furthermore if need for corrosive sub-categorization, other *in vitro* test methods should be used.	If non-corrosive outcome, an *in vitro* skin irritation test should be conducted. Furthermore if need for discrimination between Sub-cat. 1B from Sub-cat. 1C, OECD TG 435 should be considered, or the EpiSkin™ non-adopted prediction model for that purpose may be considered in a WoE approach.	If non-corrosive outcome, an *in vitro* skin irritation test should be conducted.	If positive result, an *vitro* skin corrosion test should be conducted. For countries adopting the optional UN GHS Cat. 3, additional testing in an *in vitro* skin irritation test method not adopted by the OECD may be considered in a WoE approach before an *in vivo* test is performed.
Additional non-regulatory information	Not reported	Fatty amine derivatives (characterized as cationic surfactants) risk under-prediction (Houthoff *et al.*, 2015), so that an extended exposure time is recommended (Kandárová and Liebsch, *in press*).	Not reported	Seems applicable to natural botanicals with an optimized washing procedures, and to the testing of medical device extracts (Molinari *et al.*, 2013; Casas *et al.* 2013)

(a) From the validation dataset and other published studies (OECD TG 430, 2015b); (b) Based on the Performance Standards requirements (OECD GD 219, 2015f); (c) From the validation dataset (OECD TG 435, 2015c); (d) Based on the Performance Standards requirements (OECD GD 220, 2015h)

Table 3.14. Comparison of the main protocol steps of the existing *in vivo* and *in vitro* OECD adopted test methods for skin irritation and corrosion hazard identification.

	***In vivo* skin corrosion & irritation test (OECD TG 404)**	***In vitro* skin corrosion: TER test method (OECD TG 430)**	***In vitro* skin corrosion: RhE test method (OECD TG 431)**	***In vitro* skin corrosion: membrane barrier test - Corrositex® (OECD TG 435)**	***In vitro* skin irritation: RhE test method (OECD TG 439)**
Model used	Albino rabbit (surface of exposure: 6 cm^2 of skin)	Skin disks prepared from young rats, where 10-15 skin discs can be obtained per rat skin (0.79 cm^2 surface)	3D reconstructed human epidermis (0.38 to 0.63 cm^2 surface depending on the model).	A synthetic macromolecular bio-barrier and a chemical detection system (CDS) which indicates the presence of a test chemical.	3D reconstructed human epidermis (0.3 to 0.63 cm^2 surface depending on the model).
Commercially available methods accepted	Not applicable	Not applicable	- EPISKIN™ Standard Model (SM) - EpiDerm™ Skin Corrosion Test (SCT) - SkinEthic™ RHE - epiCS® (previously named EST-1000)	- Corrositex®	- EPISKIN™ Skin Irritation Test (SIT) - EpiDerm™ SIT - SkinEthic™ SIT[42bis] - LabCyte EPI-MODEL24 SIT
Number of replicates	1 to 3 animals based on severity of effects	At least 3 skin disks	At least 2 replicates for each exposure time	2 repeats in 2 batches	At least 3 tissue replicates
Dose and application of test chemical	0.5 ml or 0.5 g (~ 83 µl or mg / cm^2 skin)	Liquids: 150 µl (~ 190 µl/cm^2) Solids: sufficient amount to cover surface	Liquids: 40-50 µl (79.4 to 131.6 µl/cm^2) Solids: 20-25 mg (39.7 to 52.6 mg / cm^2)	0.5 ml or g applied on membrane	Liquids: 10-30 µl (26 to 83 µl/cm^2) Solids: 10-25 mg (26 to 83 mg / cm^2)
Exposure time	Corrosion: 3 min, 1 hour, 4 hours depending on observed results	24 hours (RT)	3 min (RT), 1 hour (RT or 37°C depending on the model) and 4h for the Episkin™ model only (RT)	The time needed for a test chemical to penetrate the membrane barrier is used to predict corrosivity with the following cut-offs: 3 min, 30 or 60 min. (depending on predicted acidic/alkaline reserve), 60 or 240 min (depending on predicted acidic/alkaline reserve), >60 or >240 min (depending on predicted acidic/alkaline reserve)	15 to 60 min(RT or 37°C depending on the model)
Post-incubation period	14 days	n.a.	n.a.	n.a.	42 hours (37°C)
Negative control	-	H_2O	H_2O or NaCl (9 g/l)	10% citric acid or 6% propionic acid	(D)PBS or H_2O
Positive control	-	10M HCl	8 KOH or glacial acetic acid	NaOH or other	5% aq. SDS

4. Other information sources

4.1. Existing human data

4.1.1. Applicability, limitations and role within the IATA

New testing in humans for hazard identification purposes is not acceptable for ethical reasons. However existing human data retrieved from a number of sources, should be taken into account when evaluating the intrinsic hazards of substances and mixtures (ECHA, 2015a, 2015b). Two different types of human data need to be considered, namely non-standardised human data on local skin effects and data obtained from standardised skin irritation human patch testing (HPT). While the first is usually associated with a high level of uncertainty, the latter is commonly of much higher quality as it is usually acquired under standardised conditions and with strict acceptance criteria (OECD, 2014). An overview of these two types of human data is given in table 4.1 and a detailed description given below.

The quality and relevance of existing human data for hazard assessment should be critically reviewed (ECHA, 2015a). There may be a significant level of uncertainty in human data due to poor reporting and lack of specific information on exposure (ECHA, 2015b). However, well-documented existing human data from various sources can often provide very useful information on skin corrosion and irritation. The usefulness of human data will depend on the extent to which the effect, and its magnitude, can be reliably attributed to the test chemical of interest (ECHA, 2015a).

If considered suitable and adequately documented, existing human data (especially HPT) should have precedence over other data (OECD, 2014; UN, 2015a). However, although human data from accident or poison centre databases can provide evidence for classification, absence of incidents is not itself evidence for no classification as exposures are generally unknown or uncertain (UN, 2015a). In particular it does not necessarily overrule *in vitro* data or existing animal data of good quality that are positive (ECHA, 2015a). Examples of how existing human data can be used in hazard classification for irritancy are provided in ECETOC publications (ECETOC, 2002; ECETOC, 2009).

4.1.2. Human Patch Test (HPT)

Existing human data from skin irritation human patch testing (HPT) might also be available. HPT is a controlled study involving the exposure of small patches of skin of human volunteers to chemicals for which skin corrosion and other unacceptable toxicological hazards can be excluded. HPT data have been compiled for example by Jírová *et al.* (2010), Basketter *et al.* (2012), as well as Ishii *et al.* (2013). Testing with human volunteers to obtain primary hazard data on skin corrosion/irritation for regulatory purposes is discouraged. Available good quality data should nevertheless be considered as appropriate and used for C&L decision making. It should however be noted that GHS does not contain clear criteria for classification for skin irritation based on human data.

For human patch testing several high quality studies exist (Basketter *et al.*, 1994; Hall-Manning *et al.*, 1995; York *et al.*, 1996; Basketter el al., 1997; Robinson *et al.*, 1998; Robinson *et al.*, 2001; Basketter *et al.* 2004; Robinson *et al.*, 2005; Jírová *et al.*, 2007; Jírová *et al.*, 2010; Basketter *et al.*, 2012; Ishii *et al.*, 2013). The issue of use of human data has been discussed at OECD several times but did not yet result in any concrete action. A Test Guideline on HPT was proposed in 1997 and proposals for inclusion of human data in validation studies have also been discussed without success. However, OECD TG 439 (OECD, 2015d) does include references to human data in the form of HPT test results, in particular in the associated Performance Standards based on the EURL ECVAM Performance Standards for *in vitro* skin irritation testing using Reconstructed human Epidermis (RhE).

Table 4.1. Overview on the uses and applicability of existing human data within an IATA context

	Human Patch Test	**Non-standard data (clinical, occupational, poison centres, case reports, epidemiological studies)**
Mode of Action	All modes of action are potentially covered	
Applicability Domain	- Originally developed for cosmetics and household products, it was further applied to medical devices (ISO 10993-10) and for testing of chemicals - Not applicable to corrosives & unacceptable toxic hazards - Coloured materials may impair scoring	All test chemicals for which a clear and direct effect on the skin can be concluded from the data available
Predictive Capacity	Highly predictive of effects on humans	Depends on the amount and quality of available information. Usually associated with a high level of uncertainty due to lack of critical information (e.g., chemical identity, purity, exposure, health status of exposed persons, reported symptoms)
Reliability	At least similar to TG 404	Difficult to assess due to uncontrolled exposures and reporting
Strengths	- Species of interest - Highly predictive - Usually, standardized, high quality	- Species of interest - May provide information on cumulative effects
Limitations	- Should not be conducted primarily for C&L purposes - No UN GHS criteria for C&L based on human data available - Differences in population	- Not standardized - Often associated with a high level of uncertainty, unclear quality and/or incomplete data - No UN GHS criteria for C&L based on human data available - Differences in population
Role in IATA	- If high quality HPT result exists, it represents strongest basis for C&L - If different than animal result and WoE inconclusive -> *in vitro* testing - Should not be included in a strategy as a prospective testing option	- Should be used in WoE approach with other existing data - Should not over-rule existing OECD TGs based on *in vitro* and *in vivo* studies

4.1.3. Non-standardised human data on local skin effects

Existing human data on local skin effects originate from clinical and occupational studies, poison information centres, case reports and retrospective epidemiological studies. They provide information directly related to effects on the skin i.e., local skin effects, following single or repeated exposure. The exposure could be of accidental nature or prolonged (i.e., cumulative), for example in occupational settings, but it is often difficult to quantify. As such, although human data from accidents or poison centre databases can provide evidence for classification, absence of incidents is not itself evidence for

no classification as exposures are generally unknown or uncertain. It can also be anticipated that this type of human data is available in exceptional cases only and, when available, the quality, reliability and relevance of the existing data for hazard assessment should be critically reviewed before any regulatory decision is taken. Indeed, there may be a significant level of uncertainty in human data on local skin effects due to poor reporting and lack of specific information on exposure (dose and duration) and other critical aspects. For example, in case reports, information on chemical identity and purity, exposure, health status of the persons exposed and even the symptoms reported is often lacking. Specific limitations of poison centre data have been summarised by Hoffman (2007). Existing human data on local skin effects may be particularly relevant when they demonstrate effects which cannot be observed in experimental animal studies. As animal studies are designed to assess irritation as a result of acute exposure only, human data may in particular provide useful information on the cumulative effects leading to irritation (Irritant Contact Dermatitis) in humans.

It should be possible to discern corrosive properties of chemicals from mere irritation in humans based on existing human data on local skin effects, if a follow-up of the initial assessment after the accidental exposure is available. Corrosive reactions are typified by ulcers, bleeding and bloody scabs and, after recovery, the skin will be discoloured due to blanching of the skin, complete areas of alopecia and scars (see Chapter 3.2 of GHS, defining skin corrosion based on effects observed in the *in vivo* rabbit test), i.e., skin corrosion is an irreversible damage. However, human data are usually not sufficient to sub-categorise chemicals according to their corrosion potential, e.g., UN GHS Sub-categories 1A, 1B and 1C, as required in some regulatory frameworks and legislations. A clear case for Sub-cat. 1A classification (corresponding to 3 minutes in rabbits) would be an accidental splash which gave rise to necrosis of the skin. In cases where a prolonged exposure was needed before necrosis occurred (not to be confused with delayed effects), Sub-cat. 1B-and-1C seems more reasonable. The distinction between Sub-cat. 1B and Sub-cat. 1C (corresponding to 1 hour and 4 hours exposure in rabbits, respectively) may not be so obvious in practice. If the distinction between Sub-cat. 1A and Sub-cat. 1B-and-1C is not clearly apparent then a simple classification as UN GHS Cat. 1 (without sub-categorisation) should be used.

***4.2. In vivo* Acute Dermal Corrosion and Irritation test (OECD TG 404)**

4.2.1. Applicability, limitations and role within the IATA

In case of existing data based on the animal test falling within the OECD TG 404 (2015a), that is of adequate quality, these should carry a certain intrinsic weight in the context of a weight of evidence (WoE) analysis. Otherwise, the animal test falling within OECD TG 404 should be used only as a last option, after *in vitro* testing (including the use of *in vitro* test methods not adopted by the OECD), for (i) discrimination between optional Sub-categories 1B and 1C for chemicals outside of the applicability domain of OECD TG 435 when required, (ii) discrimination of optional UN GHS Cat. 3 from UN GHS No Cat. when required, or (iii) when the test chemical cannot be tested with the *in vitro* test methods currently adopted by the OECD due to limitations or non-applicability. It may in exceptional cases also be used, when *in vitro* testing is not feasible or reliable.

A wide range of chemicals (substances and mixtures) can be tested according to the OECD TG 404. The OECD TG 404 reflects all possible modes of action of skin irritant and corrosive reactions present in rabbit skin, allows to assess reversibility of effects and has been used as the basis for the traditional classification systems, so that it can provide classifications over the full range of skin irritation and corrosion effects (i.e. UN GHS No Cat., Cat. 3, Cat. 2, Sub-cat 1C, Sub-cat. 1B or Sub-cat. 1A).

The OECD TG 404 is however not applicable to the testing of gases and aerosols. Furthermore, dyes and other coloured chemicals may impair the scoring of effects, especially erythema. Similarly, physico-chemical properties such as volatility may considerably reduce the amount of chemical in contact with skin. Nevertheless, the chemical will also be volatile in a potential human exposure situation.

