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

11-Jul-2014

English - Or. English

ENV/JM/MONO(2014)19
Unclassified

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

NEW GUIDANCE DOCUMENT ON AN INTEGRATED APPROACH ON TESTING AND ASSESSMENT (IATA) FOR SKIN CORROSION AND IRRITATION

Series on Testing and Assessment

No. 203

JT03360405

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 203

GUIDANCE DOCUMENT ON AN INTEGRATED APPROACH ON TESTING AND ASSESSMENT (IATA) FOR SKIN CORROSION AND IRRITATION

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

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2014

About the OECD

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

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

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

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

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

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

or contact:

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

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

© OECD 2014
Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

FOREWORD

This Guidance Document *on Integrated Approach to Testing and Assessment for Skin Irritation and Corrosion* has two aims:

- It proposes an integrated approach on testing and assessment (IATA) for skin corrosion and irritation, in view of replacing the "testing and evaluation strategy" which is currently provided in the supplement to OECD TG 404 and which requires adaptation to scientific and technical progress.

- It provides consistent information on key performance characteristics of each of the individual information sources comprising the IATA, provides guidance on how to integrate information for decision making within the approach (including decisions on the need for further testing) and on integrating all existing and generated information on the corrosive and irritant hazard potential of test chemicals for final decisions for classification and labelling.

The Guidance Document was approved by the Working Group of the National Co-ordinators of the Test Guidelines Programme (WNT) at its 26th meeting in April 2014. The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to its declassification on 7th July, 2014.

This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

TABLE OF CONTENTS

I. INTRODUCTION TO THE IATA FOR SKIN CORROSION AND IRRITATION

1. Since 2002, the OECD OECD TG 404 on in vivo acute dermal irritation and corrosion testing (OECD, 2002) contains a supplement describing a sequential testing and evaluation strategy for skin corrosion and irritation. While this supplement is not covered by the OECD Council decision on Mutual Acceptance of Data (MAD), it has nevertheless provided valuable guidance on how to consider existing information and organise the generation of new testing data on skin corrosion/irritation. Steps 5 and 6 of this sequential testing and evaluation strategy call for validated and accepted in vitro or ex vivo test methods for skin corrosion and skin irritation, respectively, before the use of the in vivo OECD TG 404 in step 7, with the purpose of minimising animal use. However this strategy does not foresee the use of negative results from validated and accepted in vitro assays but requires confirmatory in vivo testing in such cases. Since publication of the supplement in 2002, several Test Guidelines on in vitro methods for skin corrosion or irritation have been published and/or updated, notably OECD TG 439 (OECD, 2013a) on in vitro skin irritation and OECD TGs 430 (OECD, 2013b), 431 (OECD, 2013c) and 435 (OECD, 2006) on in vitro skin corrosion. Depending on country requirements, the now available validated and OECD accepted in vitro methods may satisfy all information requirements for skin corrosion and irritation. In addition, non-standards methods (i.e. not yet validated and accepted by OECD) may provide further information required by some authorities, e.g. on full sub-categorisation of corrosives and predictions of the optional Cat. 3 for mild irritants. Although the suitability of such data for regulatory purposes needs to be judged case by case, they should be considered before conducting animal studies. For these reasons, guidance in relation to the use and generation of data for skin corrosion and irritation requires update in view of amending the possible use and usefulness of individual test methods described within this strategy and in order to avoid contradiction between the provisions of individual OECD TGs on in vitro methods and the provisions of the OECD TG 404 supplement. Moreover, in view of growing experience with the composition and use of IATAs, in particular for this specific human health endpoint, a revision in view of incorporating current scientific and regulatory considerations and practices seems timely.

2. In June 2009, during an OECD Expert Consultation Meeting on skin irritation, experts recommended that the OECD TG 404 be updated (OECD, 2010a: Annexe 7, page 158). In March 2010, WNT22 approved a project proposal from Germany to develop a Guidance Document (GD) for an Integrated Approach on Testing and Assessment (IATA) for skin corrosion and irritation. A first Expert Consultation Meeting (ECM) was held in Berlin in October 2010. The overall purpose of the first meeting was to prepare the development of a GD for such an IATA and to work towards recommendations to the WNT to revise, delete or merge any of the existing skin irritation and corrosion OECD TGs. This initial effort has been followed by Expert Consultation Meetings (ECM) held in Helsinki in January 2012, in Paris in September 2012 and in Berlin (Germany) in December 2013.

3. The general objective of the GD is to establish an IATA for hazard identification of skin corrosion or irritation potential of chemicals (or the absence thereof) that provides adequate information for classification and labelling according to the United Nations Globally Harmonised System (UN GHS). The IATA is composed of well described and characterised “Modules”, each of which containing one to several individual information sources of similar type. 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 are described with the purpose of minimizing the use of animals to the extent possible, while ensuring human safety.

The OECD Sequential Testing and Evaluation Strategy

4. The supplement of OECD TG 404 testing strategy adopted in 2002 consists of a sequential order of eight steps (OECD, 2002). If at a given step no conclusion can be reached, the next step of the strategy is considered. These steps sequentially address 1) existing human and/or animal data, 2) Structure-Activity Relationships (SAR), 3) pH, 4) systemic toxicity via dermal route, 5) the use of validated and accepted in vitro or ex vivo tests for skin corrosion, 6) the use of validated and accepted in vitro or ex vivo tests for skin irritation, and 7/8) the use of a confirmatory in vivo rabbit test in a stepwise manner if a negative result is obtained with the in vitro/ex vivo skin irritation tests. As the sequential testing strategy does not fall under MAD, it is not binding to OECD member countries and should therefore be considered only as a recommendation. Note that the testing strategy described in the supplement of OECD TG 404 has inspired the tiered testing described in Chapter 3.2 of the UN Globally Harmonised System (GHS) (UN GHS for skin irritation and corrosion).

The UN GHS Sequential Testing and Evaluation Strategy

5. The United Nations Globally Harmonised System for classification also proposed in the past a tiered testing approach which was similar to the one proposed by the OECD OECD TG 404 and included as a last step and when ethical, a human test if the test material has been shown to be non-irritant and non-corrosive in the in vivo test (UN, 2003, 2011). Such strategy has been recently considerably revised (UN, 2013), so that the UN GHS now proposes a tiered approach that provides guidance on how to organise existing information on a substance or mixture (see sections 3.2.2.2 and 3.2.3.1.1, UN, 2013) and to make a weight of evidence decision about hazard assessment and hazard classification (ideally without conducting new animal tests).

6. Such approach includes the evaluation, if available, of: 1) existing human or animal skin corrosion/irritation data, 2) other existing skin data in animals, 3) existing ex vivo / in vitro data, 4) pH-based assessment (with consideration of acid/alkaline reserve of the substance), 5) validated SAR methods, and 6) consideration of the total weight of evidence. Although information might be gained from the evaluation of single parameters within a tier, it is recommended that consideration is given to the totality of existing information and making an overall weight of evidence determination, especially when there is conflict in information available on some parameters (UN, 2013).

The ECHA Integrated Testing Strategy

7. Within the European Union, the European Chemicals Agency (ECHA) proposes a sequential strategy for skin irritation and/or corrosion in Chapter R.7a of its Guidance on information requirements and Chemical Safety Assessment under the REACH Regulation (ECHA, 2013). This Integrated Testing Strategy (ITS) has been developed during the REACH implementation project, with most of the building blocks being similar to the ones recommended within the supplement of the OECD OECD TG 404. The ITS provides guidance on how various types of available data should be evaluated, and addresses additional aspects on some elements such as the use of other toxicity data or weight of evidence (WoE) analysis of existing and relevant data. In addition, validated and accepted in vitro tests can be used to identify non-irritants and non-corrosives, in order to avoid any in vivo test for skin corrosion and irritation.

The Berlin Expert Consultation Meeting in 2010

8. In 2010, the OECD started an initiative to develop a) a GD on an IATA for skin corrosion and irritation and b) recommendations to the WNT for potential revisions, deletions and merging of existing in vivo and in vitro skin irritation and corrosion OECD TGs, i.e., OECD OECD TGs 404, 430, 431, 435 and 439. The major aspects addressed comprised:

- actual use of the OECD TGs by industry and regulatory authorities;
- strengths and limitations of the individual OECD TGs;
- the applicability domains (AD) of the OECD TGs in particular addressing chemical classes;
- suitability of the OECD TGs for mixtures and preparations;
- development of new performance standards for OECD TGs 430 & 431;
- the occurrence of false negative corrosives in OECD TGs 430, 431 and 435, and the results obtained with these chemicals using OECD TG 439;
- the adaptation of the IATA to the progress achieved with validated in vitro tests and non-testing methods (NTM), including (Quantitative) Structure-Activity Relationships ((Q)SARs).

9. The ECM agreed that in general the ITS developed during the REACH implementation project in 2006/2007 and subsequently published by ECHA (ECHA, 2013), with its step-wise procedure (data retrieval followed by WoE approach, and then, if necessary, additional testing), was suitable as a template for the development of the new OECD IATA.

II. COMPOSITION OF THE IATA FOR SKIN CORROSION AND IRRITATION

10. The ECM proposed to develop a modular approach, grouping the various individual information sources of the IATA in "Modules" according to the type of information provided. Each of the individual information sources were described in a consistent manner in terms of its applicability, limitations and performance characteristics. Eight Modules were identified as necessary elements of the IATA, which can be subsumed in three major Parts as described in Table 1.

Table 1: Parts and Modules of the IATA.

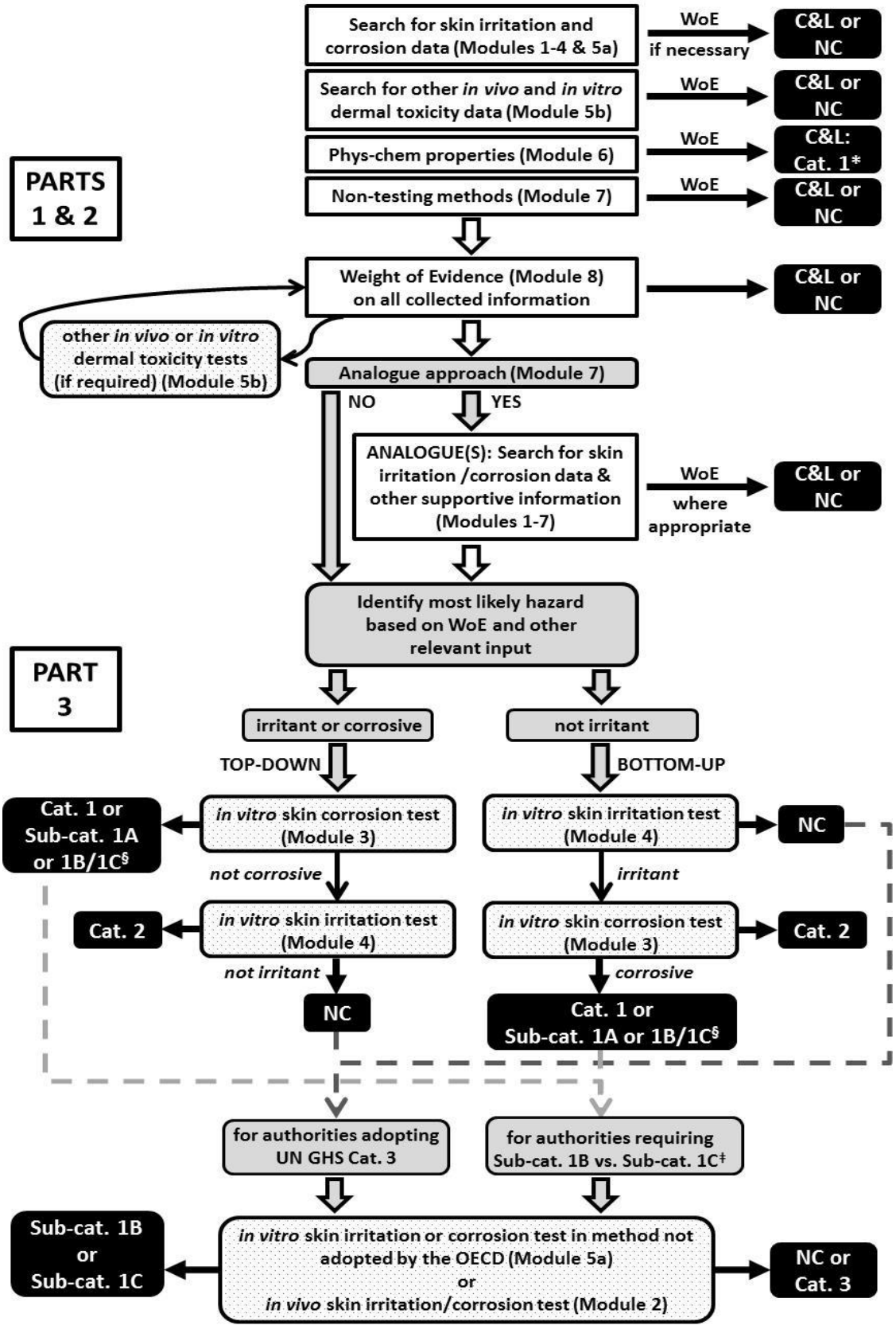
Part (*)	**Module**	**Data**
Part 1 (Existing information, physico-chemical properties and non-testing methods)	1 2 3 4 5	Existing information - Existing human data *a)* Non-standardised human data on local *skin effects* *b) Human Patch Test (HPT)* - *In vivo* skin irritation and corrosion data (OECD TG 404) - *In vitro* skin corrosion data *a) OECD TG 430* *b) OECD TG 431* *c) OECD TG 435* - *In vitro* skin irritation data (OECD TG 439) - 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 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. For more details including on Part 3, refer to Figure 1.

