EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Testing Strategies for the Prediction of Skin and Eye Irritation and Corrosion for Regulatory Purposes

Ana Gallegos Saliner & Andrew P. Worth

IHCP

2007

EUR 22881 EN

The mission of IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.
The IHCP carries out research to improve the understanding of potential health risks posed by chemicals, physical and biological agents from various sources to which consumers are exposed.

- As a Research Institute, the IHCP contributes to the improvement of scientific knowledge on health care methods and consumer issues.
- As a European Institution, the IHCP participates, at an international level and in collaboration with Member States Authorities, in R&D and regulatory actions intended to improve consumer tutelage.
- As a European Commission Service, the IHCP acts as an independent advisor for the implementation of risk assessment, monitoring and validation of procedures aimed to insure EU citizens of the use of health and consumer services or products.

European Commission
Directorate-General Joint Research Centre
Institute IHCP

Contact information
Address: European Chemicals Bureau TP581
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 789566
Fax: +39 0332 786717

http://ecb.jrc.it/QSAR
http://www.jrc.cec.eu.int

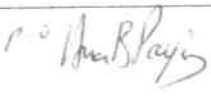
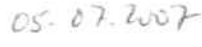
EUR 22881 EN

ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Quality control insert

	Name	*Signature*	*Date*
Report Prepared by:	**Ana Gallegos Saliner**		**30/06/07**
Reviewed by: (Scientific level)	**Andrew Worth**		**04/07/07**
Approved by: (Head of Unit)	**Steven Eisenreich**		
Final approval: (IHCP Director)	**Elke Anklam**		

ABSTRACT

This report reviews the use of stepwise testing approaches for the prediction of skin and eye irritation and corrosion in a regulatory context. It is published as a companion report to the *Review of Literature-Based Models for Skin and Eye Irritation and Corrosion*, an ECB report which reviewed the state-of-the-art of *in silico* and *in vitro* dermal and ocular irritation and corrosion human health hazard endpoints. In the former review, the focus was placed on reviewing alternative *in silico* approaches to assess acute local toxic effects, such as QSARs, SARs, chemical categories, and read-across and analogue approaches. Special emphasis was placed on literature-based (Q)SAR models for skin and eye irritation and corrosion and expert systems. In the present review, the emphasis is on different schemes (testing strategies) that have been conceived for the integrated use of different approaches, including *in silico*, *in vitro* and *in vivo* methods.

LIST OF ABBREVIATIONS

C / NC	Corrosive / Non-Corrosive
CI	Confidence Intervals
CM	Classification Model
DM	Dipole Moment
ECB	European Chemicals Bureau
ECVAM	European Centre for the Validation of Alternative Methods
EEC	European Economic Community
EHS	Environmental Health and Safety
EU	European Union
EWG	Endpoint Working Group (within REACH Implementation Project 3.3)
GHS	Globally Harmonised System
I / NI	Irritant / Non-Irritant
IRAG	Interagency Regulatory Alternatives Group
ITS	Integrated (Intelligent) Testing Strategy
Log*P*	Logarithm of the octanol/water partition coefficient
MP	Melting Point
MV	Molecular Volume
OECD	Organisation for Economic Co-operation and Development
PM	Prediction Model (for converting *in vitro* to *in vivo* data)
pK_a	-$\log_{10}$ of the acid dissociation constant
(Q)SAR	(Quantitative) Structure Activity Relationship
REACH	Registration, Evaluation, and Authorisation of Chemicals
RIP	REACH – Implementation Project
SAR	Structure-Activity Relationship
SICRET	Skin Irritation Corrosion Rules Estimation Tool
TER	(rat skin) Transcutaneous Electrical Resistance test
WoE	Weight of Evidence

TABLE OF CONTENTS

1. Introduction

Toxicological testing of chemicals for risk assessment, aiming at the prediction of adverse effects to human health and the environment, involves high costs, in terms of time, money, and animal welfare. To perform testing effectively in the regulatory context, chemicals should be adequately selected and prioritised for testing, and improved predictive test systems for new endpoints of concern should be designed. There is an urgent need for test strategies that can fill the data gaps for a large number of untested substances as efficiently as possible. Such strategies must take into account the limitations in economic resources and testing capacity. They also have to be in line with the aim to reduce the use of animals for toxicity testing.