Finally, the animal testing falling within OECD TG 404 tend to over-predict human skin corrosion and irritation and present a number of issues that can reduce the reproducibility of the test method (e.g., the subjective grading of skin responses, no standardised procedure for test chemical removal, etc – see chapter 4.2.3 for details). Furthermore, it represents an animal experiment which may potentially involve suffering due to the corrosive or the inflammatory reactions (pain, itching, etc.).

4.2.2. Method description according to OECD TG 404

The procedure is described in detail in the OECD TG 404 (2015a). Briefly, the principal steps of the *in vivo* testing are the following ones.

a) Animals used
Albino rabbit is the preferable laboratory animal.

b) Dose and application of the test substance
A dose of 0.5 ml of liquid or 0.5 g of solid or paste is applied.

The test substance is applied to a small area (approximately 6 cm^2) of skin and covered with a gauze patch, which is held in place with non-irritating tape.

Liquids are generally tested undiluted. Solids should be moistened with water to ensure good skin contact. When vehicles other than water are used, the potential influence of the vehicle on irritation of the skin by the test substance should be minimal, if any.

c) Sequential testing
An initial test using one animal is recommended. If no corrosive effects are observed but an irritant effect is observed, a confirmatory test using additional one or two animals may be conducted in a sequential manner. In any case an irritant or negative response in the initial test should be confirmed using up to two animals.

d) Exposure time
If a substance is not expected to produce corrosion but may be irritating, a single patch should be applied to one animal for four hours, and then proceed the confirmatory testing. If a substance was suspected of being corrosive but testing was still warranted after considering the IATA (OECD, 2014) and performing prospective *in vitro* testing, up to three patches are to be applied sequentially to the animal. The first patch is removed after three minutes, if no serious reactions were observed, a second patch is applied at a different site and removed after one hour. If observations indicated that exposure could humanely be allowed, a third patch is applied and removed after four hours.

e) Observation period
If no corrosive effects are observed after patch removal, the animal is observed for 14 days or until reversibility is seen. If corrosive effect were observed, the test should be immediately terminated.

f) Grading of skin reactions
Animals are examined for signs of erythema and oedema and the responses are scored at 60 min, and then at 24h, 48h and 72h after patch removal as described in table 4.2 (OECD, 2015a). The dermal irritation scores are to be evaluated in conjunction with the nature and severity of lesions and their reversibility or lack of reversibility. Examples of other observations which can be made from the study include:
- All local toxic effects such as defatting of the skin
- Any systematic adverse effects such as effects on clinical signs of toxicity and body weight
- Persistence of responses such as alopecia (limited area), hyperkeratosis, hyperplasia and scaling (in this case the substance should be considered an irritant.
Finally, histopathological examination may be carried out to clarify equivocal responses.

Table 4.2: Grading of skin reactions according to OECD TG 404 (2002)

Erythema and Eschar Formation	
No erythema	0
Very slight erythema (barely perceptible)	1
Well defined erythema	2
Moderate to severe erythema	3
Severe erythema (beef redness) to eschar formation preventing grading of erythema	4
Maximum possible: 4	

Oedema Formation	
No oedema	0
Very slight oedema (barely perceptible)	1
Slight oedema (edges of area well defined by definite raising)	2
Moderate oedema (raised approximately 1 mm)	3
Severe oedema (raised more than 1 mm and extending beyond area of exposure)	4
Maximum possible: 4	

4.2.3. Limitations of the OECD TG 404 (Draize rabbit test)

The Draize rabbit test for skin corrosion and irritation was originally developed and included in the guidelines with the purpose to identify chemicals that posed a severe hazard to the public. The test has been used for about half a century since its introduction in 1944 (Draize *et al*., 1944), and provided value in warning consumers, workers, and manufacturers of potential dangers associated with specific chemicals so that appropriate precautions could be taken. The Draize rabbit test was however not originally developed to compare products. In addition, at that time the current scientific standards required for the validation and evaluation of test methods such as those outlined in the OECD Guidance Document 34 since did not exist (OECD, 2005). Perhaps because of this, the Draize rabbit skin corrosion and irritation test has often been a target for criticism due to the several drawbacks it may present as described below.

It is generally recognised that the Draize method has errs on the safe of safety in that it over-predicts the severity of skin damage produced by chemicals in humans. Human skin is reported to be in most cases, less sensitive than rabbits (ECETOC, 2002). York and co-workers (1994) have shown that over 50% (8 out of 15) materials classified as irritants or corrosives based on the Draize rabbit test, did not show effects on humans using the human patch test. Similarly, Robinson and co-workers (2000) have showed the case study of a substance classified as corrosive with the *in vivo* skin corrosion and irritation test, whereas the *in vitro* and human studies showed no effects or irritation. More recently available Human Patch Test data seem to confirm this (Jírová *et al.*,2010; Basketter *et al.*, 2012; Ishii *et al.*, 2013). Finally, Hoffmann and co-workers (2005) have shown that the practical use of the European classification system seems to introduce a bias by itself towards over-classification of those chemicals having Draize scores close to but below the threshold for assigning skin irritation classification.

Although, no studies assessing the intra- and inter-laboratory variability in a comprehensive way exist on the Draize rabbit acute skin irritation and corrosion test (OECD, 2014), scientific concerns about its variability have been raised (Worth and Cronin, 2001a; Weil and Scala, 1971). Indeed, classification based on animal results between studies may vary due to subjective scoring, dosing by weight (i.e., not taking into account density differences), insufficiently standardised washing procedures, etc. In particular, Weil and Scala (1971) have shown that considerable variation existed between laboratories. The irritation scores given by the participating laboratories were shown to vary from the lowest non-irritation extreme to the most severe irritation/corrosion extreme in 3 out of 10 tested materials (figure 4.1). Moreover the authors found that some laboratories consistently rated materials more irritating while other laboratories just as consistently rated the same materials less irritating than the majority of the 30 participating laboratories, suggesting a subjective bias in scoring. However, an old protocol of the Draize skin irritation test was used in this study (with e.g. 24 hours exposure to the

test chemical), and it is also not clear whether all laboratories have applied the same test protocol or variants were used (Weil and Scala, 1971). As a consequence although such findings indicate potential sources of variability, they cannot be transferred to the OECD TG 404.

MINIMUM-MAXIMUM SCORE FOR REFERENCE PROCEDURE SKIN TEST—INDIVIDUAL RABBITS

Laboratory No.*	Material										Any material
	E	F	G	I	J	K	L	M	N	O	
12	0–4	21–46	2–8	1–10	5–13	7–14	41–46	12–46	4–12	6–12	0–46
14	3–10	5–22	2–7	1–5	1–22	11–16	17–28	3–46	4–10	2–8	1–46
11	5–7	6–44	0–8	0–5	6–25	6–11	12–46	8–29	0–6	3–9	0–46
13	2–19	20–46	2–16	2–30	3–28	11–23	22–36	0–4	All 0	0–9	0–46
10	0–4	23–31	1–5	2–6	4–12	6–13	20–42	2–34	0–24	0–5	0–42
25	2–9	4–44	0–3	0–3	5–23	4–9	30–42	2–35	0–6	1–5	0–44
31	0–8	31–33	0–20	0–6	1–22	6–32	All 29	0–32	All 0	0–4	0–33
23	0–4	7–30	0–3	2 15	6 25	24 27	14 31	0 24	0 1	0 3	0 30
26	1–8	6–13	1–8	2–6	1–3	7–11	6–11	3–10	2–9	2–7	1–13
5	5–7	6–11	0–4	0–1	5–9	5–9	9–12	3–5	1–4	1–4	0–12
9	1–4	6–16	2–8	0–2	7–13	9–12	15–16	0–12	All 0	0–5	0–16
2	4–9	21–38	0–1	All 0	3–7	4–8	17–21	0–25	0–3	0–2	0–38
19	3–8	6–10	1–5	1–4	2–6	4–8	10–26	1–4	0–3	0–2	0–26
24	1–4	5–9	0–2	0–3	4–8	4–7	8–11	0–4	0–3	1–4	0–11
22	0–2	8–32	0–2	1–3	4–15	9–25	25–32	0–1	0–1	0–1	0–32
7	2–5	2–6	All 0	2–19	0–21	1–20	1–21	0–2	0–2	0–2	0–21
8	2–7	2–13	0–3	0–4	4–7	4–6	3–11	0–4	All 0	0–2	0–13
4	0–3	3–6	0–1	1–3	2–10	4–9	10–13	1–4	0–1	All 0	0–13
18	0 1	0–6	All 0	All 0	0–5	2–9	9–19	0–18	0–3	0–7	0–19
30	0–2	1–5	0–1	0–9	3–7	4–8	6–13	0–2	0–1	1–3	0–13
27	0–1	5–12	0–1	0–2	0–9	0–8	8–12	0–4	All 0	0–3	0–12
21	All 0	All 0	All 0	All 0	All 0	All 0	All 0	0–6	All 0	All 0	0–6
Any laboratory	0–19	0–46	0–20	0–30	0–28	0–32	0–46	0–46	0–12	0–12	—

* Laboratories ordered by sum of ranks for primary irritation as in Table 47.

Figure 4.1. Extract from Weil and Scala (1971). Variability between-laboratories on the observed skin corrosion and irritation scores for individual rabbits (minimum score=0, maximum score = 46).

On the other hand, Hoffmann and co-workers (2005) have evaluated the *in vivo* skin irritation data for around 3000 chemicals registered in the 'New Chemicals Database' of the EC European Chemicals Bureau as notified from the 80's which made use of a more recent version of the OECD TG 404. The authors have shown that the within-test variability of Draize skin irritation test rarely resulted in misclassification. However, some principle aspects of within- and between-laboratory variability could not be assessed (Hoffmann *et al.*, 2005). Another study show that high variability was also evident in the database from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) on chemicals containing high quality data for skin irritation produced with the standardised OECD TG 404 and following Good Laboratory Practices (GLP; see figure 4.2). The authors suggest that variations may be due to subjectivity in scoring or to the intrinsic variability of responses in animals, or to both factors (OECD GD 137, 2010a).

A more recent analysis assessed the possibility of reducing the number of rabbits tested for corrosion (UN GHS Cat. 1) or irritation (UN GHS Cat. 2) from 3 to 2 based on within-test variability (Hoffmann, 2011). The study showed low variability for identification of skin corrosion, where reduction of testing from 3 to 2 animals would have no impact on classification. However, the reliability of OECD TG 404 to sub-categorise corrosive chemicals to UN GHS Sub-categories 1A, 1B and 1C has not been formally evaluated; and experience shows that the distinction between Sub-categories 1B and 1C from *in vivo* data often proves to be difficult, resulting in a limited set of well-known Sub-category 1C chemicals. The study also showed that variability was somewhat higher for skin irritation, where reduction of testing from 3 to 2 animals would have some impact on classification for skin irritation due to variability between animals.

Another criticism to the *in vivo* skin irritation and corrosion testing is that the adverse skin responses associated with repetitive, low-dose exposure to industrial chemicals and consumer products are not predicted accurately by the current regulatory assays (Patrick and Maibach, 1994). Indeed, chemically induced skin irritation can be divided into acute, cumulative and delayed acute skin irritation, where the cumulative type is the most common skin irritation and arises after repetitive exposure to mild

irritants. It often occurs in humans who do repetitive wet work and subsequently is often a cause of occupational skin disease (Welss *et al.*, 2004).

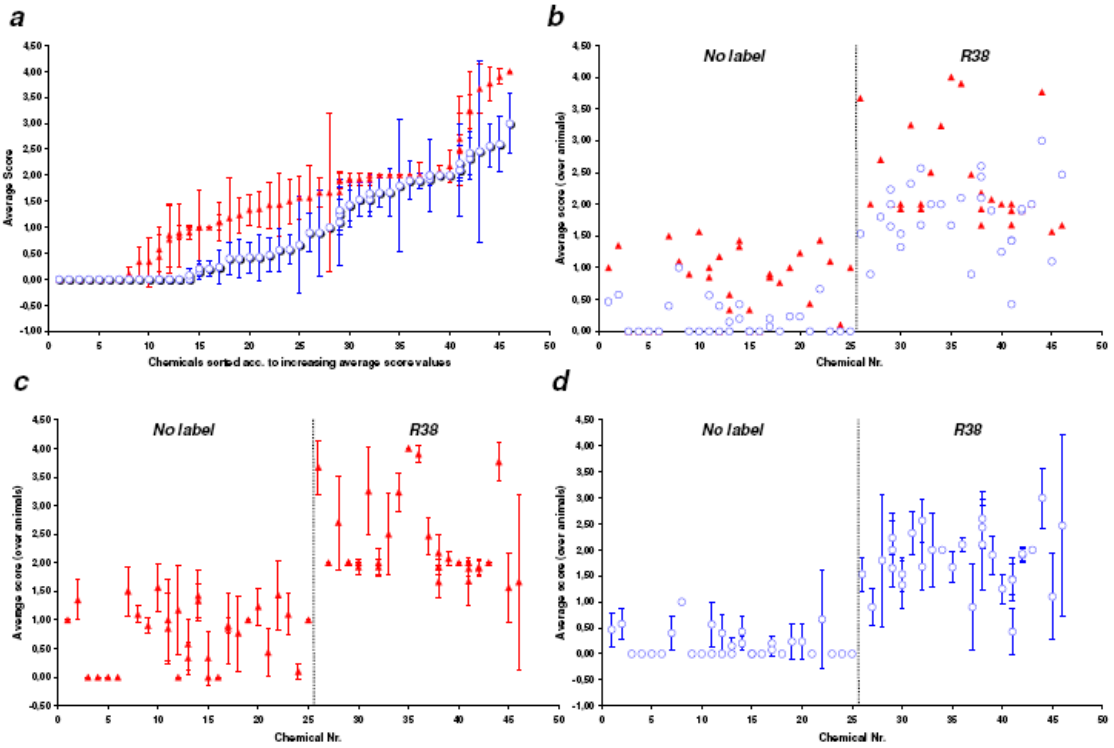

Figure 4.2. Extract from the Explanatory Background Document to the OECD Draft test guideline on *in vitro* skin irritation testing (OECD GD 137, 2010a). Erythema (red triangles) and Edema (blue circles) irritancy scores of 45 chemicals from the ECETOC skin irritancy dataset produced in agreement with OECD TG 404 and under GLP. (a) Averages of animal scores ± Standard Deviation (SD) of test results. (b-c) Scatter plots of the averages of animal scores with chemicals plotted according to the number attributed in the ECETOC report. Considerable variability is observed in particular of oedema test results (circles) for irritant (R38). The stippled line separates label from R38 substances. Average erythema and oedema scores without SD (b), average erythema (c) and oedema (d) scores across animals (± SD).