11. The three Parts guide the assessment of skin irritation and corrosion. Under Part 1 (existing data) of the IATA, existing and available information is retrieved from literature and databases and other reliable sources for Modules 1 to 5, while under Module 6 physico-chemical properties, primarily the pH, are considered. Module 7 covers non-testing methods. If the WoE (Part 2) is inconclusive regarding the skin irritation and corrosion potential, new testing, starting with in vitro methods, needs to be conducted (Part 3). Animal testing is foreseen only as a last resort (Figure 1).

12. A schematic outline of the IATA for skin irritation and corrosion focused on classification and labelling (C&L) is presented in Figure 1. Briefly, the information from Part 1 is evaluated in a weight of evidence approach. If the WoE is conclusive, decision for C&L can be carried out accordingly. If it is inconclusive, other in vivo or in vitro dermal toxicity tests (Module 5b) for which data are still not available but that may need to be conducted in some regulatory frameworks to satisfy other regulatory requirements, should be carried out first. 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 be considered to formulate a hypothesis of the most likely skin irritation/corrosion potential of the chemical. This hypothesis will then guide the sequence of prospective testing to a top-down or bottom-up approach.

Figure 1: Detailed IATA for skin irritation and corrosion.
*: If corrosive sub-categorisation is required an appropriate in vitro skin corrosion test needs to be conducted. In addition, for the case of the regulation of mixtures the use of additivity rules might also lead to classification as Cat.2 or NC.
§: Possibilities to sub-categorise depends on the specific test method used: OECD TG 435 allows for the discrimination between Sub-cat. 1A, Sub-cat. 1B and Sub-cat. 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 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 any sub-categorisation, i.e., Cat. 1.
ǂ: If outside the applicability domain of OECD TG 435


PARTS 1 & 2
Search for skin irritation and corrosion data (Modules 1-4 & 5a)
WoE if necessary
C&L or NC
Search for other in vivo and in vitro dermal toxicity data (Module 5b)
WoE
C&L or NC
Phys-chem properties (Module 6)
WoE
C&L: Cat. 1*
Non-testing methods (Module 7)
WoE
C&L or NC
Weight of Evidence (Module 8) on all collected information
C&L or NC
other in vivo or in vitro dermal toxicity tests (if required) (Module 5b)
Analogue approach (Module 7)
NO
YES
ANALOGUE(S): Search for skin irritation /corrosion data & other supportive information (Modules 1-7)
WoE where appropriate
C&L or NC
Identify most likely hazard based on WoE and other relevant input
PART 3
irritant or corrosive
not irritant
TOP-DOWN
BOTTOM-UP
Cat. 1 or Sub-cat. 1A or 1B/1C§
in vitro skin corrosion test (Module 3)
in vitro skin irritation test (Module 4)
NC
not corrosive
irritant
Cat. 2
in vitro skin irritation test (Module 4)
in vitro skin corrosion test (Module 3)
Cat. 2
not irritant
corrosive
NC
Cat. 1 or Sub-cat. 1A or 1B/1C§
for authorities adopting UN GHS Cat. 3
for authorities requiring Sub-cat. 1B vs. Sub-cat. 1C‡
Sub-cat. 1B or Sub-cat. 1C
in vitro skin irritation or corrosion test in method not adopted by the OECD (Module 5a) or in vivo skin irritation/corrosion test (Module 2)
NC or Cat. 3

13. The structure provided by the three Parts and the information on the eight Modules described above (Table 1) allow for composing an IATA. Ideally, this IATA should be universally applicable and ensure human safety, while making maximum use of existing data, being resource efficient and minimising or eliminating the requirement for animal experiments.

14. Acknowledging that there is different amount of information available on the applicability of the modules of this IATA to mixtures (e.g. see Part 3 – section on Assessment of mixtures) and that such applicability may depend on the information available in each specific case to be assessed, the IATA is considered applicable to both substances and mixtures.

15. While the three Parts are considered as a sequence, the Modules 1 to 7 of Part 1 might be arranged as appropriate. This will be especially helpful in cases in which information on one Module or a few Modules cannot be outweighed by any other information, so that a conclusion on the skin irritation and corrosion potential can be drawn without considering further Modules.

16. 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. However, it is stressed here that the following relative a priori weights are indicative only and will depend on the quality of the individual data in each specific case. 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:

- 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.

17. Furthermore, the retrieval of existing information groups Modules 1 to 4 and 5a, as they directly relate to skin irritation/corrosion. In contrast Module 5b requires a different search for other in vitro and in vivo dermal toxicity studies. Therefore, the search for existing data could be approached in a stepwise manner: only when the search for Modules 1 to 4 plus 5a does not result in information that allows concluding on skin irritation/corrosion potential/potency, a second search specifically for Module 5b would become necessary.

18. Some examples that would allow a straightforward and trivial WoE based on partial information in Part 1, i.e., the Modules 1 to 7, and considering the grouped stepwise search are given here-after:

- If it is known that the chemical being evaluated has an extreme pH (combined with high buffering capacity for mixtures) (Module 6) or contains a hydroperoxide group (Module 7), it can be concluded that this chemical is corrosive (Cat. 1) without searching for other existing information (Modules 1 to 5). However, if sub-categorisation is required further information will need to be collected.

- If HPT data (Module 1b) of good quality exist and, no in vivo or in vitro skin irritation/corrosion data are available (Modules 2 to 5a) or if available they are consistent with the HPT result, there is no need to evaluate Modules 5b to 7.

- If only in vivo data on skin irritation and corrosion (Module 2) of sufficient quality are available, there is no need to evaluate Modules 3 to 7.

- If only one reliable in vitro skin corrosion test is available indicating a corrosion potential there is no need to evaluate Modules 5 to 7.

- If skin irritation and corrosion information is only available for analogues(s) and a convincing read-across (Module 7) case can be made, there is no need to evaluate Modules 5 and 6.

19. The individual sources of information described in Modules 1-7 (Table 1) have been characterised as described below based on the Streamlined Summary Documents template developed for the in vitro eye test methods (OECD, 2013d,e) and comprise the following information headlines:

- Description/Definition

- Scientific basis including Mode of Action (MoA)

- Applicability domain

- Predictive capacity, e.g., expressed as sensitivity, specificity and accuracy

- Reliability, e.g., expressed as within- and between-laboratory reproducibility

- Strengths, weaknesses and limitations

- Potential role in the IATA

III. DESCRIPTION OF THE ELEMENTS OF THE IATA FOR SKIN CORROSION AND IRRITATION

A. Part 1: Existing Information, Physico-Chemical Properties and Non-Testing Methods

20. Evaluating existing data is key to avoiding unnecessary animal testing. It is also the fastest and cheapest way to arrive at a conclusion on skin irritation/corrosion potential, if the available data allow for it. In recent years, large databases have become available on the internet, e.g., the European C&L Inventory1 and the dissemination site for chemicals registered under REACH2. The Modules of Part 1 can be addressed in any order. It might not always be necessary to evaluate all of them, in particular, when the available data already allow for reliable classification into one of the GHS sub-categories for irritation or corrosion (or into the main categories, if sub-categorisation is not needed in a specific regulatory context). For Modules 1 to 5, existing information can be retrieved by a comprehensive literature and database search (e.g., the above databases hosted by ECHA). The search should be performed systematically using search terms such as CAS number or chemical name. Note that in case relevant information is identified, rights to use this information for regulatory purposes may need to be obtained. The OECD (Q)SAR Toolbox3 is a good starting point to retrieve information for Modules 6 and 7 on physico-chemical and non-testing data as it allows for the identification of analogues (for read-across), retrieval of a first set of existing experimental (phys.-chem. and toxicological) data on both the target chemical and the analogues and finally characterisation of these chemicals with mechanistic and other profilers, including structural alerts for skin irritation and corrosion. Further existing data on analogues identified with the Toolbox can then be retrieved by repeating the above literature and database search for these compounds. If not retrieved from database searches or available estimates are doubtful, pH and potentially acidity and alkalinity, as well as other physico-chemical parameters may also be measured.

Module 1 – Existing human data

21. 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 and can therefore rarely be used on its own for C&L decisions without a WoE assessment, the latter is commonly of much higher quality as it is usually acquired under standardised conditions and with strict acceptance criteria. If considered suitable and adequately documented human data, especially HPT data, should have precedence over other data. Examples of how existing human data can be used in hazard classification for irritancy are provided in recent ECETOC publications (ECETOC, 2002; ECETOC, 2009).

Non-standardised human data on local skin effects

22. 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

[1] http://echa.europa.eu/web/guest/information-on-chemicals/cl-inventory-database, as of 2013-09-23

[2] http://echa.europa.eu/de/information-on-chemicals/registered-substances, as of 2013-09-23

[3] http://www.oecd.org/env/ehs/risk-assessment/theoecd(Q)SARtoolbox.htm, as of 2013-09-23

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, ICD) in humans.

23. 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 Cat. 1 (without sub-categorisation) should be used.

Module 1a – Existing human data: Non-standardised data on local skin effects	
Description / Definition	Existing human data on local skin effects originate from clinical and occupational studies, poison information centres, case reports and retrospective epidemiological studies, following single or repeated exposure (accidental or prolonged exposure in e.g., occupational settings).
Scientific basis incl. MoA	As obtained from humans, all MoA are potentially covered.
Applicability domain	All chemicals for which a clear and direct effect on the skin can be concluded from the available data, but not clearly defined as most data are obtained from accidental exposure.
Predictive capacity	Depends very much on the amount and quality of the available information, but usually associated with a high level of uncertainty due to lack of critical information such as chemical identity and purity, exposure (dose and duration), health status of the persons exposed and/or the reported symptoms.
Reliability	Difficult to assess due to uncontrolled exposures (dose and timings) and reporting.

Module 1a – Existing human data: Non-standardised data on local skin effects	
Strengths, weaknesses and limitations	Strengths: - Relevance: data obtained directly from the species of interest (humans). - May provide useful information on the cumulative effects leading to irritation (Irritant Contact Dermatitis, ICD) in humans. Weaknesses: - Not standardised. - Mostly based on accidental/uncontrolled exposure, often in combination with co-exposure, leading to a high level of uncertainty. - Sufficient data to evaluate the actual exposure (duration and dose) might not be available. - Data might be incomplete, insufficient or even inaccurate (Hoffman, 2007). - Data on the reversibility of the effect might not be available, because incidents are many times not followed-up after the initial assessment following the exposure. - Data on additional, potentially confounding factors (e.g., substance purity, health status of the affected person, additional exposures) might not be available. - No GHS criteria for C&L based on human data are available. - Usually not sufficient to sub-categorise chemicals according to their corrosion potency, e.g., UN GHS Sub-categories 1A, 1B and 1C. Limitations: - Differences in populations (Robinson, 2002). - Rarely available and, if available, rarely with the necessary quality to be used for C&L decisions.
Potential role in the IATA	Should be used in a WoE with other existing data, but should not overrule high quality data obtained with OECD OECD TGs for skin irritation and/or corrosion (OECD TGs 404, 430, 431, 435 or 439) unless the human data are of high and unquestionable quality. May be particularly relevant when human data demonstrate effects which cannot be observed in experimental animal studies.

Human Patch Test (HPT)

24. 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.

25. 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, 2013a) 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).

Module 1b – Existing human data: Human Patch Test (HPT)	
Description / Definition	Controlled study involving the exposure of small patches of skin of human volunteers to chemicals that are not sensitising and not acutely toxic via the dermal route. Various appropriate protocols exist, e.g. for testing skin tolerance to cosmetic ingredients or medical devices (Basketter, 1994, Walker et al, 1997, ECETOC, 2002). Protocols described single or repeated open, occlusive or semi-occlusive exposure for 4 up to 48 hrs. The example described in more details below is the HPT protocol developed by Basketter and co-workers in 1994, which applied chemicals to the skin of the upper outer arm of human volunteers for up to 4 hr. The number of panellists with skin irritation reactions was interpreted in comparison with concurrent controls, negative or positive and/or both, run with the same panel of volunteers. In studies that included a positive control, Sodium lauryl sulphate (SLS) at 20% aq. was often used, in order to take in to account the high human variability (Basketter *et al*, 1996). However, this is not an internationally agreed guideline for human patch testing and the details above are provided for information only and for evaluation of existing data and not as guidance on how to conduct prospective testing.
Scientific basis incl. MoA	As performed in humans and all possible effects (erythema, oedema, scabbing and bleaching) are evaluated, all MoA are covered.
Applicability domain	The HPT was developed for safety testing of cosmetic and household products and has been later adopted for testing of Medical Devices according to ISO 10993-10. However, in some instances and after careful ethical review the HPT has also been used for testing of chemicals. Only chemicals for which skin corrosion and other unacceptable toxicological hazards can be excluded can be tested (only chemicals producing no effects other than skin irritation). Dyes and other coloured chemicals may impair the scoring of effects, in particular erythema.
Predictive capacity	Since skin irritation responses are determined in human volunteers and compared to controls as appropriate; it can be assumed that HPT are highly predictive of effects in humans.
Reliability	If the HPT has been performed according to an appropriate protocol and evaluated by trained assessors the reliability should be at least meet the level of the animal test according to OECD 404. Nevertheless, there is evidence for ethnic/population differences (Robinson, 2002) that might not always be captured. Such variations can obviously not be captured either with the regulatory *in vivo* or *in vitro* tests.
Strengths, weaknesses and limitations	Strengths: - Relevance (highly predictive). - Usually, standardised, high quality data. Weaknesses: - Testing with human volunteers to obtain primary hazard data on skin corrosion/irritation for regulatory purposes is discouraged - Only retrospective data should be considered. Prospective testing not recommended for ethical reasons. - No GHS criteria for C&L based on human data are available. Limitations: - Differences in populations (Robinson, 2002). - Rarely available and mostly for chemicals with intended dermal contact e.g., cosmetic ingredients.