A test system consists of a set of individual tests and a system of rules and criteria that are used to select tests for each individual substance and to determine in which order they are performed. In testing systems, relatively simple tests are performed for all the chemicals to be assessed. The outcomes of the initial tests are then used to prioritise substances for additional, more resource-intensive testing.

Two types of strategies can be formulated: animal-free strategies comprise *in silico* and *in vitro* data, whereas reductive strategies (i.e. strategies which reduce the need for animal testing) also take into account *in vivo* data. For both types of strategies, tiered and integrative strategies can be envisaged. In a tiered strategy, the data are generated stepwise. The decision at each step is based only on the newly generated data, without taking into account the previous information. Integrative strategies can either be stepwise, or a battery. In a stepwise integrated strategy, the data are also generated stepwise. However, the decision at each step is based on all the available data at that step. Finally, in a battery strategy all data are generated simultaneously. There is only one decision point, which considers all data. In this document, both tiered and integrative strategies will be considered.

The use of ITS have become widespread in the pharmaceutical field, where large numbers of candidate test substances need to be tested and screened to filter the ones likely to be toxic in subsequent testing [1]-[2]. They are also very useful in chemical assessment for the evaluation and prioritisation of large number of chemicals, based on early predictions of their potential toxicity. With the

implementation of REACH, integrated testing strategies are especially relevant within the regulatory context [3]-[4].

In the regulatory context, Integrated Testing Strategies (ITS), which are used to make classification decisions on the basis of non-animal data or to evaluate whether *in vivo* testing is needed, are widely used. These strategies consist of a series of tests performed in a defined sequential manner. The tests selected in each successive level are determined by the results in the previous level of testing in a stepwise process that leads to a decision. Testing strategies start by using existing data to enable *in silico* based toxicity predictions, including the application of (Q)SARs and decision models based on physicochemical data. In a successive step they also encompass the use of *in vitro* methods, and only if necessary they consider the application of *in vivo* tests. Special focus is made on the prospects for using (Q)SAR modelling and read-across as part of integrated testing strategies for chemical risk assessment [5]. However, future perspectives envisage the use of promising innovative techniques, such as genomics, proteomics, metabonomics for improved strategies for toxicity prediction [6]-[7].

Decision-tree schemes are also commonly used intelligent testing schemes within regulatory frameworks. In this type of strategy, at certain stages in each scheme, a decision on whether to classify and label and/or to undertake a risk assessment with respect to the test substance is made via a Weight of Evidence (WoE) process [8]. The decision on whether to stop or continue testing depends on the amount and quality of the information available, and the validation status of the tests used to generate data at each WoE evaluation step.

A rigorous strategy to combine single tests into efficient testing systems should be ideally based on standard decision theory. Such theory should include chemical, toxicological, and decision-theoretical knowledge, and take into account the optimisation of test systems, and also validity, reliability, sensitivity, and cost efficiency concepts [9]. The assessment of test performance characteristics, mechanistic understanding, extended quality assurance, formal validation and the use of integrated testing strategies should be performed to optimize the balance between safety, costs and animal welfare [10].

Acute local irritancy and corrosivity are mainly assessed in two contexts: in the hazard classification of industrial chemicals and in the safety assessment of

ingredients and mixtures used in industrial, pharmaceutical and consumer products. The intended purpose in each context is different, and therefore the considerations made in the two situations are not equivalent.