Finally, from an ethical point of view, testing for skin corrosion and irritation in laboratory animals has the potential to cause them considerable discomfort or pain. For these reasons, several alternative methods for trying to identify skin corrosives and irritants have been developed, validated and accepted for regulatory use.

4.3. Data from other dermal exposure studies

Other *in vivo* or *in vitro* toxicity data of dermal exposure may provide additional information regarding the skin effects of a test chemical. These may represent existing data or data to be generated in some regulatory frameworks to satisfy other regulatory requirements than skin irritation and corrosion. Within the IATA, if consideration of the existing data (human, animal and *in vitro* data), physico-chemical properties and non-testing data on skin irritation and corrosion is inconclusive, these other *in vivo* or *in vitro* dermal toxicity tests, for which data are not yet available but may need to be conducted in some regulatory frameworks to satisfy other regulatory requirements, should be carried out before conducting prospective testing for skin irritation and corrosion. Once available, these additional test results should be incorporated into a WoE analysis.

Data on other dermal exposure studies may be derived from one or more of the following OECD TG's:

- OECD TG 402 Acute Dermal Toxicity (OECD, 1987)
- OECD TG 406 Skin Sensitisation (Guinea Pig Maximisation Test and Buehler Test) (OECD, 1992)
- OECD TG 410 Repeated Dose Dermal Toxicity Study, 21/28 days (OECD, 1981a)
- OECD TG 411 Subchronic Dermal Toxicity Study, 90 days (OECD, 1981b)
- OECD TGs 429, 442A, 442B Skin Sensitisation, LLNA protocols (OECD, 2010b,c,d)
- OECD TG 427 Skin Absorption: *in vivo* Method (OECD, 2004a)
- OECD TG 428 Skin Absorption: *in vitro* Method (OECD, 2004b)

In systemic dermal toxicity studies, irritant and corrosive effects should be avoided. This is also particularly true for all types of sensitisation studies, for which the elicitation phase has to be performed with non-irritant concentrations of the test chemical (some level of irritation is usually required in the induction phase). Thus, positive data of these adverse effects can only be derived from pilot dose range finding studies, which are generally performed only on 1-2 animals per dose, and in general not well documented so that reporting may be incomplete. In addition, observations may be made on a species other than the rabbit, and species may differ in sensitivity in their responses.

The following considerations should be made when evaluating existing data from other dermal toxicity studies.

- In case the test chemical is classified as ***fatal in contact with skin***, further testing for skin corrosion / irritation is not required as it would result in severe suffering or death of the animal.
- The ***dosing design*** of the systemic studies mentioned above significantly differs from a local acute skin irritation / corrosion study. In a local *in vivo* skin irritation / corrosion test (OECD TG 404) the undiluted (neat) test chemical is applied to a very small area of 6 cm^2 (which equals about 0.25% of the body surface), while in systemic studies the test chemical is applied to a large area of the body surface (at least 10%; OECD Test Guidelines and Draize *et al.*, 1944), so that even the highest (limit) doses of 1000 mg / kg body weight (OECD TG 410 and OECD TG 412), or 2000 mg / kg body weight (OECD TG 402) are applied in dilutions, hampering the assessment of possible effects of the neat test chemical.
- On the other hand, ***the exposure duration*** in these studies is longer than the 4 hours required in OECD TG 404.
- Finally, the doses administered in systemic toxicity studies, including single maximum dose limit tests, are always administered as ***preparations in a vehicle/solvent***, in contrast to local acute skin irritation / corrosion studies, where vehicle/solvents are not commonly used.

In case of use of a diluted test chemical ECHA recommends to calculate or estimate the amount of test substance per cm^2 and compare this to the test substance concentration of 80 µl or 80 mg/cm^2 employed in the OECD TG 404 for dermal corrosion/irritation test with rabbits. If it is in the same range and adequate scoring of skin effects is provided, the test chemical shall be classified or not as Skin Irritant (EU CLP Cat. 2). In case conclusive negative data was obtained in rabbits, no further testing is needed. However, if the testing doses are not in the same range and in case of inadequate scoring of skin effects, the data should be used within the Weight of evidence analysis (ECHA 2015a).

In case the test was performed in other species than rabbits, which may be less sensitive (e.g. rat), the evaluation should be made with caution. Considering the fact that the rat skin is less sensitive compared to rabbit skin, much lower exposures are employed and, in general, the scoring of dermal effects is performed less accurately, according to ECHA the results of dermal toxicity testing in rats will not be adequate for classification with respect to skin irritation. Only in case of evidence of skin corrosivity in the rat dermal toxicity test can the test chemical be classified as Skin Corrosive (EU CLP/UN GHS Cat. 1). All other data should be used for Weight of Evidence (ECHA, 2015a).

Regarding data from skin sensitisation studies, the skin of guinea pigs is less sensitive than that of rats which is, in turn, less sensitive than that of rabbits. ECHA recommends that only in case of evidence of skin corrosivity in the sensitisation test (Maximisation or Buhler) with the neat material or

dilutions of solids in water, physiological saline or vegetable oil, the test chemical shall be classified as Skin Corrosive (EU CLP/UN GHS Cat. 1). However, care should be exercised when interpreting findings from guinea pig studies, particularly from maximisation protocols, as intradermal injection with adjuvant readily causes necrosis (ECHA, 2015a). All other data should be used for weight of evidence evaluation only. Similarly, information on irritant properties from skin sensitisation tests cannot be used to conclude on a specific classification regarding acute skin irritation but may be used in a weight of evidence analysis. In general, irritation data from the Local Lymph Node Assay are not usable since the test substance is applied to the dorsum of the ear by open topical application, and specific vehicles for enhancement of skin penetration are used (ECHA, 2015a).

Finally, information obtained from skin penetration studies using OECD TGs 427 or 428 may provide evidence on the skin corrosion potential of a test chemical. Thus, both rapidly penetrating and cytotoxic chemicals, or clearly corrosive chemicals, may be assumed to be corrosive and classified as UN GHS Cat. 1 if supported by other evidence in a weight of evidence assessment (UN, 2015a).

In conclusion, positive data of a dilution of a test chemical in the above mentioned studies, even with a species other than rabbit, may be used for a positive classification of corrosive potential. Although positive data may also be used for a positive classification of an irritant potential, experience has shown that such cases are very rare (OECD, 2014). Furthermore, negative results usually cannot negate any irritant potential observed with *in vitro* or *in vivo* skin irritation OECD TGs (OECD TG 439 or OECD TG 404) or justify a non-classification (OECD, 2014). In any case, data from other *in vivo* and/or *in vitro* dermal toxicity data may be used in a weight of evidence approach to help orient chemicals to a top-down or bottom-up approach in Part 3 of the IATA.

4.4. Physico-chemical properties (existing, measured or estimated)

Chemicals that spontaneously undergo rapid exothermic decomposition reactions with water or air (e.g., anhydrides, alkylated metal alkoxides or alkali metals), chemicals with a high oxidative activity like (hydro)peroxides, as well as chemicals with extreme pH, are likely to damage the integrity of the cells upon contact with human tissues, such as skin, and thus may be classified as skin corrosives (UN GHS/EU CLP Cat. 1).

In particular, test ***chemicals with oxidising properties*** can give rise to highly exothermic reactions in contact with other substances and human tissue. The high temperatures thus generated may damage/destroy biological materials. This applies, for example, to organic peroxides, which can be assumed to be skin irritants, unless evidence suggests otherwise. According to ECHA, for a hydro peroxide a classification as Skin Corrosive should be considered (UN GHS/EU CLP Sub-cat. 1B), whereas a Skin Irritation classification should be considered for peroxides (EU CLP Cat. 2). Appropriate evidence must be provided in order to consider non-classification of substances with oxidising properties (ECHA, 2015a, 2105b).

In addition, ***pH extremes like ≤ 2.0 and ≥ 11.5*** may indicate skin corrosion, especially when associated with significant acid/alkaline reserve (buffering capacity). In the absence of any other information, a test chemical is considered corrosive (UN GHS Cat. 1) if it has a pH ≤ 2.0 or pH ≥ 11.5. The determination of pH should be performed following OECD TG 122 (2013). According to Worth *et al.* (1998), pH is able to identify skin corrosive substances with a high specificity (94%; 31/33) but with rather low sensitivity (56%; 15/27), suggesting a high number of false positive predictions. Furthermore, despite in low number, some chemicals with extreme pH did not show corrosive effects in native skin (false positives) (Worth *et al.*, 1998).

The acid/alkaline reserve is a measure of the buffering capacity of chemicals and the higher the buffer capacity, the higher is in general the potential for corrosivity. The acid/alkaline reserve was shown to have a positive impact in particular for the classification of mixtures containing acidic or alkaline substances (Young *et al*, 1988; Young and How, 1994). The procedures to determine the acid

reserve or alkali reserve for chemicals are also described in TG 122 (2013). The relation is quantitatively expressed by: "IF pH +1/12 alkaline reserve ≥ 14.5 or pH -1/12 acid reserve ≤ -0.5, THEN the mixture should be considered as corrosive. However, the consideration of acid/alkali reserve should not be used alone to exonerate mixtures from classification. In particular for pure substances the sensitivity of pH for identifying skin corrosive was actually found to be reduced (29%; 7/24; with almost no change in specificity 92%; 11/12) when combined with acid/alkaline reserve information (Worth *et al.*, 1998).

Within the IATA context, the pH and buffering capacity should be used as an initial screen to identify skin corrosives. An extreme pH may be considered in a WoE together with other data, but it shouldn't necessarily result in a UN GHS Cat. 1 classification, since as mentioned above, there are cases of chemicals with extreme pH that are not skin corrosives. Furthermore, corrosive sub-categorisation is not possible based on pH. Finally, if consideration of acid/alkaline reserve suggests the test chemical may not be corrosive despite the low or high pH value, further testing should be carried out to confirm this, preferably by using an appropriate validated *in vitro* test (UN, 2015a; ECHA, 2015a, 2105b; OECD, 2014).

In the case of mixtures having pH ≤ 2.0 and ≥ 11.5, the ECHA Guidance document on the application of the CLP criteria recommends the use of a decision logic that take into consideration (ECHA, 2015b):

1. Consider the acid/alkaline reserve: if it indicates corrosivity classify as corrosive (e.g., UN GHS Cat. 1), otherwise go to step 2
2. Test the mixture in an OECD adopted *in vitro* test method for *skin corrosion*: if it indicates corrosion, classify accordingly (UN GHS / EU CLP Sub-cat. 1A, 1B, 1C), otherwise go to step 3
3. Make use of the generic concentration limits triggering classification (see chapter 4.5.2 on additivity approach), or test in a validated *in vitro* test method for *skin irritation* to classify the mixture accordingly instead of using the additivity approach.

The ECHA Guidance document on CLP criteria further states that "*where the mixture has an extreme pH value but the only corrosive/irritant ingredient present in the mixture is an acid or base with an assigned Specific Concentration Limit (either in CLP Annex VI or set by supplier), then the mixture should be classified according to the Specific Concentration Limit. In this instance, pH of the mixture should not be considered a second time since it would have already been taken into account when deriving the Specific Concentration Limit for the substance*." Furthermore it states that "*It is also important to note that the pH-acid/alkali reserve to change classification from corrosive to irritant, or from irritant to not classified, assumes that the potential corrosivity or irritancy is due to the effect of the ionic entities. When this is not the case, especially when the mixture contains non-ionic (non-ionisable) substances themselves classified as corrosive or irritant, then the pH-reserve method cannot be a basis for modifying the classification but should be considered in a weight of evidence analysis. If a mixture with corrosive constituents also contains surfactants (e.g. tensids or detergent substances), it can be assumed that corrosivity might be amplified (Kartono & Maibach 2006). Even if only one corrosive substance with an assigned Specific Concentration Limit is present in such a mixture, the possible synergistic effect has to be taken into account when classifying the mixture"* (ECHA, 2015b).

Other physico-chemical properties than those above mentioned such as melting point, molecular weight, octanol-water partition coefficient, surface tension, vapour pressure, aqueous solubility and lipid solubility, may also be used to identify chemicals with skin irritation or corrosion potential (Walker *et al.*, 2005) or chemicals not likely to cause such adverse effects (Gerner *et al.*, 2004). Such physico-chemical parameters may be measured or estimated using non-testing methods (see chapter below), e.g., (Q)SARs, and may be used to help orient chemicals to a top-down or bottom-up approach in Part 3 of the IATA.

4.5. Non-testing methods for skin corrosion and irritation

Non-testing methods exist for both substances and mixtures. For substances, non-testing methods can be divided into the following categories:

- Analogue approaches: read-across, SAR, and grouping of substances (category formation)
- (Quantitative) Structure-Activity Relationship (i.e., (Q)SARs), which quantitatively correlate activity to structure or structure-derived descriptors, and expert systems that often incorporate several SARs, (Q)SARs, expert rules and/or data.