Module 1b – Existing human data: Human Patch Test (HPT)	
Potential role in the IATA	If a high-quality HPT result is already available, it should be considered as the strongest basis for C&L decision making (subject to the ethical considerations relevant for the respective regulatory programme). When contradictory HPT and animal (OECD TG 404) data are available and WoE analysis including all other existing data and (Q)SAR profiling is not conclusive towards one or the other result, confirmatory *in vitro* testing should be considered. For ethical reasons, HPT must not be included in a strategy as a prospective testing option.

Module 2 – In vivo skin irritation and corrosion data (OECD TG 404)

26. The OECD TG 404 (OECD, 2002) on Acute Dermal Irritation/Corrosion describes an in vivo test method performed on albino rabbits. It is based on a test developed by Draize for the assessment of systemic and local toxicity to skin and mucous membranes (Draize et al. 1944). OECD TG 404 has been revised twice: first in 1992 to include the possibilities to i) waive in vivo testing based on a positive in vitro skin corrosion test result and ii) use one animal in a first step of the in vivo procedure allowing certain chemicals to be exempted from further testing; second in 2002 to include a sequential testing and evaluation strategy as a supplement to the OECD TG).

Module 2 – *In vivo* skin irritation and corrosion data (OECD TG 404)	
Description / Definition	The OECD TG 404 measures the corrosive or inflammatory response produced in reaction to exposure to corrosive or irritant chemicals in albino rabbits. The test chemical is applied in a single dose to the skin and the degree of irritation/corrosion is observed and scored at specific intervals and is further described in order to provide a complete evaluation of the effects. The duration of the study should be sufficient to evaluate whether the effects observed are reversible or irreversible.
Scientific basis incl. MoA	OECD TG 404 measures the downstream effects of the inflammatory response produced in reaction to the tissue trauma/noxious stimuli induced by irritant chemicals. Such localised cell and tissue damage leads to release of inflammatory mediators, nerve stimulation, axonal reflexes, pain and itching (Welss *et al*., 2004; Kindt *et al*., 2006; Fluhr *et al*., 2008). The inflammatory response ultimately leads to observable phenomena such as localised skin swelling (oedema) and redness (erythema). These downstream events are visually observed and scored. The rabbit model has been established as rabbit skin is assumed to be more sensitive than human skin. This increased sensitivity may at least partly result from the fact that rabbit skin bears fur. Furthermore, it can be assumed that the MoAs leading to skin corrosion or irritation are comparable between rabbits and humans. Exposure of 4 hours adds to the increased sensitivity.
Applicability domain	A wide range of chemicals (substances and mixtures) can be tested according to OECD OECD TG 404. 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. Not applicable to the testing of gases and aerosols.
Predictive capacity	Test may be over-predictive (i.e. conservative) for irritation/corrosion in humans, i.e. effects are observed with the test, that would not occur in humans, for example due to clipping of fur, interspecies differences, etc. (Philips *et al*., 1972; York et. al., 1996; Robinson *et al*., 2001; Basketter *et al*., 2004; Hoffmann *et al*., 2008; Jírová *et al*., 2010). Available Human Patch Test data seem to confirm this (Jírová *et al*., 2010; Basketter *et al*., 2012; Ishii *et al*., 2013). However, the variability between humans is high (Basketter et al, 1996). Often a positive response in HPT was defined by comparison with an internal positive control, (e.g. positive reaction = more irritating than 20% SDS as a pragmatic decision). In these cases it was the selection of the positive control that defined the sensitivity of the HPT and its comparability with animal test data (Jirova et al, 2010; Basketter et al, 2012; Ishii et al, 2013).
Reliability	No studies assessing the intra- and inter-laboratory variability in a comprehensive way exist. Note that classification based on results between studies may significantly vary due to subjective scoring, dosing by weight (ignoring density differences), insufficiently standardised washing procedures, etc. Weil and Scala (1971) have shown that considerable variation existed between laboratories. As the protocol assessed differs substantially from the OECD TG 404, their results indicate potential sources of variability, but cannot be transferred to the protocol of the OECD TG 404. Hoffmann *et al*.'s (2005) systematic analysis indicates low within-test variability of the Draize test (variability between rabbits within a test) for the prediction of skin irritation, especially when considering a dichotomous system like Cat. 2 vs. No Cat. As probabilities of incorrect classification are largest around the classification borders, the use of only one threshold by discriminating Cat. 2 from No Cat., i.e. omitting Cat. 3, may be preferred. Indeed, the UN GHS text explicitly

Module 2 – *In vivo* skin irritation and corrosion data (OECD TG 404)	
	acknowledges that "…animal responses in a test may be variable" in the context of explaining the rationale for one single irritant category (Cat. 2) (UN, 2013; paragraph 3.2.2.1.2.2., sub point b). A second analysis looked at the possibility of reducing the number of rabbits tested for corrosion (Cat. 1 vs. not corrosive) or irritation (Cat. 2 vs. No Cat.) 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-categories1B 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 variation 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.

Module 2 – *In vivo* skin irritation and corrosion data (OECD TG 404)	
Strengths, weaknesses and limitations	Strengths: - Reversibility of effects can be observed. - Reflects all possible modes of action of skin irritant and corrosive reactions present in rabbit skin. - Classification of the full irritation and corrosion potency, i.e., No Cat., Cat. 3, Cat. 2, Sub-cat 1C, Sub-cat. 1B or Sub-cat. 1A, has been based on this test, so that it can provide classifications over the entire spectrum. Weaknesses: - Not formally validated. - Animal experiment, which may potentially involve suffering due to the corrosive or the inflammatory reactions (pain, itching, etc.). - Being performed in a proxy model (the rabbit) the test may make incorrect predictions due to species differences (e.g., Philips *et al*. 1972; Basketter *et al*. 2004). - Over-prediction (i.e. conservative outcome, worst case situation) of skin irritation/corrosion in humans (e.g., York *et al*., 1996; Robinson *et al*., 2001; Basketter *et al*.,2004, Jírová *et al*. 2010), possibly caused by, e.g.: a) Clipped tight fur promoting follicular penetration ('shunt pathway') that might be excessive as compared to the human situation. b) Clipping of the fur may cause minor invisible skin abrasions, facilitating the penetration via the abrasions. - Issues reducing reproducibility: a) Subjective scoring without use of positive or benchmark controls. b) Dosing solids per weight (0.5g/6 cm²) does not consider density differences. Solids should be dosed by bulk volume with a calibrated spoon. c) No standardised procedure described for removal of the test chemical: water wipe might be insufficient, no suitable solvents recommended. d) Difficulties to apply solids directly to the skin ensuring adequate retention. - Dyes and other coloured chemicals may impair the scoring of effects, especially erythema. - Not applicable to the testing of gases and aerosols. Limitations: - Subjective grading of skin responses.
Potential role in the IATA	In case Draize test data of adequate quality are available, these should carry a certain intrinsic weight in the context of a weight of evidence (WoE) analysis. Otherwise, the Draize test 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 Cat. 3 from 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 (see also Part 1 and Part 3).

Module 3 – In vitro skin corrosion data (OECD TGs 430, 431, 435)

OECD TG 430: In vitro skin corrosion: Transcutaneous Electrical Resistance test method (TER)

27. OECD TG 430 was first adopted on 13 April 2004 together with OECD TG 431 and was revised on 26 July 2013. The revision became necessary because the first version of OECD TG 430 did not define Performance Standards (PS) allowing the assessment of methodological modifications on the predictive performance (reliability and relevance) of the TER. Since the apparatus used in the validation studies is not commercially available, it was of particular importance to define Performance Standards for OECD TG 430.

Module 3a – *In vitro* skin corrosion data: OECD TG 430 (TER)	
Description / Definition	The test material is applied for up to 24 hours to the epidermal surfaces of rat skin discs in a two compartment test system in which the skin discs function as the separation between the compartments (OECD, 2013b). The skin discs are taken from humanely killed rats aged 28-30 days. Corrosive materials 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 transcutaneous electrical resistance below a threshold level. For rat TER, a cut-off value has been selected based on extensive data for a wide range of chemicals where the vast majority of values were either clearly well above, or well below this value. Generally, materials that are non-corrosive in animals but are irritant or non-irritant do not reduce the TER below this cut-off value (OECD, 2013b).
Scientific basis incl. MoA	- Rat skin used as a model of human skin due to comparable physiology. - TER measurement as readout of corrosive effects on the skin and its barrier (stratum corneum, SC), e.g., due to erosion of the SC.
Applicability domain	Discriminates skin corrosives (Cat. 1) from non-corrosives, but not accepted for distinguishing skin corrosive sub-categories 1A, 1B and 1C. OECD TG 430 is applicable to both substances and mixtures, although only limited information on the testing of mixtures is available. It is applicable to a wide range of chemical classes and physical states including liquids, semi-solids, solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. A small number of waxes and corrosive solids were however assessed during validation. Not applicable to the testing of gasses and aerosols (although this is true for almost all tests, including OECD TG 404).
Predictive capacity	When compared to the rabbit test classifications as Cat. 1 (corrosive) and NC (not corrosive), the TER was validated with a sensitivity of 88.1%, a specificity of 72.4% and an associated accuracy of 79.4%. Based on the predictive capacity obtained with the TER for the 24 Reference Chemicals mentioned in OECD TG 430, any future similar or modified TER test method must achieve a sensitivity ≥ 90%, a specificity ≥ 75% and an accuracy ≥ 82.5% when testing those 24 Reference Chemicals.
Reliability	For prediction of GHS Cat 1 vs. non-corrosive a within-laboratory reproducibility of ≥ 90% concordant classifications between runs and a between-laboratory reproducibility ≥ 80% concordant classifications between laboratories has been demonstrated in the validation studies and recommended as a minimum requirement for future TER test methods.

Module 3a – *In vitro* skin corrosion data: OECD TG 430 (TER)	
Strengths, weaknesses and limitations	Strengths: - Officially validated test method. - Based on a different mode of action (skin barrier breakdown) than RhE (OECD TG 431) and pH-based corrosion test (OECD TG 435), and thus may be valuable to complement evidence of results from these tests. It should be noted however that all three *in vitro* skin corrosion OECD TGs (430, 431 and 435) are considered stand-alone tests that permit the detection or exclusion of corrosive effects and classification of test chemicals for skin corrosion without further testing. Weaknesses: - May be considered an *in vivo* animal experiment in some countries due to the need to shave, wash and treat the animals with antibiotics during the 4-6 days before the animal is sacrificed for the test. - Animals are sacrificed for the purpose of testing. - The TER cut-off value for predicting skin corrosion varies with age and strain of the rats (see paragraph 15-17 of revised OECD TG 430; OECD, 2013b). It is also dependent on parameters of the apparatus, and it will have to be newly established if species other than rat are used (Davies *et al*., 2004). - Gases and aerosols have not been assessed yet in validation studies. While it is conceivable that these can be tested using the TER test method, the current OECD TG does not allow testing of gases and aerosols (although this is true for almost all tests, including OECD TG 404). Limitations: - No corrosive sub-categorisation possible. Only allows the classification of chemicals identified as corrosive as Cat. 1. - Does not discriminate skin irritants (Cat. 2) from from chemicals not requiring classification for skin irritation/corrosion (No Cat.), which are identified as non-corrosives in OECD TG 430. This differentiation should be addressed by module 4 (OECD TG 439).
Potential role in the IATA	The TER may be used as a stand-alone test method for the detection or exclusion of corrosive effects of test chemicals. If corrosive sub-categorisation is required other test methods should be considered. A negative result in the TER test method will require an additional *in vitro* skin irritation test, if not performed upfront, to determine if the chemical should be classified Cat. 2 (irritant) or if it does not require classification (No Cat.), and thus replace the *in vivo* test according to OECD TG 404.

OECD TG 431: In vitro skin corrosion: Reconstructed human epidermis (RhE) test method

28. OECD TG 431 In vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method was first adopted on 13 April 2004 together with the OECD TG 430 and revised on 26 July 2013. The original OECD TG comprised two validated RhE models (EpiSkinTM and EpiDermTM). The revision in 2013 became necessary because post validation studies performed by the RhE model producers in 2012 with a refined protocol correcting interferences of unspecific MTT reduction by the test chemicals improved the performance of both, discrimination of corrosives from non-corrosives as well as sub-categorisation of corrosives in UN GHS Sub-cat. 1A and Sub-cat. 1B-and-1C. In addition, two other RhE models (SkinEthicTM RHE and Epidermal Skin Test epiCS®) were included, as well as an annexed overview on methodological differences for each of the four validated and accepted RhE models.