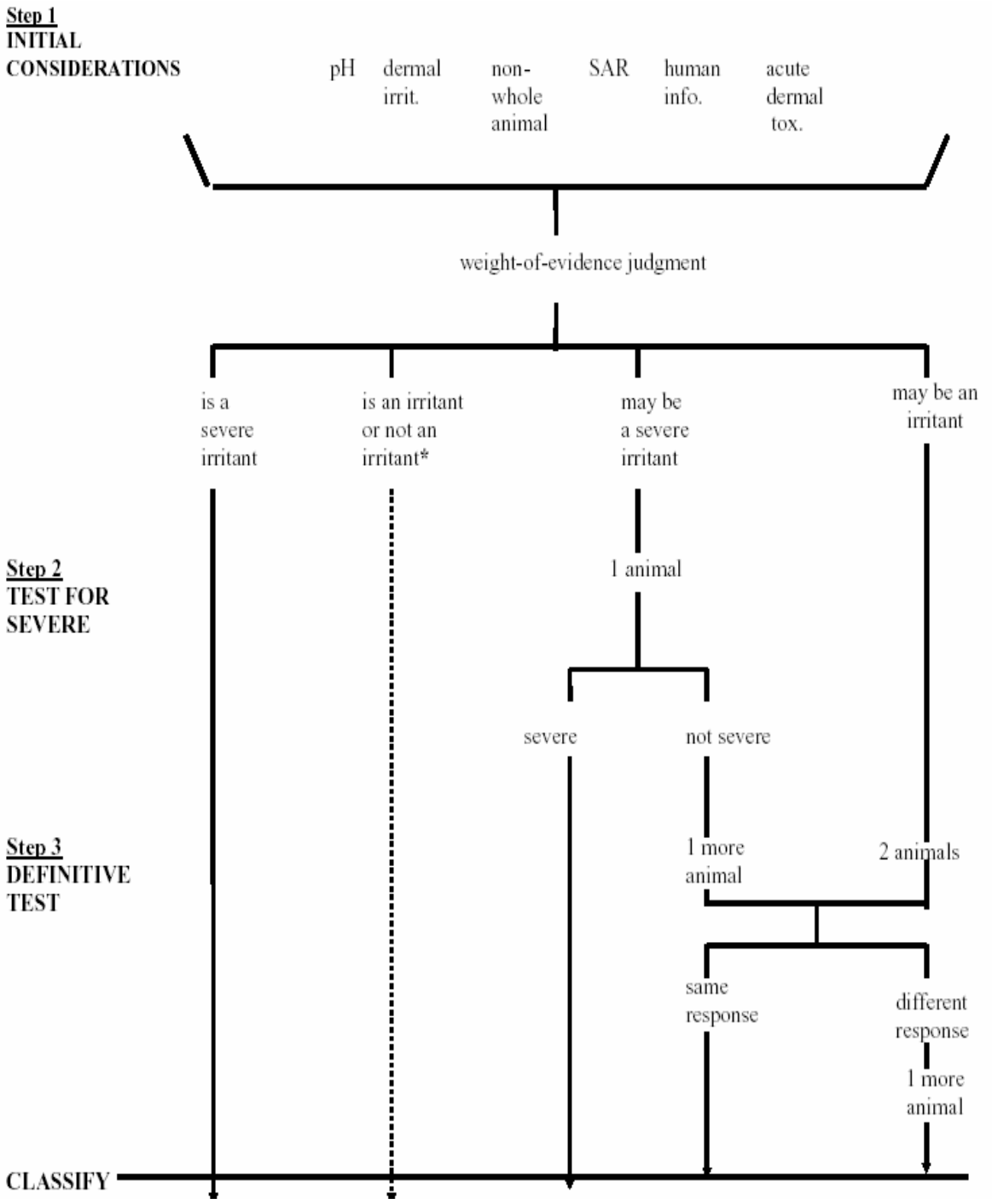
In the hazard identification of chemicals, the purpose of testing is to assess the irritation/corrosion potential according to classification schemes defined by regulatory authorities. Current regulatory proposals recommend a stepwise approach to hazard identification in which chemicals can be classified as irritating/corrosive on the basis of results from non-animal methods. Since testing in animals is only required as a last step to confirm negative results generated by non-animal tests, these testing strategies contribute to the reduction and refinement of *in vivo* tests. The non-animal methods act as partial replacements of the animal test screens, being used to place chemicals into two or more categories of irritation/corrosion potential, without generating too many false positive results. In such strategies, there is less concern about the generation of false negatives because these can be identified by the animal tests carried out in the last step of the process.

In contrast, in the safety assessment of ophthalmologic and cosmetic ingredients, mixtures and products, the purpose is to demonstrate that the products will not cause adverse effects. In this case, the placement of test substances into broad irritation/corrosion categories is often not sufficient, since it is necessary to establish the absence of adverse effects at lower concentrations. Although there is an increasing reliance on non-animal methods, it is more challenging to assess the reliability of predictions for product safety.