For mixtures, non-testing methods can be divided into (UN, 2015a; ECHA, 2015b):

- Bridging principle, when data are not available for the complete mixture but are available for similar tested mixtures and for the individual ingredients, and
- Theory of additivity, when data are not available for the complete mixture, but are available for all or some of the ingredients of the mixture.

4.5.1. (Q)SARs, read-across, grouping and prediction systems for substances

Non-testing methods for substances can be used if their proposed scientific validity has been documented according to internationally agreed procedures and if they provide adequate, relevant and reliable data for skin corrosion and irritation, for the substance of interest. Justifications for (Q)SARs and Expert Systems are provided by means of a (Q)SAR Model Reporting Format (QMRF) proposing validity of the method including consideration of the OECD (Q)SAR principles: (i) defining of the endpoint, (ii) defining the algorithm, (iii) defining the AD, (iv) defining goodness of fit and robustness, (v) defining predictivity and (vi) providing a mechanistic understanding. The Joint Research Centre (JRC) QSAR Model Database is an inventory of information on available QMRFs, freely accessible online[4]. In addition, the adequacy and reliability of individual predictions is demonstrated by means of a (Q)SAR Prediction Reporting Format (QPRF).

With the introduction of the OECD (Q)SAR Toolbox[5] in combination with the eChemPortal5[6], useful tools are provided for:

- Finding existing data on the substance under question (target),
- Identifying analogues for potential read-across and grouping and finding existing data on these analogues,
- Applying a number of SARs and other profilers for skin irritation and corrosion to the target structure,
- Grouping and deriving simple (Q)SAR or trend relationships.

Guidance on how to apply (Q)SARs for regulatory use and on how to assess the validity and suitability of (Q)SAR models and adequacy of their predictions is available from the corresponding section of the OECD website[7] and is also provided in the OECD GD 69 (OECD, 2007a).

4.5.1.1. Applicability, limitations and role within the IATA

In general, there are several different ways in which non-testing methods can be used in the context of an Integrated Approach to Testing and Assessment (IATA) (OECD, 2014), e.g.:

- for direct prediction of corrosion/irritation potential or the absence thereof,
- as part of a Weight of evidence scheme (where e.g. the information from non-testing methods alone is not sufficient for a decision), or
- in order to decide how best to proceed with further (*in vitro*) testing (i.e. via a top-down or bottom-up approach).

[4] https://eurl-ecvam.jrc.ec.europa.eu/databases/jrc-qsar-model-database-and-qsar-model-reporting-formats accessed on 18 Nov. 2016.

[5] http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm accessed on 18 Nov. 2016.

[6] http://www.echemportal.org accessed on 18 Nov. 2016.

[7] http://www.oecd.org/chemicalsafety/risk-assessment/guidancedocumentsandreportsrelatedtoqsars.htm accessed on 18 Nov. 2016.

In the case of skin corrosion and irritation, many of the non-testing models have a mechanistic basis, which provides additional information on the relevance of the model. The predictive capacity is however, model-, domain- and context-specific. Furthermore, their applicability is limited to the applicability domain of the model used. Their use is generally straight-forward for monoconstituent substances, whereas for multi-constituent substances, this only holds if the composition of the substance is known (i.e. percentage of each of the discrete organic constituents) because then predictions can be performed on each constituent and the effect of the multi-constituent substance predicted by employing a dose addition approach. QSAR models and grouping approaches have, however, been employed on multi-constituent substances and on substances of unknown or variable composition, complex reaction products or biological materials (UVCB), with partly unknown composition details for other endpoints than skin irritation/corrosivity by accepting some uncertainty and assuming that all constituents of the considered UVCBs are represented by a few known constituents/groups of constituents, on which QSAR models or grouping approaches then could be employed.

4.5.1.2. Analogue approaches (substances)

Read-across, SARs and grouping / category formation are treated together because they represent approaches based on the same basic concept and because the existence of a SAR (structural alert or set of fragments) provides one means of justifying read-across. Note that, depending on the legal framework and member country, specific requirement may be associated to the read-across and grouping approaches. For example, under the EU REACH Regulation, read-across needs to be justified, documented, and supported by reliable data on the sources substance. Furthermore, the structural similarity between the source and target substance needs to be shown. The similarity of two substances can be based for example on a common functional group, common pre-cursors or common break-down products. Grouping also requires that toxicological properties of the target substance may be predicted from the data of the source substance, basically by interpolation and/or in some cases extrapolation (OECD, 2007b).

The data from structural analogues that exhibit corrosion (or irritation) potential can be used to predict the effect of the substance of interest and derogate from further assessment. Negative data from structural analogues may also be used to make predictions in certain cases. However, the absence of one of the known structural alerts for irritation and corrosion alone does not prove absence of effect, as knowledge of structural alerts for irritation and corrosion might be incomplete. For instance, other substructures (not yet identified as structural alerts) or other properties of the substance may be responsible for a corrosive or irritant effect. As an example, irritant contact dermatitis may occur indirectly, such as in the case of exposure to organic solvents with defatting properties.

A variety of SARs for predicting the presence of irritation or corrosion have been described by Hulzebos *et al.* (2001, 2003, 2005), and some of these have been incorporated into the BfR (German Federal Institute for Risk Assessment) rulebase and the SICRET tool (Walker *et al.*, 2005). The BfR alerts ("inclusion rules") for corrosion and irritation have been incorporated into the Toxtree software[8] and into the OECD QSAR Toolbox[9].

4.5.1.3. (Q)SARs and expert systems on skin irritation and corrosion (substances)

An overview of available (Q)SARs for skin corrosion and irritation is provided in table 4.3. QSARs and expert systems for skin corrosion and irritation have been described in several reviews (Hulzebos *et al.*, 2001, 2003, 2005; Patlewicz *et al.*, 2003; Gallegos Saliner *et al.*, 2006, 2008). A comparison of the predictive capacities of three popular commercial tools is also available (Mombelli, 2008). Detailed information and a few examples on the available (Q)SAR models and expert systems is given in the

[8] https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxtree accessed on 18 Nov. 2016.

[9] https://www.qsartoolbox.org accessed on 18 Nov. 2016.

appendix R.7.2-2 of the ECHA guidance on information requirements and chemical safety assessment – chapter R.7a: endpoint specific guidance, version 4.1 (ECHA, 2015a).

Table 4.3 Overview of available (Q)SARs for skin corrosion/irritation (extract from ECHA, 2015a)

Reference or name of the model	Applicability domain
Literature models	
Barratt (*et al.*) (1995a, 1996a,b,c); Whittle *et al.* (1996)	Diverse local models for acids, bases , phenols, neutral organic and electrophiles
Hayashi *et al.* (1999)	Phenols
Kodithala *et al.* (2002)	Phenols, esters, and alcohols
Nangia *et al*. (1996)	Bases
Smith *et al*. (2000a, 2000b)	Esters
Gerner *et al*. (2004); Hulzebos *et al.* (2005); Walker *et al*. (2004)	New Chemicals Database, organic chemicals with no significant hydrolysis potential and purity > 95%
Golla *et al*. (2009)	Organic chemicals from diverse classes
Data repositories	
Danish QSAR database[10] (also included in the OECD QSAR Toolbox)	Industrial chemicals, pesticides, etc.
Computerised models	
PaDEL-DDPredictor (Liew and Yap, 2013)	Calculated by the model based on the range of descriptors
BfR rule-base, free (included in the OECD QSAR Toolbox and Toxmatch, Toxtree, ToxPredict and Ambit)	EU New chemicals (NONS) database, organic chemicals with no significant hydrolysis potential and purity > 95%
ACD/Percepta	Organic chemicals
DEREK Nexus, commercial	Organic chemicals and some metals
HazardExpert, commercial	Organic chemicals
MolCode, commercial	Organic chemicals
MultiCASE, commercial	Organic chemicals
TOPKAT, commercial	Organic chemicals
Review papers	
Hulzebos *et al*. (2001, 2003, 2005)	Not applicable
Patlewicz *et al*. (2003)	Not applicable
Gallegos Saliner *et al*. (2006, 2008)	Not applicable
Mombelli (2008)	Not applicable

Most of the (Q)SARs reported in the literature have been developed from small data sets of specific groups of substances, although in some cases more diverse and larger datasets were also examined. In general, it has been suggested that basic physico-chemical parameters such as acidity, basicity, hydrophobicity, and molecular size as well as electrophilic reactivity are useful to predict the toxic potential of homologous substances. Also the ability for skin penetration likely constitutes a relevant factor. In contrast, models intended to predict the toxic potential of heterogeneous groups of substances emphasise the commonality of structural features.

Expert systems are computer programs that guide hazard assessment by predicting toxicity endpoints of certain substance structures based on the available information. They can be based on an

[10] http://qsar.food.dtu.dk accessed on 18 Nov. 2016.

automated rule-induction system (e.g., TOPKAT, HazardExpert and MultiCASE), or on a knowledge-based system (e.g., DEREK or the BfR-DSS).

Not all of the models were developed with regulatory purposes in mind (e.g., EU CLP or UN GHS classification systems), so it is important to assess in each case whether the endpoint or effect being predicted corresponds to the regulatory endpoint of interest. The rule-base at the heart of the former BfR DSS has been developed to predict EU regulatory endpoints, however predictions refer to the former EU Dangerous Substance Directive (EU DSD) classification/labelling system used before the EU CLP regulation came into force. Due to the change of EU criteria for classification of skin irritants Cat. 2 (see chapter 1.4.3), predictions as skin Cat. 2 from models developed based on previous EU criteria (EU DSD) should be interpreted with caution since they may lead to over-prediction and should not be used for direct classification under EU CLP. These models can however be argued to be "conservative" and therefore acceptable for predicting no classification under EU CLP.

Based on the BfR rule-base, the freely downloadable OECD QSAR Toolbox software contains two profilers relevant for corrosion/irritation, which encode both the “inclusion rules” (structural alerts predicting corrosion/irritation potential) and the “exclusion rules” (“IF…THEN NOT…” rules predicting the absence of irritation/corrosion potential) due to certain physicochemical properties. The use in combination with other profilers (e.g. for skin metabolism) and data for analogues allows for the prediction of skin corrosion/irritation for new chemicals through read-across or category approaches.

In the case of classification models for skin corrosion, where it is not indicated whether the predicted classification should be Skin Corrosive UN GHS/EU CLP Sub-cat. 1A, 1B or 1C, a Cat. 1 prediction without further sub-categorisation should be used. Very few models are available (see Gallegos Saliner *et al.*, 2006, 2008 for review). Available models tend to focus on defined chemical classes (e.g., acids, bases, phenols) and may be useful as an alternative to *in vitro* testing for such classes. For classification and labelling, the BfR rule-base provides information that is the closest to the regulatory goal, since the system was designed to predict former EU Risk Phrases for skin irritation (R38) and corrosion (R34, R35) under the EU Dangerous Substance Directive (EU DSD). However, in borderline cases and as highlighted above, the prediction may not fully reflect the correct classification under EU CLP.

4.5.2. Bridging principles and theory of additivity for mixtures

4.5.2.1. Applicability, limitations and use within the IATA

According to the OECD recommended IATA (OECD, 2014), non-testing methods are usually used as supporting information in a WoE approach, e.g., to support observations from available data from *in vivo or in vitro* dermal toxicity tests (see chapter 4.3) and to support skin corrosion or irritation *in vitro* results (see chapter 3). If further testing is required, information generated with non-testing methods may be used for deciding on how to address Part 3 i.e., initiate a top-down or a bottom-up approach.

In the European Union, a decision logic for classification of mixtures is recommended in which the use of bridging principles and theory of additivity shall be used in steps 2 and 3 below, when there is no data available on the complete mixture. The decision logic procedure recommends to consider the following steps (ECHA, 2015b):

1. Data on mixture itself:
   - Physico-chemical properties such as pH and buffering capacity,
   - Existing human and animal data,
   - Data from other dermal exposure studies,
   - Data from OECD adopted and suitable *in vitro* test methods;
2. When data on the entire mixture is not available, then:
   - Make use of the bridging principles in case there are sufficient existing skin corrosion/irritation data available on similar tested mixtures and on the individual ingredients;
3. If the bridging principle cannot be applied, then:

- For extreme pH mixtures: make use of pH, buffering capacity & use of OECD adopted *in vitro* test methods as described in chapter 4.4 for mixtures;
- For non-extreme pH mixtures: make use of the additivity approach if applicable (as described in table 4.4 below but taking into account both the generic concentration limit (GCL) and the SCL), otherwise make use of the approach as described in table 4.5 below.

In any case, very limited data is available on the predictive capacity on the use of non-testing methods on mixtures for determining skin irritation/corrosion potential hazard. Only an impact assessment carried out by the International Association for Soaps, Detergents and Maintenance Products (A.I.S.E.) showed that the use of the UN GHS theory of additivity for classification of detergent and cleaning products can result in the over-labelling of products currently not requiring classification according to consistent animal, *in vitro* and human experience data (OECD, 2014).

4.5.2.2. Bridging principles (mixtures)

Bridging principles are used when the mixture itself has not been tested to determine skin corrosion/irritation potential but there are sufficient data on both the individual ingredients and similar tested mixtures to adequately characterise the hazards of the mixtures. The following bridging principles may be used: based on dilution, batching, concentration of the highest corrosion/irritation category, interpolation within one hazard, substantially similar mixtures, and aerosols (see chapter 3.2.3.2 of UN, 2015a; and chapter 1.6.3 of ECHA, 2015b). When the available identified information is inappropriate for the application of the bridging principles, then the mixture should be classified based on its ingredients and according to the additivity approach as described in chapter 4.5.2.3 below.