Module 3b – *In vitro* skin corrosion data: OECD TG 431	
Description / Definition	OECD TG 431 is based on reconstructed human epidermis (RhE), which in its overall design (the use of human derived non-transformed epidermal keratinocytes as cell source and use of representative tissue and cytoarchitecture) closely mimics the biochemical and physiological properties of the upper parts of the human skin i.e., the epidermis. The RhE models are constructed by culturing the keratinocytes at the air-liquid interface to form a multi-layered, highly differentiated model of the human epidermis. It consists of organised basal, spinous and granular layers, and a multi-layered stratum corneum containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*. Test chemicals are applied topically to the three-dimensional RhE models, and exposed for 3 min and 1 hour in all RhE test methods and also for 4 hours in the EpiSkin™ test method. Cell viability is measured immediately following chemical exposure by dehydrogenase conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (OECD, 2013c). Corrosive chemicals are identified by their ability to decrease cell viability below defined threshold levels.
Scientific basis incl. MoA	The RhE test methods are based on the premise that corrosive chemicals are able to penetrate the stratum corneum by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. Cell viability is measured by the MTT assay immediately after exposure.
Applicability domain	Discriminates skin corrosives (Cat. 1) from non-corrosives. One test method (EpiSkin™) is accepted to distinguish corrosive 1A from a combination of Sub-cat. 1B and Sub-cat. 1C corrosives (Sub-cat. 1B-and-1C), while three other test methods (EpiDerm™ SCT, SkinEthic™ RhE and epiCS®) currently are accepted to identify only Sub-cat. 1B-and1C corrosives from not-further resolved corrosives (Cat. 1). Further work has been however conducted in some of the RhE models, such as the EpiDerm™ and the SkinEthic™ models to improve their capacity to discriminate Sub-cat 1A from Sub-cat 1B-and-1C from non-corrosives (Kandárová *et al*., 2013; Alépée *et al*., 2014a; Alépée *et al*., *2014b*).OECD TG 431 does not permit at present the use of any of the methods to distinguish Sub-cat. 1B from Sub-cat. 1C corrosives due to the limited set of well-known *in vivo* corrosive Sub-cat 1C chemicals. OECD TG 431 is applicable to both substances and mixtures, although only limited information on the testing of mixtures is available. It is applicable to a wide range of chemical classes and physical states including liquids, solids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. It is however not applicable to the testing of gases and aerosols (although this is true for almost all tests, including OECD TG 404).

Module 3b – *In vitro* skin corrosion data: OECD TG 431	
Predictive capacity	For the prediction of GHS Cat. 1 vs. not corrosive, in the full validation study and catch-up validation studies a sensitivity of > 95%, a specificity of > 70% and an accuracy of ≥ 82,5% was obtained and listed as a minimum requirement for future RhE models. For discrimination of Cat. 1A from Cat. 1B-and-1C from not corrosive chemicals similar to the EpiSkinTM test method, the following predictive capacity is recommended as a minimum requirement for future RhE models (OECD, 2013c): Sensitivity (C vs NC): ≥ 95% Correctly classified 1A: ≥ 80% 1A Under-classified 1B-and-1C: ≤ 20% 1A Under-classified NC: 0% Correctly classified 1B-and-1C: ≥ 80% 1B-and-1C Over-classified 1A: ≤ 20% 1B-and-1C Under-classified NC: ≤ 5% Specificity: ≥ 70% NC Over-classified 1A ≤ 5% NC Over-classified 1B-and-1C ≤ 30% Accuracy (C vs. NC): ≥ 87% Accuracy (1A vs. 1B-and-1C vs. NC): ≥ 78%
Reliability	For prediction of GHS Cat. 1 vs. not corrosive, a within-laboratory reproducibility of ≥ 90% concordant classifications between runs and a between-laboratory reproducibility ≥ 80% concordant classifications between laboratories have been demonstrated in the validation studies and recommended as a minimum requirement for future RhE models. For the discrimination of Cat. 1A, Cat. 1B-and-1C and not corrosive chemicals, a within-laboratory reproducibility of ≥ 80% and a between-laboratory reproducibility of ≥ 70% have been demonstrated in the validation studies and recommended as a minimum requirement for future RhE models.

Module 3b – *In vitro* skin corrosion data: OECD TG 431	
Strengths, weaknesses and limitations	Strengths: - Officially validated test method. - Human-based 3D tissue model. - Several equivalent models available. - Partial sub-categorisation possible (Cat. 1A versus Cat. 1B-and-1C). Weaknesses: - Test chemicals that act directly on MTT (e.g., MTT-reducer), those that are naturally coloured, or become coloured during tissue treatment need the use of adapted controls as described in the test methods SOPs. However, test results for materials inducing non-specific MTT reduction and non-specific colour ≥ 50% of negative control should be taken with caution. Use of HPLC and photometry to detect and quantify formazan in tissue extracts may reduce the limitations observed with coloured chemicals and chemicals that became coloured during tissue treatment, but this technique is not yet mentioned in the OECD TG and therefore not necessarily accepted by authorities. - Gases and aerosols have not been assessed yet in validation studies. While it is conceivable that these can be tested using RhE technology, the current OECD TG does not allow testing of gases and aerosols (although this is true for almost all tests, including OECD OECD TG 404). - Test methods included in OECD TG 431 are able to partially sub-categorise among corrosives (1A versus 1B/1C) with a variable rate of over-classification into cat.1A depending on the method (over-classification ranges from 21 to 46%). Limitations: - OECD TG 431 does not allow at present for discriminating between UN GHS skin corrosive Sub-categories 1B and 1C due to the limited set of well-known *in vivo* corrosive Sub-cat. 1C chemicals. - Does not discriminate skin irritants (Cat. 2) from chemicals not requiring classification for skin irritation/corrosion (No Cat.), which are identified as non-corrosives in OECD TG 430. This differentiation should be addressed by module 4 (OECD TG 439).
Potential role in the IATA	The RhE test methods may be used as a stand-alone test method for the detection or exclusion of corrosive effects of test chemicals. A negative result in these test methods will require an additional *in vitro* skin irritation test, if not performed upfront, to determine if the chemical should be classified Cat. 2 (irritant) or if it does not require classification (No Cat.), and thus replace the *in vivo* test according to OECD TG 404. OECD TG 431 also allows for the sub-categorisation of corrosive chemicals into Cat. 1A or Cat. 1B-and-1C but does not permit the distinction of the latter into Cat. 1B and Cat. 1C. It is important to note however that the protocol and prediction model of the EpiSkin™ test method permits sub-categorisation of corrosive chemicals into the three Categories 1A, 1B and 1C, but its ability to discriminate between Categories 1B and 1C was never formally evaluated/validated due to the lack of high quality reference *in vivo* data against which to benchmark the *in vitro* results (Fentem et al. 1998, Alépée et al. 2014a). This method may in some casesnevertheless be considered for this purpose before any *in vivo* testing is performed if the result 1B or1C is considered in a weight of evidence approach (see Modules 5a, below). If this is not possible a cautious default classification as 1B if OECD TG431 results in 1B/1C could be decided.

OECD TG 435: In vitro Membrane Barrier test method for skin corrosion

29. OECD TG 435 In vitro Membrane Barrier Test Method for Skin Corrosion was adopted on 19 July 2006 and was the third in vitro test method for skin corrosion. To allow the assessment of similar "me-too" test methods, OECD TG 435 was the first OECD OECD TG with annexed Performance Standards, since at present the test method is only available from one commercial supplier.

Module 3c – *In vitro* skin corrosion data: OECD TG 435	
Description Definition	The test system is composed of two components, a synthetic macromolecular bio-barrier and a chemical detection system composed of pH sensitive dyes; the basis of this test method is that it detects membrane barrier damage caused by corrosive test chemicals after the application of the test chemical to the surface of the artificial membrane barrier, presumably by the same mechanism(s) of corrosion that operate on living skin. Penetration of the membrane barrier (or breakthrough) may be measured by a number of procedures, including a change in the colour of a pH indicator dye or in some other property of the indicator solution below the barrier.
Scientific basis incl. MoA	Artificial membrane as surrogate for *in vivo* membrane barrier damage, presumably by the same mechanism(s) of corrosion that operate on living skin.
Applicability domain	Accepted to identify non-corrosives and skin corrosive subcategories 1A, 1B and 1C. Test method applicable to specific classes of chemicals, i.e., organic and inorganic acids, acid derivatives, and bases (NIH, 1999; ESAC, 2001). The *in vitro* membrane barrier test methods are applicable to substances and mixtures including pure chemicals, dilutions, formulations or waste. OECD TG 435 may be used to test solids (soluble or insoluble in water), liquids (aqueous or non-aqueous), and emulsions. It is however not applicable to the testing of gases and aerosols (although this is true for almost all tests, including OECD TG 404).
Predictive capacity	When compared to the rabbit test classifications as C (corrosive) and NC (not corrosive), the test was validated with a sensitivity of 86% (54/63), a specificity of 68% (15/22) and an accuracy of 81% (69/85) for acids, bases and acid derivatives under the UN GHS classification system (NIH, 1999). For sub-categorisation, the accuracy of the method is 96% using the 40 reference chemicals of OECD TG 431 (OECD, 2006).
Reliability	The test method showed acceptable within- and between-laboratory reproducibility in a validation study (Fentem *et al*., 1998). The between-laboratory reproducibility for corrosive versus non-corrosive and UN GHS skin corrosion sub-categories of any similar or modified membrane barrier test should be at least 93%. In terms of membrane breakthrough times, the median coefficient of variation (CV) should not exceed 30% for studies conducted in different laboratories and should not exceed 5% for replicate measurements within a study.

Module 3c – *In vitro* skin corrosion data: OECD TG 435	
Strengths, weaknesses and limitations	Strengths: - Officially validated test method. - Allows full sub-categorisation into Sub-cat 1A, 1B and 1C. - Simple test method. Weaknesses: - Usually not applicable to chemicals with 4.5 < pH < 8.5 because these are not detected by the chemical detection system used to detect passage of chemicals through the bio-barrier. In the EU Validation Study (Fentem *et al*., 1998), 58% of the test chemicals were not compatible with the Chemical Detection System (CDS). - Does not contain cellular constituents but reliably detects skin corrosion based on biochemical mechanisms. - In some cases colour changes might be transient and difficult to interpret; the colour obtained should be compared to photo diagrams provided with the test method that allows direct comparison. - Gases and aerosols have not been assessed yet in validation studies. Limitations: - Method considered valid for the limited applicability domain of acids, bases and acid derivatives (NIH, 1999; ESAC, 2001). - OECD TG 435 does not discriminate skin irritants (Cat. 2) from chemicals not requiring classification for skin irritation/corrosion (No Cat.), which is addressed by module 4 (OECD TG 439).
Potential role in the IATA	Considering that the RhE test methods can now also differentiate between sub-cat. 1A and Sub-cat. 1B-and-1C corrosives, the membrane barrier test may potentially be of particular value where discrimination between sub-categories 1B and 1C is required. It may also be particularly useful to sub-categorise corrosive chemicals identified on the basis of extreme pH (see Module 6 below). A negative result in the membrane barrier test method will require an additional *in vitro* skin irritation test, if not performed upfront, to determine if the chemical should be classified Cat. 2 (irritant) or if it does not require classification (No Cat.), and thus replace the *in vivo* test according to OECD TG 404.

Module 4 – In vitro skin irritation data (OECD TG 439)

30. OECD TG 439 on In vitro Skin Irritation: Reconstructed Human Epidermis (RhE) Test Method was first adopted on 22 July 2010 comprising three validated RhE models (EpiSkinTM, EpiDermTM and SkinEthicTM RHE). It constitutes the first in vitro test for skin irritation. A revised version was adopted on 26 July 2013, comprising a fourth validated RhE model (LabCyte EPI-MODEL24) as well as an annexed overview on methodological differences for each of the four validated and accepted RhE models. A further Annex of the OECD TG includes Performance Standards (PS). The updated Test Guideline will allow performance assessments of possible future RhE models used for the purpose of skin irritation testing and an easy update / revision of the current OECD TG 439.

Module 4 – *In vitro* skin irritation data: OECD TG 439	
Description / Definition	OECD TG 439 is based on RhE, which in its overall design (the use of human derived non-transformed epidermis keratinocytes as cell source and use of representative tissue and cytoarchitecture) closely mimics the biochemical and physiological properties of the upper parts of the human skin i.e., the epidermis. The RhE models are constructed by culturing the keratinocytes at the air-liquid interface to form a multi-layered, highly differentiated model of the human epidermis. It consists of organised basal, spinous and granular layers, and a multi-layered stratum corneum containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*. Test chemicals are applied topically to the three-dimensional RhE models, and exposed for 15 min to EpiSkin™ and LabCyte EPI-MODEL24, for 42 min to SkinEthic™ RHE and for 60 min to EpiDerm™. Cell viability is measured after a 42 hour post-treatment incubation period by dehydrogenase conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (OECD, 2013a). Irritant chemicals are identified by their ability to decrease tissue viability below 50% of the negative control.
Scientific basis incl. MoA	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. Stressed, damaged or dying keratinocytes release mediators that initiate an inflammatory reaction, 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 *in vivo*. The RhE-based test methods measure the initiating events in the cascade i.e., cell and tissue damage measured through decreased tissue viability *in vitro*. OECD TG 439 also addresses reversibility of the irritation effect by determining tissue viability 42 h after the end of exposure.
Applicability domain	Discriminates skin irritants (Cat. 2) from chemicals not classified for skin irritation (No Cat.). Not designed to classify chemicals to the optional GHS Cat. 3 (mild irritants). In the EU, where Cat. 3 has not been adopted and all Cat. 3 chemicals are considered not classified (No Cat.), the RhE-based test methods can be used as a skin irritation replacement test methods. However, a result indicating skin irritation (Cat. 2) does not allow excluding corrosion (Cat. 1), unless combined with results of other methods that discriminate corrosives from non-corrosives. Applicable to both substances and mixtures, although only limited information on the testing of mixtures is available. In particular, further investigations would be beneficial on agrochemicals due to the contradictory limited information reported and difficulty to interpret the data as the composition of the mixtures has not been identified (Eskes *et al*., 2012; Kolle *et al*., 2013). OECD TG 439 is applicable to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. It is however not applicable to the testing of gases and aerosols (although this is true for almost all tests, including OECD TG 404).
Predictive capacity	For the prediction of GHS Cat 2 vs. No Cat., in the full validation study and catch-up validation studies a sensitivity of ≥ 80%, a specificity of ≥ 70% and an accuracy of ≥ 75% was obtained and listed as minimum requirement for future RhE models.
Reliability	The test methods showed acceptable within- and between-laboratory reproducibility in full and catch-up validation studies, with within-laboratory reproducibility of ≥ 90% concordant classifications between runs and between-laboratory reproducibility of about ≥ 80% concordant classifications between laboratories.