2. Seminal testing strategies for eye irritation

In 1991, the Interagency Regulatory Alternatives Group (IRAG), an *ad hoc* organization composed of staff from U.S. regulatory agencies, sponsored an international workshop to propose updates for *in vivo* eye irritation test methods. As a result, a testing and evaluation scheme for the determination of eye irritation potential of chemicals was proposed. The IRAG testing process starts with initial considerations before commencing animal testing on a chemical, specifically on physicochemical properties including pH extremes, and the use of potential buffering capacity information; evidence of dermal irritation or corrosivity; validated and accepted non-whole animal (*in vitro*) alternatives; structure-activity relationships (SAR); and human experience. A sixth parameter, acute dermal toxicity, was considered because any acute toxic agent via dermal route is assumed to be also toxic to the eye. The resulting scheme, which takes into account a Weight of Evidence (WoE) approach for judgement, is shown in Figure 1.

Figure 1. IRAG tiered scheme for eye irritation testing. [11].



In April 1993, the German competent authority for the regulatory assessment of chemicals presented a decision-tree scheme for eye irritancy testing at a symposium in Ottawa [12]. The scheme, based on the experience evaluating test reports submitted under the EU New Chemicals notification procedure, is shown in Figure 2. Briefly the steps of this strategy can be summarised as follows:

Measurement of pH (Step 1). Eye irritancy properties of strongly acidic (pH < 2) or alkaline substances (pH > 11.5) need not be tested because of their probable corrosive properties. Buffering capacity is also taken into account.

Evaluation of skin corrosivity/irritancy (Step 2). Eye irritant properties of substances known to be corrosive to skin are not tested for ethical reasons, even though a number of skin corrosive substances demonstrate only mild eye irritation.

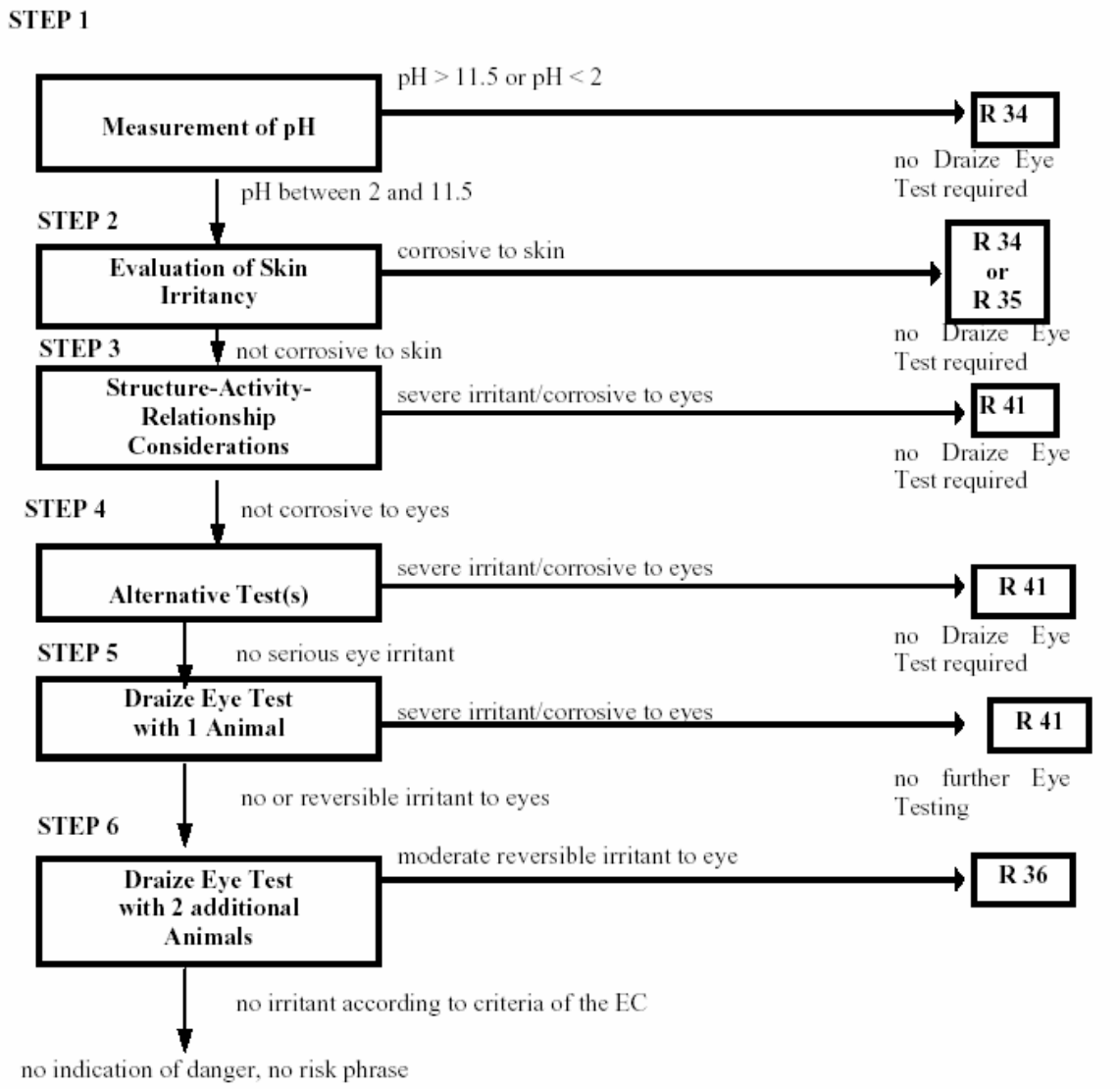
Structure-activity-relationships considerations (Step 3). Theoretical considerations on qualitative structure-activity-relationships or computational chemical modelling to predict eye irritant properties are considered.