Dilution: Where the tested mixture is diluted with a substance (diluent) that has an equivalent or lower hazard category than the least hazardous original ingredient substance and which is not expected to affect the skin corrosivity/irritancy of other ingredients, then it can be assumed that the hazard of the new mixture is equivalent to that of the original tested mixture, and the new diluted mixture may be classified as equivalent to the original tested mixture.

Batching: Where a batch of a mixture is produced under a controlled process and by or under the control of the same manufacturer, then it can be assumed that the skin corrosin/irritation potential of the tested production batch is equivalent to those of untested batches of the same commerical product. This method must not be used where there is reason to believe that the composition may vary significantly, affecting the skin corrosin/irritation potential of the untested batches. If this occurs, a new classification is necessary.

Concentration of the highest corrosion/irritation category: Where a tested mixture is classified in the highest Sub-category for skin corrosion, an untested mixture which contains a higher concentration of those ingredient substances that are in that Sub-category should also be classified in the highest corrosion Sub-category without further testing (UN, 2015a; ECHA, 2015b). If a tested mixture classified for skin irritation (UN GHS Cat. 2) is concentrated and does not contain skin corrosive ingredients, a more concentrated untested mixture should be classified for skin irritation (UN GHS Cat. 2) without additional testing (UN, 2015a).

Interpolation within one hazard: In the case there are three mixtures (A, B and C) which contain identical hazardous ingredients, if mixtures A and B have been tested and are in the same skin corrosion/irritation hazard category, and mixture C is not tested and has concentrations of those hazardous components intermediate to the concentrations in mixtures A and B, then mixture C is assumed to be in the same skin corrosion/irritation category as A and B.

Substantially similar mixtures: Two mixtures contain an identical ingredient at the same concentration. Each of the two mixtures contains an additional ingredient which is not identical with each other; however they are present in equivalent concentrations and the hazard category of these two ingredients is the same and neither of them is expected to affect the hazard classification of the other. If one of the mixtures is classified based on test data it may be assumed that the hazard category of the other mixture is the same.

Aerosols: A mixture in aerosol form is considered to have the same classification as the non-aerosolised form of the mixture, provided that the propellant used does not affect skin corrosion/irritation properties of the mixture upon spraying and data demonstrating that the aerosolised form is not more hazardous than the non-aerosolised form is available (UN, 2015a; ECHA, 2015b).

4.5.2.3. Additivity approach (mixtures)

The theory of additivity is used when data are available for all or only some of the ingredients, but not on the mixture as a whole. It assumes that each skin corrosive or irritant ingredient contributes to the overall corrosive or irritant properties of the mixture in proportion to its potency and concentration. The mixture is classified as corrosive or irritant to skin when the sum of the concentrations of the relevant ingredients exceeds a cut-off value / concentration limit (see chapter 3.2.3.3 of UN, 2015a; and chapter 3.2.3.2.3 of ECHA, 2015b).

Both the UN GHS and the EU CLP consider that the 'relevant ingredients' of a mixture are those which are present in concentrations of 1% (w/w for solids, liquids, dusts, mists and vapours and v/v for gases) or greater, unless there is a presumption (e.g., in the case of corrosive ingredients) that an ingredient present at a concentration of less than 1% can still be relevant for classifying the mixture for skin irritation/corrosion (UN, 2015a; ECHA 2015b).

Indeed before using the additivity approach, the supplier must ascertain that it is applicable. The first step for doing so is to identify all the ingredients in the mixture (i.e. their name, chemical type, concentration level, hazard classification and any specific concentration limits) and the pH of the mixture. In addition to that, for example surfactant interaction, neutralisation of acids/bases could also occur in a mixture, which also makes it important to consider effects of the entire mixture (i.e. pH and the acid/alkaline reserve) rather than considering contributions of individual ingredients (ECHA, 2015b).

In cases where the additivity approach applies for skin corrosion/irritation to a mixture with two or more substances some of which may have SCLs assigned, the following formula should be used (ECHA, 2015b):

The mixture is classified for skin corrosion/irritation if the:

$$\text{Sum of } (\text{ConcA} / \text{clA}) + (\text{ConcB} / \text{clB}) + \ldots + (\text{ConcZ} / \text{clZ}) \text{ is} \geq 1$$

Where ConcA = the concentration of substance A in the mixture;
clA = the concentration limit (either specific or generic) for substance A;
ConcB = the concentration of substance B in the mixture;
clB = the concentration limit (either specific or generic) for substance B; etc.

Table 4.4 provides the generic concentration limits to be used to determine if the mixture is considered to be corrosive or irritant to the skin according to the EU CLP classification scheme. However, the specific concentration limits (SCLs) take precedence over the generic concentration limits GCLs. Thus, if a given substance has a SCL (e.g. in CLP Annex VI or set by supplier), then this limit has to be taken into account when applying the additivity approach for skin corrosion/irritation (ECHA, 2015b). Furthermore, a weighting factor of 10 is used for corrosive components when they are present at a concentration below the generic concentration limit for the UN GHS/EU CLP Cat. 1 classification, but are at a concentration that will contribute to the classification of the mixture as an irritant (UN, 2015a; ECHA, 2015b).

Table 4.4. When the additivity approach is applicable: generic concentration limits of ingredients classified for skin corrosion/irritant hazard (EU CLP Cat. 1 or 2) that trigger classification of the mixture as corrosive/irritant to skin (extract from ECHA, 2015b)

Sum of ingredients classified as:	**Concentration triggering classification of a mixture as:**	
	Skin Corrosive EU CLP Cat. 1*	**Skin Irritant EU CLP Cat. 2**
Skin corrosive EU CLP Sub-cat. 1A, 1B, 1C	≥ 5%	≥ 1% but < 5%
Skin irritant EU CLP Cat. 2		≥ 10%
(10 x Skin corrosive EU CLP Sub-cat. 1A, 1B, 1C) + Skin irritant EU CLP Cat. 2		≥ 10%

* The sum of all ingredients of a mixture classified as Skin Corrosive Sub-cat. 1A, 1B or 1C respectively, shall each be ≥ 5% respectively in order to classify the mixture as either Skin Corrosive Sub-cat. 1A, 1B or 1C. If the sum of the Skin Corrosive Sub-cat. 1A ingredients is < 5% but the sum of Sub-cat. 1A+1B ingredients is ≥ 5%, the mixture shall be classified as Skin corrosive Sub-cat. 1B. Similarly, if the sum of Skin corrosive Sub-cat. 1A+1B ingredients is < 5% but the sum of Sub-cat. 1A+1B+1C ingredients is ≥ 5% the mixture shall be classified as Skin Corrosive Sub-cat. 1C. Where at least one relevant ingredient in a mixture is classified as Cat. 1 without sub-categorisation, the mixture shall be classified as Cat. 1 without sub-categorisation if the sum of all ingredients corrosive to skin is ≥ 5%.

4.6. Weight of evidence evaluation

Current regulatory frameworks require that consideration is given to the totality of the available information by making a weight of evidence (WoE) assessment (UN, 2015a; OECD, 2014; ECHA, 2015a, 2015b). This is especially true when there is conflict in information available on some parameters (UN, 2015a). The WoE should be carried out before any additional *in vitro* or *in vivo* testing is performed (OECD, 2014). In case of skin corrosion/irritation information to be considered include existing *in vivo*, *in vitro* and/or human data, physico-chemical properties (e.g., pH, acid/alkaline reserve), and non testing methods such as (Q)SAR, read-across and grouping information (OECD, 2014).

The quality and consistency of the data should be taken into account when weighing each piece of available information (OECD, 2014). Both positive and negative results can be assembled together in a single weight of evidence determination. Evaluation must be performed on a case-by-case basis and with expert judgement, which shall be exercised prior to making such a determination (OECD, 2014; UN, 2015a). Normally positive results that are adequate for classification should not be overruled by negative findings (UN, 2015a; ECHA, 2015b). The WoE assessment needs to be transparently explained and documented to enable a logical flow leading to the decision/conclusion (OECD, 2014).

As described above, the WoE approach involves an ***assessment of the relative values/weights*** of different pieces of the available information that has been retrieved and gathered in previous steps (for an example cf. Hulzebos and Gerner, 2010). These weights/values can be assigned either in a more objective way by applying a formalised procedure (e.g., based on Bayesian logic, as in Rorije *et al.*, 2013) or by using expert judgement. The weight given to the available evidence will be influenced by factors such as the quality of the data, consistency of results/data, nature and severity of effects, relevance of the information for the given regulatory endpoint. In all cases the relevance, reliability and adequacy for the purpose have to be considered.

Since it quality will contribute to the value/weight of each data element, the ***quality of the data*** that is obtained for a WoE needs to be assessed (OECD, 2014). In case the quality of a certain study is deemed to be inappropriate, it is recommendable not to consider those data in the WoE, but focus on other pieces of information which are of sufficient quality (OECD, 2014). Quality might be inappropriate e.g., due to missing validation of the methodology, “non-adherence” to the relevant test guideline/method, lack of adequate controls, deficiencies in data reporting etc. Examples of tools to evaluate the quality include the Klimisch scores (Klimisch *et al.*, 1997) and Hill’s criteria for evaluation

of epidemiological data (Hill, 1965), as well as the JRC's ToxRTool for scoring *in vivo* and *in vitro* data (Schneider *et al.*, 2009).

The ***evaluation of adequacy*** of test results and documentation for the intended purpose is particularly important for chemicals where there may be (a number of) test results available, but where some or all of them have not been carried out according to current standards (OECD, 2014). Adequacy defines the usefulness of information for the purpose of hazard and risk assessment, in other words whether the available information allows clear decision-making about whether the chemical is non-irritant, irritant or corrosive and an adequate classification can be derived (OECD, 2014). Where there is more than one study, the greatest weight is attached to the studies that are the most relevant and reliable. For each endpoint, robust summaries need to be prepared for the key studies. Sound scientific judgment is an important principle in considering the adequacy of information and determining the key study (OECD, 2014).

The ***consistency of the available data*** from various sources is crucial and should therefore be thoroughly evaluated in WoE (OECD, 2014). In case the data elements are of comparable weight but give inconsistent evidence (e.g., (Q)SAR is positive and available limited human data is negative), usually WoE analysis will not be conclusive and prospective *in vitro* and/or *in vivo* testing will have to be conducted (Part 3 of the IATA). In case the weights of the individual pieces of evidence differ considerably (e.g., where irritation is observed in a local lymph node assay (LLNA) as a piece of evidence with lower weight and existing human data of good quality indicate lack of irritancy as evidence with higher weight), a WoE conclusion may be drawn according to the evidence carrying the highest weight. Consistent data, on the other hand, which come from several studies/sources may be considered sufficient for regulatory purposes. If high quality HPT, *in vitro* (Modules 3 and 4) and/or *in vivo* (Module 2) data are available, these should carry the highest weight in the WoE assessment (OECD, 2014).

Another important element of WoE is to consider ***to what extent the parameters and observations were addressed*** by each data element of the WoE (OECD, 2014). While in a standard *in vivo* test guideline the required parameters / observations have been specified and often build the basis for decision making (e.g., C&L for skin irritation is mainly directly derived from Draize scores), it is not always possible to extract information equivalent to those parameters from non-testing data.

Taking into account the above elements, in the final WoE assessment, each data element will be characterised for its quality, relevance, coverage (e.g., irritation and/or corrosion) and associated uncertainty. The assessor would either decide to include or exclude the existing information based on these (OECD, 2014). Two conclusions may be possible on the WoE evaluation:

i) When consistency is seen among "qualified" data elements, WoE may reach a conclusion that the relevant endpoint or information requirement has been sufficiently covered and further testing is not necessary.

ii) When on the other hand, insufficient information remains after the "non-qualified" data have been rejected/put aside and/or when the remaining information is inconsistent or contradictory, WoE would reach to a conclusion that the relevant endpoint or information requirement has not been sufficiently covered and further testing is necessary, depending on the specific legal/regulatory framework, and inform on which test to conduct to fill the data gap. In these cases, all available information and the WoE assessment should be used to formulate a hypothesis of the most likely skin irritation/corrosion potential of the chemical. This hypothesis and the regulatory context under which a decision must be taken should then guide the choice of test methods to be used and the sequence of the prospective testing in either a top-down or a bottom-up approach (see chapter 3.1).

References

Alépée N., Tornier C. Rober C., Amsellem C., Roux M.H., Doucet O., Pachot J., Méloni M., de Brugerolle de Fraissinette A. (2010). A catch-up validation study on reconstructed human epidermis (SkinEthic RHE) for full replacement of the Draize skin irritation test. *Toxicology In Vitro* 24, 257-266.

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

Alépée N., Rober C., Tornier C. and Cotovio J. (2014b). The usefulness of the validated SkinEthic™ RHE test method to identify skin corrosive UN GHS subcategories. *Toxicology In Vitro* 28, 616-625.

Alépée N., Grandidier M.H., Tornier C., Cotovio J. (2015a). An integrated testing strategy for *in vitro* skin corrosion and irritation assessment using SkinEthic™ Reconstructed Human Epidermis. *Toxicology In Vitro* **29**, 179-1792.

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. (2015b). Use of HPLC/UPLC-spectrophotometry for detection of formazan in *in vitro* Reconstructed Human Tissue (RhT)-based test methods employing the MTT-reduction assay to expand their applicability to strongly coloured test chemicals. *Toxicology In Vitro* **29**, 741-761.