Module 4 – *In vitro* skin irritation data: OECD TG 439	
Strengths, weaknesses and limitations	Strengths: - Officially validated test method. - Human-based 3D tissue model. - Several equivalent models available. - Accepted for identification of UN GHS classification Cat. 2 versus No Cat. Weaknesses: - Test chemicals that act directly on MTT (e.g., MTT-reducer), those that are naturally coloured, or become coloured during tissue treatment need the use of adapted controls as described in the test methods SOPs. Use of HPLC and photometry to detect and quantify formazan in tissue extracts may reduce the limitations observed with coloured chemicals and chemicals that became coloured during tissue treatment, but this technique is not yet mentioned in the OECD TG and therefore not necessarily accepted by authorities. - Gases and aerosols have not been assessed yet in validation studies. While it is conceivable that these can be tested using RhE technology, the current OECD TG does not allow testing of gases and aerosols (although this is true for almost all tests, including OECD TG 404). Limitations: - Not designed to classify chemicals to the optional GHS Cat. 3 (mild irritants). However, in countries not adopting this optional category, such as the EU, the RhE-based test methods can be used as a skin irritation replacement test methods. - OECD TG 439 does not provide adequate information on skin corrosion (Cat. 1), which is covered by the OECD TG described in module 3 (OECD TG 430, 431 and 435).
Potential role in the IATA	The RhE-based test methods are able to identify Cat. 2 and No Cat. chemicals and can thus serve as stand-alone skin irritation methods for non-corrosives in countries where optional Cat. 3 is not implemented. For authorities adopting Cat. 3, additional testing in an *in vitro* skin irritation test method not adopted by the OECD (see Module 5a below) or in the *in vivo* test method (see Module 2 above) may be required to resolve Cat. 3 from No Cat. In case RhE-based test methods result in Cat. 2, an *in vitro* skin corrosion test, if not performed upfront, is required to determine the final classification (Cat. 2 (irritant) or Cat. 1(A, B or C) (corrosive)).

Module 5 – Other in vivo and in vitro data

a) In vitro skin irritation or corrosion data from test methods not adopted by the OECD

31. Data from test methods not yet adopted by the OECD may also be considered in view of supporting WoE assessments. The relative weight of such data for integration within an WoE approach will depend on several factors, including

- the status of validation of the test methods used (if applicable),
- the quality and comprehensiveness of the available documentation on the test methods in peer-reviewed or other suitable publications allowing, for example, an appraisal of their predictive capacity, reproducibility, biological and mechanistic relevance etc.
- The quality and completeness of the available data generated by the test method in question.

32. Data from such methods may exist already or may be generated by prospective testing before conducting animal studies. Use of data from non-standard methods should be considered in cases where such methods are able to provide specific information on classification and labelling needs that may be required by some authorities and which cannot be generated currently by adopted (i.e. guideline) in vitro test methods.

33. Prospective testing with such methods may provide supportive information

- for discrimination between optional Sub-categories 1B and 1C for chemicals outside of the applicability domain of OECD TG 435,
- for discrimination of optional Cat. 3 from No Cat,
- if the test chemical cannot be tested with the in vitro test methods currently adopted by the OECD due to limitations or non-applicability.

34. Below is a short description of currently available in vitro methods capable of addressing at least one of the points listed above. This section should be updated as new test methods become available that are sufficiently well documented to be considered as non-standard information within this IATA.

35. Full sub-categorisation of corrosive chemicals: While OECD TG 431 can be used to sub-categorise corrosive chemicals into Sub-cat. 1A and a combination of Sub-cat. 1B and 1C (referred to as "1B-and-1C"), it can currently not be used to distinguish Sub-cat.1B from Sub-cat.1C chemicals. Nevertheless, although not validated due to lack of a sufficient number of Sub-cat. 1B and 1C test chemicals having high quality reference in vivo data against which to benchmark the in vitro results, the protocol and prediction model of the RhE EpiSkinTM test method permits discrimination of 1B from 1C Sub-categories (Fentem et al. 1998, Alépée et al. 2014a). Furthermore, it is conceivable and plausible that the protocols and prediction models of other RhE models could be adapted to provide reproducible predictions also for discrimination of 1B from 1C Sub-categories. Scientific evaluation of the capacity of such protocols will likely be hampered by the lack of relevant reference data (Fentem et al. 1998). This lack may indicate that the level of resolution requested by three subcategories may actually not be reliably provided by the reference test method itself. Nevertheless, it should be noted that the protocol and prediction model of the EpiSkin^{TM} test method, as evaluated in the ECVAM validation study (Fentem et al, 1998; Barratt et al., 1998) and as described in OECD TG 431, was originally developed for sub-categorisation of corrosive chemicals into the three Sub-categories 1A, 1B and 1C. The capacity to discriminate between Sub-categories 1B and 1C could however not be validated at the time due to the lack of high quality reference in vivo data against which to benchmark the in vitro results (Fentem et al., 1998). A recent study using a refined EpiSkin^{TM} protocol correcting interferences of unspecific MTT reduction by test chemicals showed results for all (sub-) categories relating to skin corrosion: Sub-cat. 1A vs. Sub-cat. 1B vs. Sub-cat. 1C vs. Not Corrosive (NC) as well as the reproducibility of the protocol (Alépée et al., 2014a). This method may in some cases be considered for discriminating Sub-Cat. 1B and 1C before any *in vivo* testing is performed if the result 1B or1C is considered in a weight of evidence approach (see Modules 5a, below). If this is not possible a cautious default classification as 1B if OECD TG431 results in 1B/1C could be decided.

36. Information on optional Category 3 (UN GHS) for classification of mild skin irritants: UN GHS foresees one category for irritant chemicals: Cat. 2 but allows the use of a further optional category (Cat. 3) to classify substances with intermediate irritancy potency ('mild irritants') with in vivo scores between 1.5 and 2.3. It is up to the regulatory authorities to decide whether or not they wish to implement this category. If not implemented, Cat. 3 chemicals are considered 'non irritants' (No Cat.). In the EU Cat. 3 has not been implemented (EC, 2008), while other regions may require information on optional Cat. 3. Currently,

alternative methods for skin irritation testing (OECD TG 439) provide information on Cat. 2 and No Cat., but cannot resolve Cat. 3 chemicals. There is however indication that novel protocols based on the measurement of parameters other than cell viability may be able to resolve Cat. 3 chemicals. For example, the IRR-IS assay, exploiting quantitative analysis of expression profiles of relevant genes appears to be a promising methodology to contribute to the determination of skin irritancy potential, i.e. 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 EpiSkinTM test system in response to chemical exposure (Groux et al, 2012). Before embarking on animal testing to generate information on Cat.3 chemicals to satisfy requirements of authorities implementing this category, the use and/or generation of data from non-standard methods able to provide such information should be considered.

b) Other in vivo and in vitro dermal toxicity data

37. Other in vivo or in vitro toxicity data of dermal exposure may provide additional information regarding the skin effects of a test chemical. Such data may be derived from one or more of the following OECD OECD TG's:

- OECD TG 402 Acute Dermal Toxicity (OECD, 1987)
- OECD TG 406 Skin Sensitisation (GPMT and Buehler Test) (OECD, 1992)
- OECD TG 410 Repeated Dose Dermal Toxicity Study, 21/28 days (OECD, 1981a)
- OECD TG 412 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)

38. 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.

39. In case acute local dermal toxicity data are available from the above mentioned studies, a number of considerations should be well thought-out when evaluating the data:

40. 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 cm² (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 b.w. (OECD TG 410 and OECD TG 412), or 2000 mg / kg b.w. (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.

41. In conclusion, although positive data obtained with a dilution of a test chemical in the above mentioned systemic studies, even with a species other than rabbit, may be used for a positive classification of an irritant potential, the authors of this document were unable to identify such cases. Positive data from range finder experiments for systemic studies or sensitisation studies may, however, be used in a weight of evidence approach (see Part 2). Negative results from other in vivo and/or in vitro dermal toxicity data can, however, not 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.

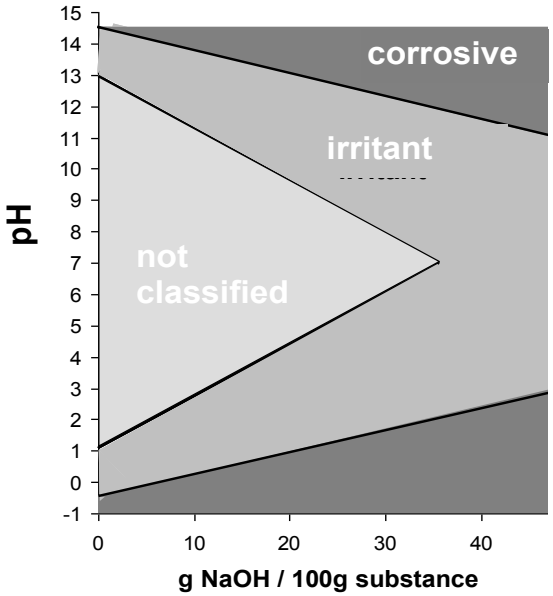
42. Finally, information obtained from skin penetration studies using OECD TGs 427 or 428 may provide evidence on the skin corrosion potential of a test chemical. Thus, both rapidly penetrating and cytotoxic chemicals, or clearly corrosive chemicals, may be assumed to be corrosive and classified as Cat. 1 if supported by other evidence in a WoE assessment. Data obtained with OECD TGs 427 or 428 may also be used to help orient chemicals to a top-down or bottom-up approach in Part 3 of the IATA.

Module 6 – Physico-chemical properties (existing or measured)

43. 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 (Cat. 1).

44. For chemicals with pH $\leq$ 2.0 or pH $\geq$ 11.5, skin corrosion could be expected. However, using an extreme pH for classification of a substance or a mixture as skin corrosive (Cat. 1) is a worst case assumption that should only be considered if no further data are available. As mentioned in OECD OECD TG 404, where extreme pH is the only basis of classification as corrosive, it may also be important to take into consideration the acid/alkaline reserve (a measure of the buffering capacity of a chemical), especially for classification of mixtures containing acidic or alkaline substances (Young et al., 1988) (Figure 2). However, it should be noted that for pure substances the sensitivity of pH for identifying skin corrosive may actually be significantly reduced when combined with acid/alkaline reserve information (Worth et al., 1998).

Figure 2: Relationship of pH, acid/alkaline reserve and classification of corrosive, irritant or not classified chemicals, according to UN GHS. Figure modified after Young et al. (1988).



45. The determination of pH should be performed following OECD TG 122 (2013f). This Test Guideline also describes procedures to determine acid reserve or alkali reserve for chemicals that are acidic (pH < 4) or alkaline (pH > 10) by titration with standard sodium hydroxide or sulphuric acid solution using electrometric endpoint detection.

46. However, the pH or pH in combination with buffering capacity should not be used alone to exonerate from classification as corrosive. Indeed, when the pH or pH in combination with acid/alkaline reserve, suggests that the chemical might not be corrosive, further in vitro testing should be considered.

Module 6 – Physico-chemical properties: pH	
Description / Definition	pH measurement (considering buffering capacity, if relevant).
Scientific basis incl. MoA	Chemicals exhibiting extreme pH (either pH ≤ 2.0 or pH ≥ 11.5), with high buffering capacity when relevant, are likely to produce visible necrosis of the skin.
Applicability domain	OECD guideline 122 describes the procedure to determine pH, acidity and alkalinity of aqueous solutions or aqueous dispersions in the range of 0 ≤ pH ≤ 14. Although OECD TG 122 allows pre-treatment with acetone to avoid plugging of the electrodes, it is apparent that some chemical properties, such as low water solubility or rapid hydrolysis, might impair pH measurements.
Predictive capacity	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). Worth *et al*. (1998) further reported that when the acid/alkaline reserve is also considered in combination with pH, the sensitivity is significantly decreased (29%; 7/24), with almost no change in specificity (92%; 11/12). The acid/alkaline reserve was however shown to have a positive impact with buffering mixtures (Young *et al*., 1988). It should also be noted that despite in low number, some chemicals with extreme pH did not show corrosive effects in native skin (false positives) (Worth *et al*., 1998).
Reliability	The studies were performed in single laboratories (Young *et al*., 1988; Worth *et al*., 1998). Therefore the reliability cannot be assessed.

Module 6 – Physico-chemical properties: pH	
Strengths, weaknesses and limitations	Strengths: - Simplicity. - Low cost. Weaknesses: - No information available on the test method reliability (reproducibility). - Detects skin corrosion induced by pH effects but not by other mechanisms. - Low sensitivity for identifying skin corrosion (high number of false negatives, i.e., there are several skin corrosives without an extreme pH). - An extreme pH may be considered in a WoE together with other data, but it shouldn't necessarily result in a classification to Cat. 1, since there are cases of chemicals with extreme pH that are not skin corrosives. Limitations: - No corrosive sub-categorisation possible. Only allows the classification of chemicals identified as corrosive as Cat. 1. - For extreme pH mixtures having low or no buffering capacity suggesting the mixture may not be corrosive despite the low or high pH value, the non-corrosive classification still needs to be confirmed by other data (preferably by data from an appropriate validated *in vitro* test).
Potential role in an IATA	Initial screen to identify skin corrosives based on extreme pH. Could be followed by an *in vitro* membrane barrier test (Module 3c) if sub-categorisation is required.