***In vitro* methods or other alternative tests** (Step 4). Results of validated alternative methods predicting serious eye irritant/corrosive effects to eyes may be sufficient for labelling the substances as R41.

Draize eye test with one animal (Step 5). In the case of results showing severe irritant/corrosive effects, the substance is labelled as R41 and no further testing is required.

Draize eye test with two additional animals (Step 6). If the outcome is negative after the last step, there is no indication of danger and, subsequently, no risk phrase is assigned.

Figure 2. Eye Irritancy Testing Strategy for New Chemicals within the Notification Procedure of the European Community [11].



As a result of discussions at the 1991 IRAG workshop, it became apparent that there were areas of consensus and areas where there were different opinions as to how the test should be performed and evaluated. Consequently, a second set of proposals for modifications to the 1987 OECD eye irritation test guideline was sent out for review and comment. After an international discussion, the proposals were submitted to OECD for additions and modifications, which resulted in a proposal for the revised TG 405 [13], and a tiered scheme for eye irritation testing to be annexed to the updated OECD guideline was worked out [14].

3. Tiered testing strategy adopted by the GHS

The recommended testing strategy in the Globally Harmonised System (GHS) for the Classification of Chemical Substances [15] is based on a collection of test guidelines and classification schemes [16]. The IRAG scheme of the U.S. regulators (Figure 1), the experiences of the German regulators based on the EU chemicals notification procedure (Figure 2), and the outcome of the OECD Workshop on *Harmonisation of Validation Criteria for Alternative Tests / Harmonisation and Acceptance Criteria for Alternative Toxicological Test Methods* held in Solna, Sweden, in January 1996, were integrated into a single proposal for a testing strategy to be included in TG 405 [17].

Test methods in Annex V to Directive 67/548 for skin irritation/corrosion [18] and for eye irritation/corrosion [19] focus on possible improvements through the evaluation of all existing information on test substances. The sequential testing strategy provides a WoE approach for the evaluation of existing data on the irritation/corrosion properties of substances and a tiered approach for the generation of relevant data on substances for which additional studies are needed or for which no studies have been performed. For new substances this stepwise testing approach for developing scientifically sound data on the corrosivity/irritation of the substance is recommended. For existing substances with insufficient data on skin and eye corrosion/irritation, the strategy should be used to fill missing data gaps.