Alépée N., Hibatallah J., Klaric M., Mewes K.R., Pfannenbecker U., and McNamee P. (2016). Assessment of cosmetic ingredients in the *in vitro* reconstructed human epidermis test method EpiSkinTM using HPLC/UPLC-spectrophotometry in the MTT-reduction assay. *Toxicology In Vitro* **33**, 105-117.

Barratt M.D. (1995). Quantitative structure activity relationships for skin corrosivity of organic acids, bases and phenols. Toxicology Letters 75, 169-176.

Barratt M.D. (1996a). Quantitative structure-activity relationships (QSARs) for skin corrosivity of organic acids, bases and phenols: Principal components and neural network analysis of extended datasets . Toxicology *in vitro* 10, 85-94.

Barratt M.D. (1996b). Quantitative structure-activity relationships for skin irritation and corrosivity of neutral and electrophilic organic chemicals. Toxicology *in vitro* 10, 247-256.

Barratt M.D., Dixit M.B., Jones P.A. (1996c). The use of *in vitro* cytotoxicity measurements in QSAR methods for the prediction of the skin corrosivity potential of acids. Toxicology *in vitro* 10, 283-290.

Barratt M. D., Branton P.G., Fentem J.H., Gerner I., Walker A.P., Worth A.P. (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.

Basketter D.A., Whittle E., Griffiths H.A., York M. (1994). The identification and classification of skin irritation hazard by human patch test. Food Chem. Toxico. 32, 769–775.

Basketter D.A., Chamberlain M., Griffiths H.A., Rowson M., Whittle E., York M. (1997). The classification of skin irritants by human patch test. Food Chem. Toxico. 35, 845-852.

Basketter D.A., York M., McFadden J.P., Robinson M.K. (2004). Determination of skin irritation potential in the human 4-h patch test. Contact Dermatitis 51, 1 – 4.

Basketter D., Jírová D, Kandárová H (2012). Review of skin irritation/corrosion hazards on the basis of human data: a regulatory perspective. *Interdiscip. Toxicol*. 5, 98-104.

Borlon C, Godard P, Eskes C, Hartung T, Zuang V, Toussaint O (2007). The usefulness of toxicogenomics for predicting acute skin irritation on *in vitro* reconstructed human epidermis. *Toxicology* **30**, 157-166

Botham PA, Chamberlain M, Barratt MD, Curren RD, Esdaile DJ, Gardiner JR, Gordon VC, Hildebrand B, Lewis RW, Liebsch M, Logemann P, Osborne R, Ponec M, Régnier J-F, Steiling W, Walker AP & Balls M (1995) A prevalidation study on *in vitro* skin corrosivity testing. The report and recommendations of ECVAM workshop 6. *ATLA* **23**, 219-255.

Botham PA, Earl LK, Fentem JH, Roquet R, Van de Sandt JJM (1998). Alternative methods for skin irritation testing: the current status. ECVAM Skin Irritation Task Force Report 1. *ATLA* **26**, 195-211.

Bytheway H., Fuchs D., Fuchs H., Guest R., Hoffmann J., Warren N., Whittingham A. (2009). An Initial Evaluation of the CellSystems EST-1000 Reconstructed Human Skin Model for distinguishing R34 and R35 corrosives *in vitro*. Poster SOT 2009, USA.

Casas JW, Lewerenz GM, Rankin EA, Willoughby JA Sr, Blakeman LC, McKim JM Jr, Coleman KP (2013). *In vitro* human skin irritation test for evaluation of medical device extracts. Toxicol *In vitro* 27, 2175-2183.

Corrositex (2008). DB-ALM Protocol No. 116. SOP on the Corrositex® continuous time monitor assay. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/skin-corrosion-docs/116.pdf. Accessed on 11.05.2016.

Cotovio J., Grandidier M.-H., Portes P., Roguet R., Rubinstenn G. (2005). The *in vitro* acute skin irritation of chemicals : optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. *ATLA* **33**, 329-349.

de Brougerolle de Fraissinette A, Tornier C, Cotovio J, Alépée N (2009). *In vitro* predictivity of the validated SkinEthic™ RHE “42 bis” Skin Irritation Test method using the UN GHS (in EU: CLP) classification. Poster presented at the 7th World Congress, 31 August – 3 September, Rome, Italy.

Deshmukh GR, Kumar KH, Reddy PV, Rao BS (2012). *In vitro* skin corrosion: Human skin model test - A validation study. Toxicol *In vitro* 26,1072-1074.

Desprez B., Barroso J., Griesinger C., Kandárová H., Alépée N., Fuchs H.W. (2015). Two novel prediction models improve predictions of skin corrosive sub-categories by test methods of OECD Test Guideline No. 431. *Toxicology in vitro* **29**, 2055-2080.

Draize, J.H., Woodard, G. & Calvery, H.O. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *Journal of Pharmacology and Experimental Therapeutics* **82**, 377-390.

EC (1999). Directive 1999/45/EC of the European Parliament and of the Council of 31 May 1999 concerning the approximation of laws, regulations and administrative provisions of the Member States relating to the classification, packaging and labelling of dangerous preparations. *Official Journal of the European Communities* **L 200**, 1-68.

EC (2000). Annex I to Commission Directive 2000/33/EC adapting to technical progress for the 27th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Communities* **L136**, 91-97.

EC (2001). Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Communities* **L 225**, 1-333.

EC (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union **L 396**, 1-849.

EC (2008a). Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of substances and mixtures , amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. *Official Journal of the European Union* **L353**, 1-1355.

EC (2008b). Council Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal of the European Union* **L142**, 1-739.

EC (2009a). Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. *Official Journal of the European Union*, **L342**, 59-209.

EC (2009b). Annex III. Method B.46. *In vitro* Skin Irritation: Reconstructed Human Epidermis Model Test. pp. 24-35. In: Commission Regulation (EC) No 761/2009 of 23 July 2009 amending for the purpose of its adaptation to technical progress, Regulation No. 440/2008 laying down test methods pursuant to Regulation (EC) No. 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal of the European Union* **L220**, 1-94.

EC (2010). Directive 2010/63/EU of the European Parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union*, **L276**, 33.

EC (2016a). Regulation (EU) 2016/918 of 19 May 2016 amending, for the purposes of its adaptation to technical and scientific progress, Regulation (EC) No 1272/2008 of the European Parliament and of the Council on classification, labelling and packaging of substances and mixtures. *Official Journal of the European Union* L 156, 1-103.

EC (2016b). Regulation (EU)2016/863 of 31 May 2016 amending Annexes VII and VIII to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards skin corrosion/irritation, serious eye damage/eye irritation and acute toxicity. *Official Journal of the European Union* L 144, 27-31.

ECETOC (2002). Monograph 032 - Use of Human Data in Hazard Classification for Irritation and Sensitisation. Available at: [http://www.ecetoc.org/monographs].

ECETOC (2009). TR 104: Framework for the Integration of Human and Animal Data in Chemical Risk Assessment. Available at: [http://www.ecetoc.org/technical-reports].

ECHA (2011): Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.4: Evaluation of available information. Version 1.1. Available at: http://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf. Accessed on 25.04.2016.

ECHA (2015a). Guidance on information requirements and chemical safety assessment. Chapter R.7a: Endpoint Specific Guidance. Version 4.1. Subchapter R.7.2. Skin corrosion/irritation, serious eye damage/eye irritation and respiratory tract corrosion/ irritation, p. 169-200. Available at: http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf. Accessed on 20.12.2016.

ECHA (2015b). Guidance on the Application of the CLP criteria. Guidance to regulation EC N. 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Version 4.1. Chapter 3.2. Skin corrosion/irritation. Available at: https://echa.europa.eu/documents/10162/13562/clp_en.pdf. Accessed on 26.04.2016.

El Ghalbzouri A, Siamari R, Willemze R, Ponec M. (2008). Leiden reconstructed human epidermal model as a tool for the evaluation of the skin corrosion and irritation potential according to the ECVAM guidelines. *Toxicology In vitro* **22**, 1311-1320.

Engelking O., Fuchs H.W., Weisensee D. (*in press*). EpiCS Skin Irritation Test (SIT) method. In: Alternatives for dermal toxicity testing. Eds. Eskes C., van Vliet E., Maibach H. Springer, *in press*.

epiCS® 4.1 (2012). SOP version *In vitro* Skin Corrosion: Human Skin Model Test Epidermal Skin Test epiCS® CellSystems. Available at: http://cellsystems.de/files/epiCS/SOP-epiCS-skin-corrosion.pdf. Accessed on 10.05.2016.

EpiDerm™ EPI-200-SIT (2009). SOP version MK-24-007-0023. Protocol for: *In vitro* EpiDerm™ Skin Irritation Test (EPI-200-SIT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm (EPI-200). Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-skin-irritation-1/DOC13-SOP%20modified%20EpiDerm%20MTT.pdf Accessed on 11.05.2016.

EpiDerm™ SCT (2012). SOP version MK-24-007-0024. Protocol for: *In vitro* EpiDerm™ Skin Corrosion Test (EPI-200-SCT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/skin-corrosion-docs/EpiDerm%20EPI-200-SCT%20%20Protocol%20MK-24-007-0024.pdf. Accessed on 10.05.2016.

EpiSkin™ SIT (2009). SOP Version 1.8. ECVAM Skin Irritation Validation Study: Validation of the EpiSkin™ Test Method 15 min - 42 hours for the Prediction of acute Skin Irritation of Chemicals. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-skin-irritation-1/DOC11-SOP%20EpiSkin_revised_MTT.pdf. Accessed on 11.05.2016.

EpiSkin™ SM (2011). INVITTOX Protocol No. 118. EpiSkin™ Skin Corrosivity Test. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/skin-corrosion-docs/EpiSkin_INVITOX%20protocol%20corrosivity%20test_OECD.pdf. Accessed on 10.05.2016.

ESAC (1998a). Statement on the scientific validity of the EPISKIN™ test (an *in vitro* test for skin corrosivity). *ATLA* **26**, 277-280.

ESAC (1998b). Statement on the scientific validity of the rat skin transcutaneous electrical resistance (TER) test (an *in vitro* test for skin corrosivity). *ATLA* **26**, 275-277.

ESAC (2000). Statement on the application of the Epiderm™ human skin model for skin corrosivity testing. *ATLA* **28**, 365-366.

ESAC (2001). Statement on the application of the CORROSITEX® assay for skin corrosivity testing. *ATLA* **29**, 96-97.

ESAC (2006). Statement on the application of the SkinEthic™ human skin model for skin corrosivity testing. Available at: http://ecvam.jrc.ec.europa.eu/ under "Publications", "ESAC statements". Accessed on 18.09.2009.

ESAC (2007) Statement on the validity of *in vitro* tests for skin irritation. Available at: http://ecvam.jrc.ec.europa.eu/ under "Publications", "ESAC statements". Accessed on 18.09.2009.

ESAC (2008). Statement on the Scientific Validity of *In vitro* tests for Skin Irriation testing (SkinEthic and modified EpiDerm). Available at: http://ecvam.jrc.ec.europa.eu/ under "Publications", "ESAC statements", accessed on 18.09.2009.

ESAC (2009a). ESAC Statement on the scientific validity of an *in vitro* test method for skin corrosivity testing. Available at: http://ecvam.jrc.ec.europa.eu/ under "Publications", "ESAC statements". Accessed on 18.09.2009.

ESAC (2009b). Statement on the Performance under UN GHS of three *in vitro* assays for Skin Irritation testing and the Adaptation of the Reference chemicals and defined accuracy values of the ECVAM Skin Irritation Performance Standards. Available at: http://ecvam.jrc.ec.europa.eu/ under "Publications", "ESAC statements". Accessed on 18.09.2009.

Eskes C, Cole T, Hoffmann S, Worth A, Cockshott A, Gerner I, Zuang V (2007). The ECVAM international validation study on *in vitro* tests for acute skin irritation: selection of test chemicals. *ATLA* **35**, 603-619.

Eskes C., Detappe V., Koeter H., Kreysa J., Liebsch M., Zuang V., Amcoff P., Barroso J., Cotovio J., Guest R., Hermann M., Hoffmann S., Masson P., Alepee N., Arce L.A., Bruschweiler B., Catone T., Cihak R., Clouzeau J., D'Abrosca F., Delveaux C., Derouette J.P., Engelking O., Facchini D., Frohlicher M., Hofmann M., Hopf N., Molinari J., Oberli A., Ott M., Peter R., Sa-Rocha V.M., Schenk D., Tomicic C., Vanparys P., Verdon B., Wallenhorst T., Winkler G.C. and Depallens O. (2012). Regulatory Assessment of *In Vitro* Skin Corrosion & Irritation Data within the European Framework: Workshop Recommendations. *Regulatory Toxicology & Pharmacology,* **62**, 393-403.

Fentem, J. H., Archer, G. E. B., Balls, M., Botham, P. A., Curren, R. D., Earl, L. K., Esdaile, D. J., Holzhütter, H. G. & Liebsch, M. (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.

Fentem J.H., Briggs D., Chesné C., Elliott G.R., Harbell J.W., Heylings J.R., Portes P., Roguet R., Van De Sandt J.J.M., Botham P.A. (2001). A prevalidation study on *in vitro* tests for acute skin irritation: results and evaluation by the Management Team. *Toxicology in vitro* **15**, 57-93.

Gallegos Saliner A., Patlewicz G., Worth A.P. (2006). Review of literature-based models for skin and eye irritation and corrosion. JRC report EUR 22320 EN. European Chemicals Bureau, Ispra, Italy.