47. Other physico-chemical properties 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 health effects (Gerner et al., 2004). Such physico-chemical parameters may be measured or estimated using non-testing methods (see module 7), 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 (Figure 1).

Module 7 – Non-testing methods

48. Non-testing methods exist for both substances and mixtures. For mixtures, non-testing methods are described within the UN GHS chapter 3.2 Skin Corrosion/Irritation health hazards (UN, 2013), and can be divided into:

- Bridging principle, when data are not available for the complete mixture, and
- Theory of additivity, when data are available for the ingredients of the mixture.

49. For substances, non-testing methods can be divided into three different categories:

- Analogue approaches: Read-Across, SAR, and grouping (category formation)
- 'Classical' (Q)SARs, which quantitatively correlate activity to structure or structure-derived descriptors, and
- Expert and other prediction systems that often include several SARs, (Q)SARs, expert rules and/or data.

50. The above-mentioned 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. In addition, the adequacy and reliability of individual predictions is demonstrated by means of a (Q)SAR Prediction Reporting Format (QPRF) (see http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/(Q)SAR_tools/QRF).

51. With the introduction of the OECD (Q)SAR Toolbox4 in combination with the eChemPortal5, 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.

52. 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 website6 and is also provided in the OECD GD 69 (OECD, 2007a).

53. First the model should be described in accordance with OECD principles on (Q)SARs (OECD, 2004c), and documented by means of a QMRF. Interpretation of the model is additionally needed. For example a model based on the logarithm of the octanol/water partition coefficient (Kow) may indicate how the log Kow should be derived, measured, calculated, with which program, whether ionised substances can be used as well. For more complicated parameters e.g., the quantum descriptors HOMO (Highest Occupied Molecular Orbital energy) and LUMO (Lowest Unoccupied Molecular Orbital energy), this is even more crucial as the calculation outcome depends on the configuration state of the molecule. The performance parameters for the model (i.e., correlation coefficient, sensitivity/specificity, etc.) have to be reported. When the predictivity of a model is assessed, it should be assessed whether the test set is within the applicability domain of the model. The guidance given by the authors/builders of the model should be a starting point.

54. The second step is to evaluate the prediction for a specific substance. The OECD principles on (Q)SARs again apply. One of the most important principles is the substances' fit in the applicability domain (i.e., is the substance within the applicability domain of the model, and does information exist on the predictivity?). The outcome of the prediction should be assessed and documented in the form of a QPRF.

55. The third and last step of the evaluation explicitly needs to meet regulatory requirements. In this last evaluation the (Q)SAR prediction is weighed against the possible mechanism of skin irritation and corrosion. It has to be compared with the effects that can be observed in the in vivo test, to evaluate whether all skin irritation/corrosion pathways are covered. In this last step, the hazard of defatting properties has to be assessed as well. (Q)SAR models have to be evaluated considering the possible mechanism and how this would relate to GHS hazard classification.

[4] http://www.oecd.org/env/ehs/risk-assessment/theoecd(Q)SARtoolbox.htm, as of 2013-09-23

[5] http://www.echemportal.org, as of 2013-09-23

[6] http://www.oecd.org/env/ehs/risk-assessment/guidancedocumentsandreportsrelatedto(Q)SARs.htm, as of 2013-09-23

Module 7 – Non-testing methods	
Description / Definition	**Substances:** - Analogue approaches (read-across, SARs, and grouping). - (Q)SARs. - Expert and other prediction systems that often include several (Q)SARs, expert rules and data. **Mixtures:** - Bridging principles - Theory of additivity
Scientific basis incl. MoA	**Substances:** Mainly correlative approaches based on the general assumption that substances with comparable structural properties have comparable skin corrosion and irritation properties. However this might change once the Adverse Outcome Pathway (AOP) project (OECD, 2013g) has made further progress or more (Q)SARs might become available built on mechanistically based high-throughput *in vitro* data. **Mixtures:** Bridging principles are used when 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. The theory of additivity is used when data are available on 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, 2013).
Applicability domain	**Substances:** Model-specific and needs to be defined in a QMRF. Also QPRF are used to describe whether a prediction for a specific substance should be regarded as within the Applicability Domain or not. Application of these non- testing approaches is rather straight-forward for mono-constituent 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. For UVCB substances, by definition, not all of the constituents are known with respect to their identity and/or their relative concentrations. QSAR models and grouping approaches have, however, been employed on multi-constituent substances and UVCBs 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. **Mixtures:** The bridging principle is applicable to mixtures having data on both their individual ingredients and similar tested mixtures. The theory of additivity is applicable to mixtures that have data available for all or for some ingredients.

Predictive capacity	**Substances:** Model-, domain- and context-specific. **Mixtures:** Only limited data available. An impact assessment carried out by 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 many products currently not requiring classification according to consistent animal, *in vitro* and human experience data.
Reliability	Not applicable.
Strengths, weaknesses and limitations	Strengths: Substances and mixtures: - Ease of application. - Low cost. Weaknesses: Substances: • Results may be less relevant compared to experimental data, depending on the substance as well as the non-testing method and its underlying (model development / validation) data set. Limitations: Substances: - Applicability limited to the applicability domain of the model. Mixtures: - For extreme pH mixtures having low or no buffering capacity suggesting the mixture may not be corrosive despite the low or high pH value, the non-corrosive classification still needs to be confirmed by other data (preferably by data from an appropriate validated *in vitro* test).
Potential role in an IATA	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 (Module 5b) and to support skin corrosion or irritation *in vitro* results (Modules 3, 4 and/or 5a). If further testing is required, information generated with this Module may be used for deciding on how to address Part 3 i.e., initiate a top-down or a bottom-up approach (Figure 1).

Bridging approaches and theory of additivity (mixtures)

56. Bridging principles are used when 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, 2013).

57. 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, 2013).

Analogue approaches (substances)

58. Read-across, SARs and Grouping/Category formation are treated together because they all represent approaches based on the same basic concept. 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).

59. 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, as indicated in the OECD testing strategy for skin irritation/corrosion (OECD, 2002). Negative data from structural analogues may also be used to make predictions in certain cases, provided that there are no other substructures in the substance that are considered likely to cause the effect.

60. A variety of SARs for predicting the presence of irritation or corrosion have been described by Hulzebos et al. (2001, 2003, 2005a), and others have been incorporated into the BfR rulebase and the SICRET tool (Walker et al., 2005). These alerts have later been incorporated into the Toxtree software as well as into the OECD (Q)SAR Toolbox.

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

61. 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.

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

63. In the case of classification models for skin corrosion, where it is not indicated whether the predicted classification should be 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 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 substances.

Table 2: Overview of available (Q)SARs for skin irritation/corrosion. Note that this list is likely to be non-exhaustive and does not imply endorsement by OECD of any of the listed models for a particular prediction (internet links accessed in Dec. 2013).

Source	**Chemical domain**
Literature Models	
Barratt (*et al*.), 1995, 1996a, b, c, Whittle *et al*. 1996	Diverse local models for acids, bases, phenols, neutral organics, electrophiles
Golla *et al*., 2009	Organic chemicals from diverse classes
Hayashi *et al*., 1999	Phenols
Kodithala *et al*., 2002	Phenols, ethers, and alcohols
Nangia *et al*., 1996	Basic compounds
Smith *et al*., 2000 a, b	Esters
Data repositories for pre-calculated (Q)SAR predictions - Free	
Danish QSAR database (http://qsar.food.dtu.dk/) Also available as part of the OECD QSAR Toolbox (http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm)	Industrial chemicals, pesticides etc.
Computerised Models – Free	
BfR rule base (Gerner *et al*., 2004, 2007a,b; Hulzebos *et al*., 2005a; Rorije and Hulzebos, 2005; Walker *et al*., 2004, 2005; Gallegos et al., 2007), as part of: OECD QSAR Toolbox (http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm) Toxmatch, Toxtree, ToxPredict, and Ambit (http://www.ideaconsult.net/products)	EU New Chemicals (NONS) database, organic chemicals with no significant hydrolysis potential and purity > 95 %
PaDEL-DDPredictor (Liew and Yap, 2013) (http://padel.nus.edu.sg/software/padelddpredictor/)	
Computerised Models - Commercial	
ACD/Percepta (http://www.acdlabs.com/products/percepta/)	Organic chemicals
Derek Nexus (http://www.lhasalimited.org/products/derek-nexus.htm)	Organic chemicals and some metals
HazardExpert (http://www.compudrug.com/hazardexpertpro)	Organic chemicals
Molcode (http://reachqsar.com/)	Organic chemicals
MultiCASE (http://www.multicase.com/products/products.htm)	Organic chemicals
TopKat (http://accelrys.com/solutions/scientific-need/predictive-toxicology.html)	Organic chemicals
Review papers	
Gallegos Saliner *et al* 2006, 2008	N.A.
Hulzebos *et al*. 2001, 2003, 2005b	N.A.
Mombelli 2008	N.A.

Source	Chemical domain
Patlewicz *et al*. 2003	N.A.

N.A. – Not Applicable. A detailed description of the above models is given in Appendix R.7.2-2 of the ECHA IR/CSA guidance 7a (ECHA, 2013).

B. Part 2: Weight of Evidence Analysis

Module 8 – Phases and elements of weight of evidence approaches

64. A weight of evidence (WoE) determination means that all available and scientifically justified information bearing on the determination of hazard is considered together. In case of skin corrosion/irritation this includes structural information, information on physico-chemical parameters (e.g., pH, acid/alkaline reserve), information from category approaches (e.g., grouping, read-across), (Q)SAR results, the results of suitable in vitro tests, relevant animal data, skin irritation information/data on other similar chemicals, and human data. The quality and consistency of the data should be taken into account when weighing each piece of available information. 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. However, normally positive results that are adequate for classification should not be overruled by negative findings.

65. A 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.

66. 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).

67. Under the GHS (UN, 2013), in sub-chapter 3.2.2.2 a weight of evidence approach is recommended. All available information that can contribute to the determination of classification for an endpoint is considered together.

68. In the following paragraphs a suggestion of the steps and elements of WoE is given.

Place/role of WoE in the IATA

69. WoE should be carried out before any additional in vitro or in vivo testing is performed. Physico-chemical information, (Q)SAR, read-across, grouping information and/or existing in vivo, in vitro and/or human data might be considered sufficient to conclude on skin corrosion and irritation.

Coverage of relevant sources of information

70. The IATA specifies several types of existing information that can be used, provided these are of sufficient quality. Structural information, physico-chemical properties, data on structurally-related chemicals obtained by read-across or grouping approaches, (Q)SAR modelling data, existing human data and data from acute or sub-acute dermal toxicity studies in laboratory animals as well as in vitro data are listed. In the WoE analysis, the availability of specified types of data should be checked. The sources of those data obviously vary, ranging from clinical study reports, scientific publications, data from poison information centres, guideline tests, up to worker surveillance data of the chemical companies.

Assessment of data quality

71. The quality of the data that is obtained for a WoE needs to be assessed, since the quality will contribute to the value/weight of each data element. 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. 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.

72. The quality of toxicological studies is usually described by assigning Klimisch scores. The process of score assignment was originally described by Klimisch et al. (1997). In order to reduce the subjectivity and to increase the transparency in Klimisch score assignment, Schneider et al. (2009) proposed a scoring tool in form of a questionnaire, called the 'ToxRTool'7, as a convenient means of summarising and assessing study quality based on the Klimisch system. Epidemiological data can be evaluated using Hill's criteria (Hill, 1965).

73. The quality of the study, the method, the reporting of the results, and the conclusions that are drawn, must be evaluated carefully. Reasons why existing study data may vary in quality include the use of outdated test guidelines, the failure to characterise the test chemical properly (in terms of purity, physical characteristics, etc.) and the use of crude techniques/procedures that have since become refined, Moreover, other reasons could be poor reporting of information and poor quality assurance.

74. For many existing chemicals, at least some of the available information could have been generated prior to the requirements of Good Laboratory Practice (GLP) and the standardisation of testing methods. While such information may still be usable, both the data and the methodology used must be evaluated in order to determine their reliability. Such an evaluation would ideally require an evidence-based evaluation i.e., a systematic and consistent evaluation following pre-defined, transparent and independently reviewed criteria before making decisions. These should always include justifications for the use of particular data sets on the basis of the criteria-based evaluation. For some chemicals, information may be available from tests conducted according to OECD Test Guidelines (or other standards like CEN, ISO, ASTM, OSPAR methods, national standard methods), and in compliance with the principles of GLP or equivalent standards.

[7] http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/archive-publications/toxrtool, as of 2013-09-23

Adequacy and relevance of information

75. 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. 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. 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 judgement is an important principle in considering the adequacy of information and determining the key study.

Non-testing data

(Q)SAR data

76. It is important to distinguish between the proposed validity of the (Q)SAR model per se, and the reliability and adequacy of an individual (Q)SAR estimate (i.e., the application of the (Q)SAR model to a specific substance), and the appropriateness of the documentation (e.g., QMRF) associated with models and their predictions.

77. Guidance on how to characterise (Q)SARs according to the OECD (Q)SAR validation principles is provided in the OECD GD 69 (OECD, 2007a).