Prior to undertaking tests as part of the sequential testing strategy, all available information should be evaluated to determine the need for *in vivo* testing. Although significant information might be gained from the evaluation of single parameters, the totality of existing information should be assessed. All relevant data on the effects of the substance in question, and its structural analogues, should be evaluated in making a WoE decision, and a rationale for the decision should be presented. Primary emphasis should be placed upon existing human and animal data on the substance, followed by the outcome of *in vitro* or *ex vivo* testing. *In vivo* studies of corrosive substances should be avoided as far as possible. The testing strategies for eye irritation (Figure 3) and for skin irritation/corrosion (Figure 4) include the following steps:

Evaluation of existing human and animal data (Step 1). Existing human data, e.g. clinical and occupational studies, and case reports, and/or animal test data

should be considered first. Substances with sufficient evidence of non-corrosivity and non-irritancy from previously performed studies should also not be tested in *in vivo* studies.

Analysis of structure activity relationships (SAR) (Step 2). The results of testing of structurally related chemicals should be considered, if available. When sufficient human and/or animal data are available on structurally related substances or mixtures of such substances to indicate their eye corrrosion/irritancy potential, it can be presumed that the test substance will produce the same responses. In those cases, the substance may not need to be tested. Conversely, negative data from studies of structurally related substances or mixtures of such substances do not constitute sufficient evidence of non-corrosivity/ non-irritancy of a substance. Valid and accepted SAR approaches should be used to identify the corrosion and irritation potential for both dermal and ocular effects.

Physicochemical properties and chemical reactivity (Step 3). Substances exhibiting pH extremes such as pH $\leq$ 2.0 or pH $\geq$ 11.5 may have strong local effects. If extreme pH is the basis for identifying a substance as corrosive or irritant to the eye, then its acid/alkaline reserve (buffering capacity) may also be taken into consideration. If the buffering capacity suggests that a substance may not be corrosive, then further testing should be undertaken to confirm this, preferably by the use of a validated and accepted *in vitro* or *ex vivo* test.

Consideration of other existing information (Step 4). All available information on systemic toxicity via the dermal route should be evaluated at this stage. If the test substance has been shown to be very toxic by the dermal route, it may not need to be tested in the eye. Although there is not necessarily a relationship between acute dermal toxicity and eye irritation/corrosion, it can be assumed that if an agent is very toxic via the dermal route, it will also exhibit high toxicity when instilled into the eye.

Alternatively, for skin irritation/corrosion, if a chemical has proven to be very toxic by the dermal route, an *in vivo* dermal irritation/corrosion study may not be practicable because the amount of test substance normally applied could exceed the very toxic dose and, consequently result in the death or severe suffering of animals.

Results from *in vitro* or *ex vivo* tests (Steps 5 and 6). Substances that have demonstrated corrosive or severe irritant properties in an *in vitro* or *ex vivo* test that