Gallegos Saliner A, Patlewicz G & Worth AP (2008). A Review of (Q)SAR Models for Skin and Eye Irritation and Corrosion. QSAR & Combinatorial Science 27, 49-59.

Gerner I., Schlegel K., Walker J.D., Hulzebos E. (2004). Use of physicochemical property limits to develop rules for identifying chemical substances with no skin irritation or corrosion potential. QSAR Comb. Sci. 23, 726-733.

Garle MJ, Fry JR. (2003). Sensory nerves, neurogenic inflammation and pain: missing components of alternative irritation strategies? A review and a potential strategy. *ATLA* **31**, 295-316.

Golla S., Madihally S., Robinson Jr., Gasem K.A.M. (2009). Quantitative structure-property relationships modeling of skin irritation. Toxicology *in vitro* 23, 176-184.

Groeber F, Schober L, Schmid FF, Traube A, Kolbus-Hernandez S, Daton K, Hoffmann S, Petersohn D, Schäfer-Korting M, Walles H, Mewes KR. (2016). Catch-up validation study of an *in vitro* skin irritation test method based on an open source reconstructed epidermis (phase II). *Toxicol In Vitro* **36**, 254-261.

Groux H., Lelievre D., Cottrez F., Auriault C., Alépée N., Meunier J.R., Cotovio J. (2012). Evaluation of IRR-IS®, an EpiSkinTM based model for quantifying chemical irritation potency. Oral Communication Nº 1.5, ESTIV2012, 17th October 2012, Lisbon, Portugal. Poster presented at 51st Annual Meeting of the Society of Toxicology, March 11-15, 2012, San Francisco, USA. The Toxicologist, p. 286. Available at: https://www.toxicology.org/pubs/docs/Tox/2012Tox.pdf. Accessed on 10.11.2016.

Hall-Manning T.J., Holland G.H., Basketter D.A., Barratt M.D. (1995). Skin irritation potential of mixed surfactant systems in a human 4 hour covered patch test. Allergologie 18, 465.

Hayashi M., Nakamura Y., Higashi K., Kato H., Kishida F., Kaneko H. (1999). A quantitative structure-activity relationship study of the skin irritation potential of phenols. Toxicology *in vitro* 13, 915-922.

Hewitt NJ, Edwards RJ, Fritsche E, Goebel C, Aeby P, Scheel J, Reisinger K, Ouédraogo G, Duche D, Eilstein J, Latil A, Kenny J, Moore C, Kuehnl J, Barroso J, Fautz R, Pfuhler S. (2013). Use of human *in vitro* skin models for accurate and ethical risk assessment: metabolic considerations. *Toxicol. Sci.* **133**, 209-217.

Hill A.B. (1965). The environment and disease: association or causation? Proc. R. Soc. Med. 58, 295−300.

Hoffmann J., Heisler E., Karpinski S., Losse J., Thomas D., Siefken W., Ahr H.-J., Vohr H.-W., Fuchs H.W. (2005). Epidermal-skin-test 1000 (EST-1000)—A new reconstructed epidermis for *in vitro* skin corrosivity testing. *Toxicology in Vitro* **19**, 925–929.

Hoffman R.S. (2007). Understanding the limitations of retrospective analyses of poison center data. Clin. Toxicol. 45, 943–945.

Hoffmann S, Saliner AG, Patlewicz G, Eskes C, Zuang V, Worth AP (2008). A feasibility study developing an integrated testing strategy assessing skin irritation potential of chemicals. *Toxicol Lett.* **180**, 9-20.

Hofmann S. (2011). Analysis of the ECETOC database on skin irritation and corrosion exploring the effects of a sample size reduction. Document prepared for the OECD ECM on skin irritation corrosion at ECHA, Helsinki, on 18-19 January 2012. Available at: https://community.oecd.org/docs/DOC-36259.

Houthoff E, Rugen P, Hart D (2015). Predictability of *in vitro* dermal assays when evaluating fatty amine derivatives. Toxicol *In vitro* 29, 1263-1267.

Hulzebos E.M., Janssen P.A.H., Maślankiewicz L., Mijerink M.C.M., Muller J.J.A., Pelgrom S.M.G., Verdam L., Vermeire T.G. (2001). The application of structure-activity relationships in human hazard assessment: a first approach. National Institute of Public Health and the Environment. RIVM report 601516 008, Bilthoven, The Netherlands.

Hulzebos E.M., Maslankiewicz L., Walker J.D. (2003). Verification of literature-derived SARs for skin irritation and corrosion. QSAR Comb. Sci. 22, 351–363.

Hulzebos E., Walker J.D., Gerner I., Schlegel K. (2005). Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. QSAR Comb. Sci. 24, 332-342.

Hulzebos E. and Gerner I. (2010). Weight factors in an Integrated Testing Strategy using adjusted OECD principles for (Q)SARs and extended Klimisch codes to decide on skin irritation classification. *Regulatory Toxicology and Pharmacology* **58**, 131-144.

Ishii S., Ishii K., Nakadate M., Yamasaki K. (2013). Correlation study in skin and eye irritation between rabbits and humans based on published literatures. Food and Chemical Toxico. 55, 596-601.

Jírová D., Liebsch M., Basketter D., Spiller E., Kejlova K., Bendova H., Marriott M., Kandarova, H. (2007). Comparison of human skin irritation and photo-irritation patch test data with cellular *in vitro* assays and animal *in vivo* data. AATEX 14, 359-365.

Jírová D, Basketter D, Liebsch M, Bendová H, Kejlová K, Marriott M, Kandárová H. (2010). Comparison of human skin irritation patch test data with *in vitro* skin irritation assays and animal data. *Contact Dermatitis* **62**, 109-116.

Kartono F & Maibach H. (2006) Irritants in combination with a synergistic or additive effect on 28 the skin response: an overview of tandem irritation studies. Contact Dermatitis 54(6), 303-12.

Kandárová, H., Liebsch, M., Spielmann, H., Genschow, E., Schmidt, E., Guest, R., Whittingham, A., Warren, N., Gamer, A., Remmele, M., Kaufmann, T., Wittmer, E., De Wever, B., Rosdy, M., (2006). Assessment of the SkinEthic Reconstituted Human Epidermis for skin corrosion testing according to OECD guideline 431. *Toxicology in vitro* **20**, 547–559.

Kandárová H, Hayden P, Klausner M, Kubilus J, Sheasgreen J. (2009a). An *in vitro* skin irritation test (SIT) using the EpiDerm reconstructed human epidermal (RHE) model. *J Vis Exp*. **13**, pii: 1366. doi: 10.3791/1366.

Kandárová H, Hayden P, Klausner M, Kubilus J, Kearney P., Sheasgreen J. (2009b). *In vitro* skin irritation testing: improving the sensitivity of the EpiDerm skin irritation test protocol. *ATLA* **37**, 671-689.

Kandárová H., Kaluzhny Y., Kubilus J., Hayden P., Sheasgreen J., Klausner M. (2011), Retrospective Analysis of the EpiDerm 3-Minute Prediction Model for Assessment of GHS Skin Corrosion Packing Group Sub-Category 1A. MatTek Corporation, Poster 8th World Congress, Montréal, Canada.

Kandárová H., Liebsch M. (*in press*). EpiDerm Skin Corrosion Test (EpiDerm SCT). In: Alternatives for dermal toxicity testing. Eds. Eskes C., van Vliet E., Maibach H. Springer, *in press*.

Katoh M, Hamajima F, Ogasawara T, Hata K. (2009). Assessment of human epidermal model LabCyte EPI-MODEL for *in vitro* skin irritation testing according to European Centre for the Validation of Alternative Methods (ECVAM)-validated protocol. *J Toxicol Sci*. **34**,327-34.

Katoh, M. and Hata K. (2011). Refinement of LabCyte EPI-MODEL24 skin Irritation test method for adaptation to the requirements of OECD test guideline 439. *AATEX* 16, 111-122.

Kim H, Choi J, Lee H, Park J, Yoon BI, Jin SM, Park K (2016). Skin Corrosion and Irritation Test of Nanoparticles Using Reconstructed Three-Dimensional Human Skin Model, EpiDermTM. Toxicol Res 32, 311-316.

Klimisch H.J., Andreae M. , Tillmann U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology 25, 1-5.

Kodithala K., Hopfinger A.J., Thompson E.D., Robinson M.K. (2002). Prediction of skin irritation from organic chemicals using membrane-interaction QSAR analysis. Toxicol.Sci. 66, 336-346.

Kojima, H., Ando, Y., Idehara, K., Katoh, M., Kosaka, T., Miyaoka, E., Shinoda, S., Suzuki, T., Yamaguchi, Y., Yoshimura, I., Yuasa, A., Watanabe, Y. and Omori, T. (2012). Validation Study of the *In vitro* Skin Irritation Test with the LabCyte EPI-MODEL24. *ATLA* 40, 33-50.

Kojima, H., Katoh, M. Shinoda, S., Hagiwara S., Suzuki, T., Izumi R., Yamaguchi, Y., Nakamura M., Kasahawa T., Shibai A. (2014). A catchu-up validation study of an *in vitro* skin irritation test method using reconstructed human epidermis LabCyte EPI-MODEL24. *J. Appl. Toxicol.* 34, 766-774.

Kolle S.N., Sullivan K.M., Mehling A., van Ravenzwaay B., Landsiedel R. (2013). Erratum to "Applicability of *in vitro* tests for skin irritation and corrosion to regulatory classification schemes: Substantiating test strategies with data from routine studies". *Regulatory Toxicology and Pharmacology* 65, 366-378

Khammo N, Ogilvie J, Clothier RH (2007). Development of an innervated model of human skin. *ATLA* **35**, 487-491.

Kinsner-Ovaskainen A, Akkan Z, Casati S, Coecke S, Corvi R, Dal Negro G, De Bruijn J, De Silva O, Gribaldo L, Griesinger C, Jaworska J, Kreysa J, Maxwell G, McNamee P, Price A, Prieto P, Schubert R, Tosti L, Worth A, Zuang V. (2012) Report of the EPAA-ECVAM workshop on the validation of Integrated Testing Strategies (ITS). *ATLA* **40**, 175-181.

Labcyte EPI-MODEL 24SIT (2011). EPI-MODEL24 SIT SOP, Version 8.3, Skin Irritation Test Using the Reconstructed Human Model "LabCyte EPI-MODEL24". Available at: http://www.jpte.co.jp/english/business/LabCyte/Testprotocol.pdf. Accessed on 11.05.2016.

Liebsch, M., Traue, D., Barrabas, C., Spielmann, H., Uphill, P., Wilkins, S., Wiemann, C., Kaufmann, T., Remmele, M. & Holzhütter, H. G. (2000). The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. *ATLA* **28**, 371-401.

Liew C.Y., Yap C.W. (2013). QSAR and predictors of eye and skin effects. Molecular Informatics 32, 281-290.

Lönnqvist S, Briheim K, Kratz G (2016). Non-occlusive topical exposure of human skin *in vitro* as model for cytotoxicity testing of irritant compounds. Toxicol Mech Methods **26**, 82-87.

Louekari K., Sihvonen K., Kuittinen M., Somnes V. (2006). *In vitro* tests within the REACH information strategies. *ATLA* **34**, 377-389.

Macfarlane M., Jones P., Goebel C., Dufour E., Rowland J., Araki D., Costabel-Farkas M., Hewitt N.J., Hibatallah J., Kirst A., McNamee P., Shellauf F., Scheel J. (2009). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: Skin irritation. *Regulatory Toxicology and Pharmacology* **54**, 188-196.

Mewes KR, Fischer A, Zöller NN, Laubach V, Bernd A, Jacobs A, van Rompay A, Liebsch M, Pirow R, Petersohn D. (2016). Catch-up validation study of an *in vitro* skin irritation test method based on an open source reconstructed epidermis (phase I). Toxicol *In Vitro* **36**, 238-253.

Miles A, Berthet A, Hopf NB, Gilliet M, Raffoul W, Vernez D, Spring P (2014). A new alternative method for testing skin irritation using a human skin model: a pilot study. Toxicol *In Vitro* 28, 240-247.

Miyani VA, Hughes MF (2016). Assessment of the *in vitro* dermal irritation potential of cerium, silver, and titanium nanoparticles in a human skin equivalent model. Cutan Ocul Toxicol 8,1-7.

Molinari J, Eskes C, Andres E, Remoué N, Sá-Rocha VM, Hurtado SP, Barichello C. (2013). Improved procedures for *in vitro* skin irritation testing of sticky and greasy botanicals. TIV 27: 441-450.

Mombelli E. (2008). An evaluation of the predictive ability of the QSAR software packages, DEREK, HAZARDEXPERT and TOPKAT, to describe chemically-induced skin irritation. *ATLA* 36, 15-34.

Nangia A., Anderson P.H., Berner B., Maibach H.I. (1996). High dissociation constants (pKa) of basic permeants are associated with *in vivo* skin irritation in man. Contact Dermatitis 34, 237-242.

NIH (1999). *Corrositex®: an in vitro test method for assessing dermal corrosivity potential of chemicals*. NIH Publication No. 99-4495. Research Triangle Park, NC, USA: NIEHS.

Niwa M, Nagai K, Oike H, Kobori M (2009). Evaluation of the skin irritation using a DNA microarray on a reconstructed human epidermal model. Biol Pharm Bull **32**, 203-208.

OECD (1981a). Repeated Dose Dermal Toxicity: 21/28-day Study. OECD Guideline for the Testing of Chemicals No. 410, OECD, Paris.

OECD (1981b). Subchronic Dermal Toxicity: 90-day Study. OECD Guideline for the Testing of Chemicals No. 411, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (1987). Acute Dermal Toxicity. OECD Guideline for the Testing of Chemicals No. 402, OECD, Paris.