78. The information in the QMRF and QPRF should be used when assessing whether a prediction is adequate for the purpose of classification and labelling and/or risk assessment. The assessment will also need to take into account the regulatory context. This means that the assessments of (Q)SAR validity (typically proposed in scientific publications) and (Q)SAR estimate reliability need to be supplemented with an assessment of the relevance of the prediction for the regulatory purposes, which includes an assessment of completeness, i.e., whether the information is sufficient to make the regulatory decision, and if not, what additional (experimental) information is needed. The decision will be taken on a case-by-case basis.

79. (Q)SAR predictions may be gathered from databases (in which the predictions have already been generated and documented) or generated de novo through the available models.

Data obtained by grouping approaches

80. Conclusions about the likely properties of a substance can also be based on the knowledge of the properties of one or more similar chemicals, by applying grouping methods.

81. The corresponding OECD guidance provides information on the use of grouping of chemicals and read-across approaches (OECD, 2007b – currently being updated)

82. As with (Q)SARs, grouping approaches can be used to indicate either the presence or the absence of an effect.

Existing human data

83. The strength of the epidemiological evidence for specific health effects depends, among other things, on the type of analyses and on the magnitude and specificity of the response. Human data other than epidemiological studies can come from e.g., case reports, clinical studies, occupational disease

registries or other occupational surveillance schemes and from poison centre information. In principle all types of toxic effects can be reported in such studies; however, in many cases they address acute and/or local effects. Confidence in the findings is increased when comparable results are obtained in several independent studies on populations exposed to the same agent under different conditions. Other characteristics that support a causal association are presence of a dose-response association, a consistent relationship in time and (biological) plausibility, i.e., aspects covered by epidemiological criteria such as those of Hill (1965).

84. A comprehensive guidance of both the evaluation and use of epidemiological evidence for risk assessment purposes is provided by Kryzanowski et al. (WHO, 2000).

85. High quality human data may also be obtained from historical HPT studies (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). High quality HPT data may be considered as one of the strongest basis for C&L decision making (subject to the ethical considerations relevant for the respective regulatory programme). However, when contradictory HPT and animal (OECD TG 404) data are available and WoE analysis including all other existing data and (Q)SAR profiling is not conclusive towards one or the other result, confirmatory in vitro testing should be performed.

86. It is emphasised that testing with human volunteers is strongly discouraged for ethical reasons, but when there are good quality data already available they may be used as appropriate, in well justified cases.

Evaluation of consistency of the data

87. The consistency of the existing data from various sources is crucial and should therefore be thoroughly evaluated in WoE. 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 an 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.

Assessment of the coverage of relevant parameters and observations

88. 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. Therefore, an important element of WoE is to consider to what extent the parameters and observations were addressed by each data element of the WoE.

Conclusions of WoE

89. In the final analysis of the WoE, 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. 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. 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.

90. The WoE assessment needs to be transparently explained and documented to enable a logical flow leading to the decision/conclusion. An example for a simple approach to the documentation of the WoE is presented in Annex II.

C. Part 3: Additional Testing

91. In case the existing information and the WoE does not allow for an unequivocal decision regarding the skin corrosion and/or irritation potential/potency of the chemical, the generation of additional non-testing data (i.e., (Q)SAR and read across for substances as well as bridging principles and additivity approach for mixtures), or relevant physico-chemical data should be considered. If data from several (Q)SAR models on a substance are already available and are known to disagree, it may not be helpful to generate other (Q)SAR predictions but to carefully consider how well the prediction from each (Q)SAR model can be concluded to be within the applicability domain of that model. If however no (Q)SAR analysis has been performed, the generation of (Q)SAR information might just be sufficient to supplement the existing data and come to a conclusion on C&L. If the WoE considering the additional physico-chemical and non-testing data is still inconclusive, other in vivo or in vitro dermal toxicity tests (Module 5b) for which data are not yet available but that may need to be conducted in some regulatory frameworks to satisfy other regulatory requirements, should be carried out first. Once available, these additional test results should be incorporated into a new WoE analysis. If the WoE is however still inconclusive or no other in vivo or in vitro dermal toxicity tests need to be conducted, additional testing will be required (Part 3 of the IATA). 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 (Figure 1).

92. Testing options include adopted in vitro skin corrosion test methods (Module 3: OECD TGs 430, 431 and 435), adopted in vitro skin irritation test methods (Module 4: OECD TG 439) and in vitro skin irritation or corrosion test methods not adopted by the OECD (Module 5a). It is generally acknowledged that when limitations and domain of the in vitro tests adopted by OECD 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 vivo testing may be considered only 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 properties of these tests have been described in the respective Modules above. 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 test method to be used.

93. The top-down approach (start with an in vitro skin corrosion test followed by an in vitro skin irritation test in case the chemical is identified as not being corrosive in the first test) should be used when all available collected information and the WoE assessment result in a high a-priori probability of the chemical being an irritant or a corrosive. The bottom-up approach, on the other hand (start with an in vitro skin irritation test followed by an in vitro skin corrosion test in case the chemical is identified as being irritant in the first test) should be followed only when all available collected information and the WoE

assessment result in a high a-priori probability of the chemical being not an irritant to skin (Figure 1). This approach is recommended due to the difference in exposure times between the in vitro RhE-based skin irritation tests and the in vitro RhE-based skin corrosion tests. While the former has exposure times varying from 15 min to 1 hour (see Module 4 above) and a unified classification cut-off at 50% tissue viability, the latter have maximum exposure times of 1 to 4 hours (see Module 3b above) and classification cut-offs for these maximum exposures at 15% or 35% tissue viability. Based on these characteristics, it cannot be excluded that in some situations a skin corrosive chemical is correctly identified as corrosive in the in vitro RhE-based skin corrosion test methods but identified as being non-irritant in the in vitro RhE-based skin irritation test methods. It is plausible that the probability for this situation to occur increases as the exposure time in the in vitro RhE-based skin irritation test methods decreases. However, if existing information and WoE assessment point to the chemical being non-irritant, it should be safe to start a bottom-up approach without risking to identify a corrosive chemical as non-irritant.

Assessment of mixtures

94. Mixtures are defined as "a mixture or a solution composed of two or more substances in which they do not react" (UN, 2013). Whereas mixtures cover a wide spectrum of categories and composition, the type of regulatory testing required may depend on the type of mixture. For example, cosmetic formulations can no longer be tested using animal studies in some geographical regions (EC, 2009). In contrast biocidal products including mixtures may be subject to specific testing requirements (e.g., EU, 2012). As such, depending on the field and/or sector, the use of validated in vitro assays to assess mixtures is of relevance. Examples where in vitro testing of preparations and/or mixtures could be useful and/or relevant include cosmetics, detergents and cleaning products, biocides, and plant protection products. Furthermore, in cases where testing is not required and information on the irritating or corrosive properties of the mixture is only required for classification and labelling, the bridging principles or the theory of additivity based on the percentage and skin irritation/corrosivity properties of all constituents in the mixture should be applied (sections 3.2.3.2. and 3.2.3.3. UN GHS, 2013).

95. Most of the currently adopted in vitro test methods for skin irritation or corrosion (i.e., OECD OECD TGs 430, 431 and 439) have undergone scientific validation studies that, although covering a wide range of chemical classes and physical states, were conducted mainly based on substances (Fentem et al., 1998; Liebsch et al., 2000; Kandárová et al., 2006; Tornier et al., 2010; Spielmann et al., 2007; Eskes et al., 2007; Kandárová et al., 2009; Kojima et al., 2012; Kojima et al., 2013). The only exception is OECD TG 435, for which a number of mixtures (n=152) were reported to be tested (NIH, 1999).

96. Moreover, 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 (Eskes et al., 2012; Kolle et al., 2013). The applicability of the test methods within OECD TGs 430, 431 and 439 for the assessment of mixtures may depend on: i) the types and categories of products tested, ii) the endpoint(s) assessed (corrosion versus irritation), and iii) the adopted in vitro test method protocol used. It is therefore not possible to generalise the applicability of the currently adopted in vitro skin irritation or corrosion OECD TGs based on the types of mixtures assessed. Furthermore, it is not possible to define generalised criteria on the amount of evidence needed to demonstrate the applicability of an adopted in vitro assay to test mixtures, as it may depend on the availability of in vivo (animal and/or human) data, as well as on the variety, category and types of mixture evaluated.

97. 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. Further investigations would be beneficial in particular on the in vitro skin irritation testing of

agrochemicals due to contradictory limited information reported (Eskes et al., 2012; Kolle et al., 2013). In cases where evidence can be demonstrated on the non-applicability of the Test Guideline to a specific category of mixtures (e.g., following a strategy as described by Eskes and co-workers, 2012), 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.).

IV. REFERENCES

Alépée N., Grandidier M.H., 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. Toxicol In Vitro 28, 131-145.

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

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 ((Q)SARs) 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 (Q)SAR 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., Griffiths HA, Wang XM, Wilhelm KP, McFadden J. (1996). Individual, ethnic and seasonal variability in irritant susceptibility of skin: the implications for a predictive human patch test. Contact Dermatitis; 35(4): 208-13.

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.

Davies D.J., Ward R.J., Heylings J.R. (2004). Multi-species assessment of electrical resistance as a skin integrity marker for in vitro percutaneous absorption studies. Toxicol. In Vitro 18, 351-358.

Draize J.H., Woodard G., Clavery H.O. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J. Pharmacol. Exp. Ther. 82, 377–390.

EC (2008). 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 (2009). 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.

ECETOC (2002). Mono 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 (2013). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7a: Endpoint specific guidance. Version 2.2. August 2013 [http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf].

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

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., Depallens O. (2012). Regulatory Assessment of In vitro Skin Corrosion & Irritation Data within the European Framework: Workshop Recommendations. Regulatory Toxicology & Pharmacology 62, 393-403.

EU (2012). Regulation (EU) N 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. Official Journal of the European Union L 167, 1-123.

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

Fluhr J.W., Darlenski R., Angelova-Fischer I., Tsankov N., Basketter D. (2008). Skin irritation and sensitization: mechanisms and new approaches for risk assessment. 1. Skin irritation. Skin Pharmacol Physiol. 21,124-135.

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. Available at : [http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/doc/QSAR_Review_Irritation.pdf].

Gallegos Saliner A., Patlewicz G., Worth A.P. (2008). A review of (Q)SAR models for skin and eye irritation and corrosion. (Q)SAR & Combinatorial Science 27, 49-59.

Gallegos Saliner A., Tsakovska I., Pavan M., Patlewicz G., Worth A.P. (2007). Evaluation of SARs for the prediction of skin irritation/corrosion potential - structural inclusion rules in the BfR decision support system. SAR and (Q)SAR in Environmental Research 18, 331-342.

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. (Q)SAR Comb. Sci. 23, 726-733.

Gerner I., Hulzebos E., Rorije E., Hakkert B., Walker J.D., Herzler M., Spielmann H.(2007a). Potential regulatory use of (Q)SARs to develop dermal irritation and corrosion assessment strategies. In: Dermal Absorption and Toxicity Assessment, edited by M. S. Roberts and K. A. Walters. Informa Healthcare, New York, London. p. 495-506.

Gerner I., Hulzebos E., Rorije E., Herzler M., Liebsch M., Walker J.D., Spielmann H. (2007b). Development of (Q)SARs for dermal irritation and corrosion assessment using European Union New Chemicals notification data. In: Dermal Absorption and Toxicity Assessment, edited by M. S. Roberts and K. A. Walters, Informa Healthcare, New York, London, p. 507-522.

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: [http://www.toxicology.org/AI/Pub/Tox/2012Tox.pdf].

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.

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.

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

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

Hoffmann S., Cole T., Hartung T. (2005). Skin irritation: prevalence, variability, and regulatory classification of existing in vivo data from industrial chemicals. Regulatory Toxicology and Pharmacology 41, 159-166.

Hoffmann S., Saliner A.G., Patlewicz G., Eskes C., Zuang V., Worth A.P. (2008). A feasibility study developing an integrated testing strategy assessing skin irritation potential of chemicals. Toxicology Letters 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.

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. Available at: [http://rivm.openrepository.com/rivm/bitstream/10029/9562/1/601516008.pdf].

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

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

Hulzebos, E., Sijm, D., Traas, T., Posthumus, R., Maslankiewicz, L. (2005b). Validity and validation of expert (Q)SAR systems. SAR and (Q)SAR in Environmental Research 16, 385-401.

Hulzebos E., 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.

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., Kearney P., Sheasgreen J. (2009). In vitro Skin Irritation Testing: Improving the Sensitivity of the EpiDerm Skin Irritation Test Protocol. ATLA 37, 671–689.

Kandárová H., Letasiova S., Milasova T., Hayden P., Klausner M. (2013). Analysis of the Validated Epiderm Skin Corrosion Test (EpiDerm SCT) and a Prediction Model for Sub-Categorization According to the UN GHS and EU CLP. ALTEX Proceedings 2, LINZ 2013, p. 47. Available at: [http://www.altex.ch/resources/linz_2013_entire_issue.pdf].

Kindt T.J., Osborne B.A., Goldsby R.A. (2006). Immunology, 6th edition. Freeman and Company, New York.

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 (Q)SAR 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., Omori T. (2012). Validation Study of the In vitro Skin Irritation Test with the LabCyte EPI-MODEL24. Altern. Lab. Anim. 40, 33-50.

Kojima H., Katoh M., Shinoda S., Hagiwara S., Suzuki T., Izumi R., Yamaguchi Y., Nakamura M., Kasahawa T., Shibai A. (2013). A catch-up validation study of an in vitro skin irritation test method using reconstructed human epidermis LabCyte EPI-MODEL24. J. Appl. Tox., in press.

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.

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

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

Mombelli E. (2008). An evaluation of the predictive ability of the (Q)SAR software packages, DEREK, HAZARDEXPERT and TOPKAT, to describe chemically-induced skin irritation. ATLA 36, 15-24.

NAFTA (North American Free Trade Agreement) (2012). Technical Working Group on Pesticides - (Quantitative) Structure Activity Relationship ((Q)SAR) Guidance Document. 186pp. Available at: [http://www.epa.gov/oppfead1/international/naftatwg/guidance/qsar-guidance.pdf].

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. The results of an independent peer review evaluation coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No. 99-4495. Research Triangle Park, NC, USA: NIEHS. Available at: [http://iccvam.niehs.nih.gov/docs/dermal_docs/corprrep.pdf].

OECD (1981a). Repeated Dose Dermal Toxicity: 21/28-day Study. OECD Guideline for the Testing of Chemicals No. 410, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (1981b). Subchronic Dermal Toxicity: 90-day Study. OECD Guideline for the Testing of Chemicals No. 412, 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. Available at: [http://www.oecd.org/env/testguidelines].

OECD (1992). OECD Guidelines for the Testing of Chemicals No. 406: Skin Sensitisation. OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2002). Acute dermal irritation/corrosion. OECD Guideline for the Testing of Chemicals No. 404, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2004a). Guidelines for the Testing of Chemicals 427. Skin absorption: in vivo method. OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2004b). Guidelines for the Testing of Chemicals 428. Skin absorption: in vitro method. OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2004c). Report from the Expert Group on (Quantitative) Structure Activity Relationships ((Q)SARs) on the Principles for the Validation of (Q)SARs. OECD Series on Testing and Assessment No. 49, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, OECD Series on Testing and Assessment No. 34, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2006). In vitro membrane barrier test method. OECD Guideline for the Testing of Chemicals No. 435, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

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. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2007b). Guidance on Grouping of Chemicals. OECD Series on Testing and Assessment No. 80, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2010). Explanatory Background Document to the OECD Test Guideline on In Vitro Skin Irritation Testing OECD Series on Testing and Assessment No. 137, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2010b).OECD Guidelines for Chemical Testing No. 429. Skin sensitization: Local Lymph Node assay. OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2010c). OECD Guidelines for Chemical Testing No. 442A. Skin sensitization: Local Lymph Node assay: DA. OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2010d). OECD Guidelines for Chemical Testing No. 442B. Skin sensitization: Local Lymph Node assay: BrdU-ELISA. OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013a). In vitro skin irritation: Reconstructed Human Epidermis Test Method. OECD Guideline for the Testing of Chemicals No. 439, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013b). In vitro skin corrosion: Transcutaneous Electrical Resistance (TER). OECD Guideline for the Testing of Chemicals No. 430, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013c). In vitro skin corrosion: reconstructed human epidermis (RHE) test method. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013d). Streamlined Summary Document supporting the Test Guideline 437 for eye irritation/corrosion. Series on Testing and Assessment, No.189. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013e). Streamlined Summary Document Supporting OECD Test Guideline 438 on the Isolated Chicken Eye for Eye Irritation/Corrosion. Series on Testing and Assessment no. 188 (Part 1 and Part 2), OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013f). Determination of pH, Acidity and Alkalinity. OECD Guideline for the Testing of Chemicals No. 122, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines].

OECD (2013g). Guidance Document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment No. 184. Accessible at: [http://www.oecd.org/env/testguidelines].

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

Phillips L., Steinberg M., Maibach H.I., Akers W.A. (1972) A comparison of rabbit and human skin response to certain irritants. Toxicol. Appl. Pharmacol. 21, 369–382.

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., 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., 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., Hulzebos E.M. (2005). Evaluation of (Q)SARs for the prediction of skin irritation/corrosion potential. Bilthoven, The Netherlands: National Institute of Public Health and Environment (RIVM-SEC)/European Chemicals Bureau (ECB). 1-45. Available at: [http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/information-sources/qsar-document-area/Evaluation_of_Skin_Irritation_qsars.pdf].

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.

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.

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.

Spielmann H., Hoffmann S., Liebsch M., Botham P., Fentem J., 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.

Tornier C., Roquet M., Fraissinette A.B. (2010). Adaptation of the validated SkinEthicTM Reconstructed Human Epidermis (RHE) skin corrosion test method to 0.5 cm2 tissue sample. Toxicol. in vitro 24, 1379-1385.

United Nations-Economic Commission for Europe (UN) (2003). Globally Harmonised System of Classification and Labelling of Chemicals (GHS). Part 3 Health and Environmental Hazards – Chapter 3.3 Serious eye damage / eye irritation. New York, USA, and Geneva, Switzerland: United Nations. p.137-150.

United Nations-Economic Commission for Europe (UN) (2011). Globally Harmonised System of Classification and Labelling of Chemicals (GHS), Fourth revised edition, UN New York and Geneva, 2011. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev04/04files_e.html].

United Nations-Economic Commission for Europe (UN) (2013). Globally Harmonised System of Classification and Labelling of Chemicals (GHS), Fifth revised edition, UN New York and Geneva, 2013. Available at: [http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev05/English/03e_part3.pdf].

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

Walker J.D., Gerner I., Hulzebos E., Schlegel K. (2005). The skin irritation corrosion rules estimation tool (SICRET)". (Q)SAR 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. Toxicol in vitro 18, 231-243.

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 (Q)SAR studies. Toxicology in Vitro 10, 95-100.

WHO (2000). Working Group report on Evaluation and use of epidemiological evidence for environmental health risk assessment: WHO Guideline Document. Environm. Health Perspect. 108, 997-1002.

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.

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. Toxicol. In Vitro 2, 19-26.

ANNEX I

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance, to mean the proportion of correct outcomes of a test method (OECD, 2005).

Acid/alkaline reserve: A measure of the strength of an acidic or alkaline mixture, sometimes also called "buffering capacity". For acidic mixtures this is the amount of sodium hydroxide in gram per 100 g of the acidic mixture required to produce a pH of 4. For alkaline mixtures this is the amount of sulphuric acid in gram per 100 g of the alkaline mixture required to produce a pH of 10.

Acidity / Alkalinity: terms used in OECD Test Guideline 122 and in this document interchangeably with "acid / alkaline reserve".

Applicability Domain: A description of the physicochemical or other properties of the substances for which a test method is applicable for use (OECD, 2005). For (Q)SAR models, the applicability domain (AD) is the response and chemical structure space in which the model makes predictions with a given reliability. The AD of a (Q)SAR can be thought of as a theoretical region in multi-dimensional space in which the model is expected to make reliable predictions. Thus, information on the AD helps the user of the model to judge whether the prediction for a new chemical is reliable or not. The region depends on the nature of the chemicals in the training set, and the method used to develop the model. The development and assessment of methods for defining the domain of applicability is an important area of (Q)SAR research (OECD, 2007a).

"Catch-up" validation study: A validation study for a test method that is structurally and functionally similar to a previously validated and accepted reference test method. The candidate test method should incorporate the essential test method components included in performance standards developed for the reference test method, and should have comparable performance when evaluated using the reference chemicals provided in the performance standards (OECD, 2005).

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

Chemical: means a substance or a mixture.

ESAC: ECVAM Scientific Advisory Committee.

ESAC Statement: Statement on the scientific validity of a new test method, following the peer review of a prospective or retrospective validation study; often associated with a recommendation on the use of this method in regulatory context.

Expert system: A computer-based tool that generates predictions of endpoints by applying (Q)SARs and/or rules designed to recreate the reasoning of experts. Expert systems may also contain a database of experimental data which may be consulted directly and which may be used during the application of the rules (OECD, 2005).

Formulation: see "mixture".

Hazard: In this context, inherent property of a chemical having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that chemical. While "hazard identification" and "hazard characterisation" describe two levels of the process of "hazard assessment", "hazard classification" is a regulatory act.

IATA (Integrated Approach to Testing and Assessment): Integrate existing knowledge based on classes of chemicals with the results of biochemical and cellular assays, computational predictive methods, exposure studies, and other sources of information to identify requirements for targeted testing or develop assessment conclusions. In some cases, the application of IATA could lead to the refinement, reduction, and/or replacement of selected conventional tests (e.g., animal toxicity tests). IATA also has the potential to further enhance the understanding of mode/mechanism of action including the consideration of relevant adverse outcome pathways (AOPs) that provide biological linkages between molecular initiating events to adverse outcomes in individual organisms and populations that are the bases for risk assessments (NAFTA, 2012).

Mixture: A mixture or a solution composed of two or more substances in which they do not react (UN, 2013). According to this definition, also highly complex preparations / formulations of products should be called "mixtures". At a few instances, however, this document uses the terms "preparations" and "formulations" to better describe the limited publicly available experience with these complex products and the new test methods.

Performance Standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (i) essential test method components; (ii) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of reference chemicals (OECD, 2005).

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

Preparation: see "mixture".

(Q)SAR, (Quantitative) Structure Activity Relationship: is a quantitative relationship between a biological activity (e.g. toxicity) and one or more molecular descriptors that are used to predict the activity (OECD, 2007a).

QMRF ((Q)SAR Model Reporting Format): is a harmonised template for summarising and reporting key information on (Q)SAR models, including the results of any validation studies. The information is structured according to the OECD (Q)SAR validation principles. The (Q)SAR Prediction Reporting Format (QPRF) is a harmonised template for summarising and reporting substance-specific predictions generated by (Q)SAR models. See also:
[http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/(Q)SAR_tools/QRF].

Relevance: In context of test methods and non-testing methods of this document, description of the relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (OECD, 2005).

Reliability: measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (OECD, 2005).

Reliable (Q)SAR: is a (Q)SAR that is considered to be "reliable" or "valid" for a particular purpose is a model that exhibits an adequate performance for the intended purpose. The criteria for determining whether the model performance is "adequate" will depend on the particular purpose and are highly context-dependent (OECD, 2007a).

Replacement test: A test which is designed to substitute for a test that is in routine use and accepted for hazard identification and/or risk assessment, and which has been determined to provide equivalent or improved protection of human or animal health or the environment, as applicable, compared to the accepted test, for all possible testing situations and chemicals (OECD, 2005).

Reproducibility: The agreement among results obtained from testing the same substance using the same test protocol (see reliability) (OECD, 2005).

Risk Assessment: A process intended to calculate or estimate the risk to a given target organism, system or (sub)population , including the identification of attendant uncertainties, following exposure to a particular agent, taking into account the inherent characteristics of the agent of concern as well as the characteristics of the specific target system. The Risk Assessment process includes four steps: hazard identification, hazard characterisation (related term: dose-response assessment), exposure assessment, and risk characterisation. It is the first component in a risk analysis process. The definition of risk assessment may vary between Member countries (OECD, 2005).

Sensitivity: The proportion of all positive / active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (OECD, 2005).

Skin corrosion *in vivo*: The production of irreversible damage of the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions (UN, 2013).

Skin irritation *in vivo*: The production of reversible damage to the skin following the application of a test chemical for up to 4 hours (UN, 2013). Skin irritation is a locally arising reaction of the affected skin tissue and appears shortly after stimulation. It is caused by a local inflammatory reaction involving the innate (non-specific) immune system of the skin tissue. Its main characteristic is its reversible process involving inflammatory reactions and most of the clinical characteristic signs of irritation (erythema, oedema, itching and pain) related to an inflammatory process.

Specificity: The proportion of all negative / inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (OECD, 2005).

Substance: Chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition (UN, 2013).

Test chemical: According to an OECD agreement, in OECD Test Guidelines "test chemical" means what

is being tested in the test method.

Tiered testing: Testing which uses test methods in a sequential manner; the test methods selected in each succeeding level are determined by the results in the previous level of testing (OECD, 2005).

United Nations Globally Harmonised System of Classification and Labelling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (UN, 2013).

ANNEX II

EXAMPLE OF MATRIX FOR WEIGHT OF EVIDENCE ANALYSES.

For those modules having available data, entries are filled in the respective cases. For the rest of the entries, NA shall be indicated in column 2. It is recommended to use short and conclusive wording. For assessment of the evidence, refer to the Part 2 of this guidance document. Note that WoE should be assessed before any new experimental data is generated.

Module	**Title of document / full reference; or data not available (NA)**	**Study result and/or positive or negative evidence obtained**	**Data quality,** according to the Klimisch score, when appropriate *	**Adequacy and relevance,** short statement	**Coverage of relevant para-meters and observations,** Yes/No	**Consistency with other information****	**Conclusive remark*****
1. Existing human data							
2.In vivo study							
3.In vitro corrosion data							
4.In vitro irritation data							
5. Other *in vivo* and *in vitro* data							
6.Physico-chemical properties							
7.Non-testing methods ((Q)SAR, grouping, bridging & additivity approaches)							
Overall conclusion	1. WoE allows decision/assessment of the skin irritation/corrosion potential of the substance. The substance should be classified as non-irritant, irritant, corrosive, (non-corrosive), **or** 2. WoE does not allow decision/assessment of skin irritation/corrosion potential of the substances. Recommendation or specification of the most appropriate additional testing.						

*) An electronic tool supporting the quality assessment of in vivo and vitro data through the application of consistent criteria leading to scored results has been developed by EURL ECVAM (described in Schneider et al., 2009). The ToxRTool can be downloaded from the EURL ECVAM page: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/archive-publications/toxrtool

**) For example: “This data (any entry except 3 and 4) is consistent with the existing in vitro studies”.

***) For example: “The existing human data suggest that the substance is an irritant. Due to poor reporting of this data, and low quality in terms of exposure information, the data is inconclusive, and has a low weight in the final evaluation.”