OECD (1992). OECD Guidelines for the Testing of Chemicals No. 406: Skin Sensitisation. OECD, Paris.

OECD (2004a). Guidelines for the Testing of Chemicals 427. Skin absorption: *in vivo* method. OECD, Paris.

OECD (2004b). Guidelines for the Testing of Chemicals 428. Skin absorption: *in vitro* method. OECD, Paris.

OECD (2005) Guidance Document No. 34: Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environmental Health and Safety Monograph Series on Testing and Assessment. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on 10.05.2016.

OECD (2007a). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models OECD Series on Testing and Assessment No. 69, OECD, Paris.

OECD (2007b). Guidance on Grouping of Chemicals. OECD Series on Testing and Assessment No. 80, OECD, Paris.

OECD. (2010a) Document No. 137: Explanatory background document to the OECD Test Guideline on *in vitro* skin irritation testing. Environment, Health, and Safety Publications, Series on Testing and Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on: 10.05.2016.

OECD (2010b).OECD Guidelines for Chemical Testing No. 429. Skin sensitization: Local Lymph Node assay. OECD, Paris.

OECD (2010c). OECD Guidelines for Chemical Testing No. 442A. Skin sensitization: Local Lymph Node assay: DA. OECD, Paris.

OECD (2010d). OECD Guidelines for Chemical Testing No. 442B. Skin sensitization: Local Lymph Node assay: BrdU-ELISA. OECD, Paris.

OECD. (2011) Document No. 155: Peer Review Report of Validation of the Skin Irritation Test Using LabCyte EPI-MODEL24. Environment, Health and Safety Publications, Series on Testing and Assessment, Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on: 10.05.2016.

OECD. (2013) Document No. 190: 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. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on: 10.05.2016.

OECD (2014). Guidance document No. 203: Integrated Approached to Testing and Assessment for Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on: 25.4.2016.

OECD (2015a). *OECD Guidelines for the Testing of Chemicals No. 404. Acute Dermal Irritation/Corrosion.* Paris, France: Organisation for Economic Cooperation and Development. Originally adopted in 1981. Available at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Accessed on 10.05.2016.

OECD (2015b). *OECD Guidelines for the Testing of Chemicals No. 430. In vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER).* Paris, France: Organisation for Economic Cooperation and Development. Originally adopted in 2004. Available at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Accessed on 10.05.2016.

OECD (2015c) *OECD Guidelines for the Testing of Chemicals No. 435. In vitro Membrane Barrier Test Method for Skin Corrosion.* Paris, France: Organisation for Economic Cooperation and Development. Originally adopted in 2006. Available at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Accessed on 10.05.2016.

OECD (2015d). *OECD Guidelines for the Testing of Chemicals No. 439. In vitro Skin Irritation: Reconstructed Human Epidermis Test Method.* Paris, France: Organisation for Economic Cooperation and Development. Originally adopted in 2010. Available at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Accessed on 10.05.2016.

OECD (2015e). Document No. 218: Performance Standards for the Assessment of Proposed Similar or Modified Test Methods for Skin Corrosion Testing as Described in TG 430. OECD Environmental Health and Safety Publications, Series on Testing and Assessment. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on 10.05.2016.

OECD (2015f). Document 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. OECD Environmental Health and Safety Publications, Series on Testing and Assessment. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on 10.05.2016.

OECD (2015g). Assessment of the performance characteristics of proposed *in vitro* membrane barrier test method for skin corrosion. Document under revision. Available at: http://www.oecd.org/chemicalsafety/testing/PerfStand-TG435-June2014.pdf. Accessed on 11.05.2016.

OECD (2015h). Document No. 220: Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human Epidermis (RHE) Test Methods for Skin Irritation Testing as Described in TG 439. OECD Environmental Health and Safety Publications, Series on Testing and Assessment. Available at: http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm. Accessed on 10.05.2016.

OECD (2016) OECD Guidelines for the Testing of Chemicals No. 431. *In vitro* skin corrosion: reconstructed human epidermis (RHE) test method. Paris, France: Organisation for Economic Cooperation and Development. Originally adopted in 2004. Available at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788. Accessed on 11.11.2016.

Patlewicz G., Rodford R., Walker J.D. (2003). Quantitative structure-activity relationships for predicting skin and eye irritation. Environ. Toxicol. Chem. 22,1862-1869.

Patrick and Maiback H (1994). Chapter 21. Dermatology. In Principles and methods of toxicology. Third edition (Hayes A.W. ed.). Raven Press, New York, 767-803.

Portes P., Grandidier M.H., Cohen C., Roguet R. (2002). Refinement of the EPISKIN protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study. *Toxicology In vitro* **16**, 765-770.

Robinson M .K., Perkins M.A., Basketter D.A. (1998). Application of a four hour human patch test method for comparative and investigative assessment of skin irritation. Contact Dermatitis 38, 194 - 202.

Robinson M.K., Osborne R., Perkins M.A. (2000). *In vitro* and human testing strategies for skin irritation. Ann. NY Acad. Sci. 919, 192-204.

Robinson M.K., McFadden J.P., Basketter D.A. (2001). Validity and ethics of the human 4 hour patch test as an alternative method to assess acute skin irritation potential. Contact Dermatitis 45, 1–12.
Robinson M.K. (2002). Population differences in acute skin irritation responses. Race, sex, age, sensitive skin and repeat subject comparisons. Contact Dermatitis 46, 86–93.

Robinson M.K., Cohen C., de Brugerolle de Fraissinette A., Ponec M., Whittle E., Fentem J.H. (2002). Non-animal testing strategies for assessment of the skin corrosion and skin irritation potential of ingredients and finished products. *Food and Chemical Toxicology* **40**, 573-592.

Robinson M.K., Kruszewski F.H., Al-Atrash J., Blazka M.E., Gingell R., Heitfeld F.A., Mallon D., Snyder N.K., Swanson J.E., Casterton P.L. (2005). Comparative assessment of the acute skin irritation potential of detergent formulations using a novel human 4-h patch test method. Food Chem Toxicol 43, 1703–1712.

Rorije E., Aldenberg T., Buist H., Kroese D., Schüürmann G. (2013). The OSIRIS Weight of Evidence approach: ITS for skin sensitization. Regul Toxicol Pharmacol. 67, 146-156.

Scheel J, Heppenheimer A, Lehringer E, Kreutz J, Poth A, Ammann H, Reisinger K, Banduhn N. (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.

Schneider K., Schwarz M., Burkholder I., Kopp-Schneider A., Edler L., Kinsner-Ovaskainen A., Hartung T., Hoffmann S. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol Lett. 189, 138-144.

SkinEthic™ RHE SIT-42bis (2009). SOP Version 2.0. SkinEthic Skin Irritation Test-42bis Test Method for the Prediction of Acute Skin Irritation of Chemicals: 42 Minutes Application + 42 Hours Post-Incubation. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/docs-skin-irritation-1/DOC14-SOP%20SkinEthic%20MTT.pdf. Accessed on 11.05.2016.

SkinEthic™ RHE (2012). INVITTOX Protocol SkinEthic™ Skin Corrosivity Test. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/eurl-ecvam/validation-regulatory-acceptance/skin-corrosion-docs/SkinEthic_INVITTOX_protocol_corrosivity_test_OECD_Jan2012.pdf. Accessed on 10.05.2016.

Smith J.S., Macina O.T., Sussman N.B., Luster M.I., Karol M.H. (2000a). A robust structure-activity relationship (SAR) model for esters that cause skin irritation in humans. Toxicological Sciences 55, 215-222.

Smith J.S., Macina O.T., Sussman N.B., Karol M.H., Maibach H.I. (2000b). Experimental validation of a structure-activity relationship model of skin irritation by esters. Quantitative Structure-Activity Relationships 19, 467–474.

Spiekstra SW, Dos Santos GG, Scheper RJ, Gibbs S. (2009). Potential method to determine irritant potency *in vitro* - Comparison of two reconstructed epidermal culture models with different barrier competency. *Toxicology In vitro* **23**, 349-355.

Spielmann H, Hoffmann S, Liebsch M, Botham P, Fentem JH, Eskes C, Roguet R, Cotovio J, Cole T, Worth A, Heylings J, Jones P, Robles C, Kandárová H, Gamer A, Remmele M, Curren R, Raabe H, Cockshott A, Gerner I, Zuang V (2007). The ECVAM international validation study on *in vitro* tests for acute skin irritation: report on the validity of the EPISKIN and EpiDerm assays and on the Skin Integrity Function Test. *ATLA* **35**, 559-601.

Swiss Federal Council (2015). Ordinance n. 813.11 on Protection against Dangerous Substances and Preparations (Chemicals Ordinance, ChemO), of 5 June 2015 (status as of 1 February 2016). Available at : https://www.admin.ch/opc/en/classified-compilation/20141117/index.html. Accessed on 10.11.2016.

TER (2008). DB-ALM Protocol No. 115. SOP on the rat skin Transcutaneous Electrical Resistance (TER) Test. Available at: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance/skin-corrosion-docs/115.pdf. Accessed on 11.05.2016.

Tornier C. Amsellem C., de Brugerolle de Fraissinette A. Alépée N. (2010). Assessment of the optimized SkinEthic Reconstructed Human Epidermis (RHE) 42bis skin irritation protocol over 39 substances. *Toxicology In Vitro* 24, 245-256.

United Nations (UN) (2015a). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Sixth revised edition. Part 3: Health Hazards – Chapter 3.2 Skin corrosion/irritation. New York, USA, and Geneva, Switzerland: United Nations. p.125-136. Available at: http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev06/English/ST-SG-AC10-30-Rev6e.pdf. Accessed on 25.04.2016.

UN (2015b). Volume I - Part 2. Classification. In Recommendations on the Transport of Dangerous Goods - Model Regulations. Nineteenth revised edition. Available at: http://www.unece.org/trans/danger/publi/unrec/rev19/19files_e.html. Accessed on 26.04.2015.

Van de Sandt J., Roguet R., Cohen C., Esdaile D., Ponec M., Corsini E., Barker C., Fusenig N., Liebsch M., Benford D., de Brugerolle de Fraissinette A., Fartasch M. (1999). The Use of Human Keratinocytes and Human Skin Models for Predicting Skin Irritation. The Report and Recommendations of ECVAM Workshop 38. *ATLA* **27**, 723-743.

Walker J.D., Gerner I., Hulzebos E., Schlegel K. (2004). (Q)SARs for Predicting Skin Irritation and Corrosion: Mechanisms, Transparency and Applicability of Predictions. QSAR & Combinatorial Science 23, 721-725.

Walker J.D., Gerner I., Hulzebos E., Schlegel K. (2005). The skin irritation corrosion rules estimation tool (SICRET)". QSAR Comb. Sci. 24, 378-384.

Weil C.S., Scala A. (1971). Study of intra- and inter- laboratory variability in the results of rabbit eye and skin irritation tests. Toxicology and Applied Pharmacoogy 19, 276-360.

Welss T., Basketter D.A., Schroder K.R. (2004). *In vitro* skin irritation: facts and future. State of the art review of mechanisms and models. Toxicology *in vitro* 18, 231-243.

Welss T, Matthies W, Schroeder KR. (2007). Compatibility testing *in vitro*: a comparison with *in vivo* patch test data. *Int J Cosmet Sci* **29**, 143.

Whittle E., Barratt M.D., Carter J.A., Basketter D.A., Chamberlain M. (1996). Skin corrosivity potential of fatty acids: *In vitro* rat and human skin testing and QSAR studies. Toxicology *in vitro* 10, 95-100.

Worth A.P., Fentem J.H., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdail D.J., Liebsch M. (1998). An evaluation of the proposed OECD Testing Strategy for skin corrosion. ATLA 26, 709-720.

Worth A.P., Cronin M.T.D. (2001). The use of bootstrap resampling to assess the variability of Draize tissue scores. ATLA 29, 557-573.

York M., Wilson A.P., Newsome C.S. (1994). The classification of soluble silicates for eye hazard using the enucleated rabbit eye test; Toxicology *in vitro*; 8 (6); 1265-1268.

York M., Griffiths H.A., Whittle E., Basketter D.A. (1996). Evaluation of a human patch test for the identification and classification of skin irritation potential. Contact Dermatitis 34, 204– 212.

Young J.R., How M.J., Walker A.P., Worth W.M.H. (1988). Classification as corrosive or irritant to skin of preparations containing acidic or alkaline substances, without testing on animals. Toxicology *in vitro* 2, 19-26.

Young J.R., How M.J. (1994). Product classification as corrosive or irritant by measuring pH and acid/alkali reserve. In Alternative Methods in Toxicology vol. 10 - *In vitro* Skin Toxicology: Irritation, Phototoxicity, Sensitization, eds. A.Rougier, A.M. Goldberg and H.I Maibach, Mary Ann Liebert, Inc. 23-27.

Zhang Q, Dai T, Zhang L, Zhang M, Xiao X, Hu H, Zou P, Liu X, Xiang Q, Su Z, Huang Y, He QY (2011). Identification of potential biomarkers for predicting acute dermal irritation by proteomic analysis. J Appl Toxicol *31*, 762-772.

Zuang V., Balls M., Botham P.A., Coquette A., Corsini E., Curren R.D., Elliott G.R., Fentem J.H., Heylings J.R., Liebsch M., Medina J., Roguet R., van de Sandt H., Wiemann C., Worth A.P. (2002). Follow-up to the ECVAM Prevalidation Study on *In vitro* Tests for Acute Skin Irritation. ECVAM Skin Irritation Task Force Report 2. *ATLA* **30**, 109-129.