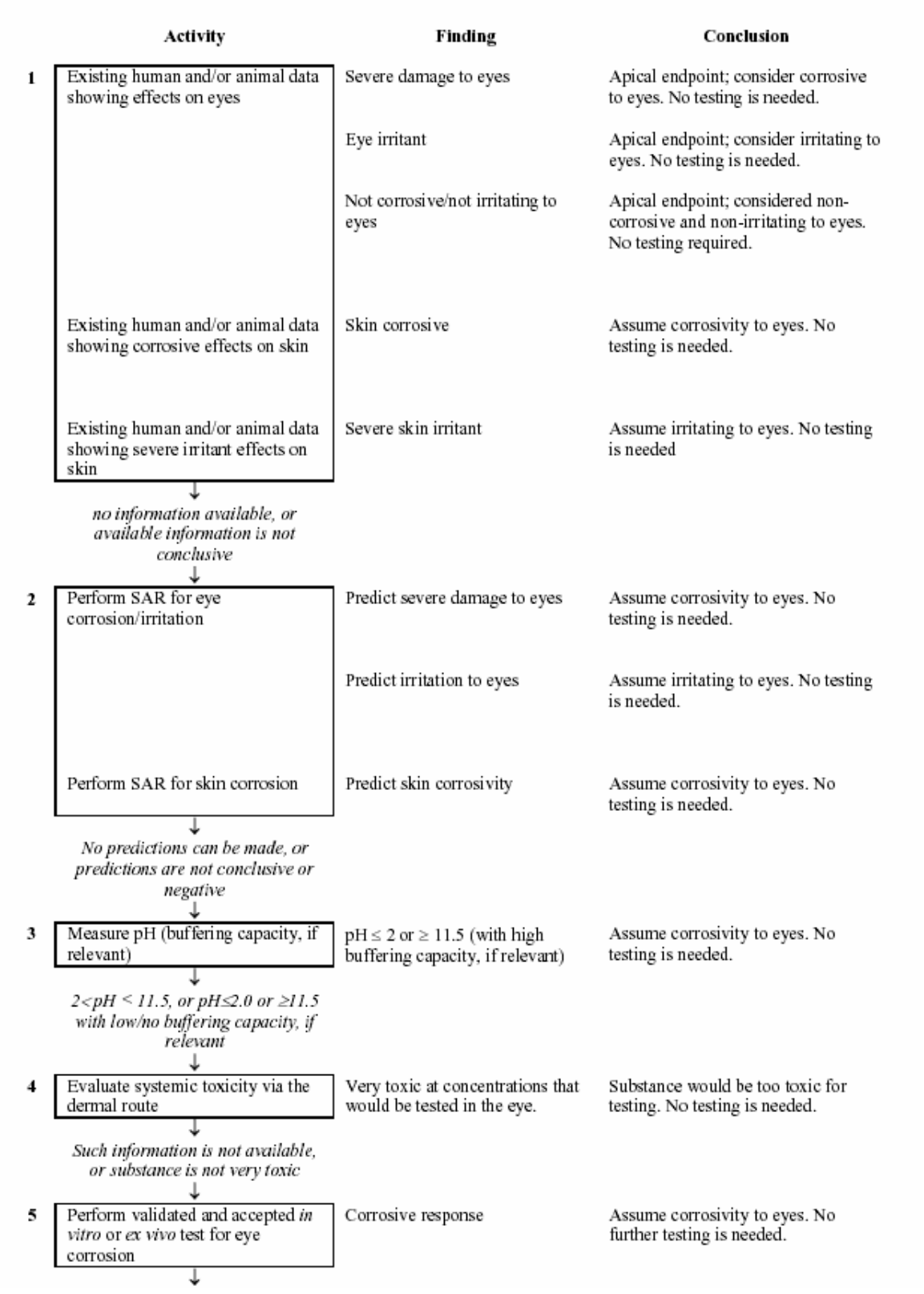
has been validated and accepted for the assessment of eye or skin corrosivity/irritation do not need to be tested in animals. It can be presumed that such substances will produce similar severe effects *in vivo*.

Assessment of *in vivo* dermal irritancy or corrosivity of the substance (Step 7 for eye). This step is only carried out in the case of eye irritation/corrosion. When insufficient evidence exists with which to perform a conclusive WoE analysis of the potential eye irritation/corrosivity of a substance based upon data from the studies listed above, the *in vivo* skin irritation/corrosion potential should be evaluated first. If the substance is shown to produce corrosion or severe skin irritation, it should be considered to be a corrosive eye irritant unless other information supports an alternative conclusion. Thus, an *in vivo* eye test does not need to be performed. If the substance is not corrosive or severely irritating to the skin, an *in vivo* eye test should be performed.

***In vivo* test in rabbits** (Step 8 and 9 for eye; Step 7 and 8 for skin). *In vivo* ocular testing should begin with an initial test using one animal. If the results of this test indicate the substance to be a severe irritant or corrosive to the eyes, further testing should not be performed. If that test does not reveal any corrosive or severe irritant effects, a confirmatory test is conducted with two additional animals.

In the case of skin irritation/corrosion, this corresponds to Steps 7 and 8. Should a WoE decision be made to conduct *in vivo* testing, it should also begin with an initial test using one animal. If the results of this test indicate the substance to be corrosive to the skin, further testing should not be performed. If a corrosive effect is not observed in the initial test, the irritant or negative response should be confirmed using up to two additional animals for an exposure period of four hours. If an irritant effect is observed in the initial test, the confirmatory test may be conducted in a sequential manner, or by exposing the two additional animals simultaneously.

Figure 3. Testing and Evaluation Strategy for Eye Irritation/Corrosion [19].



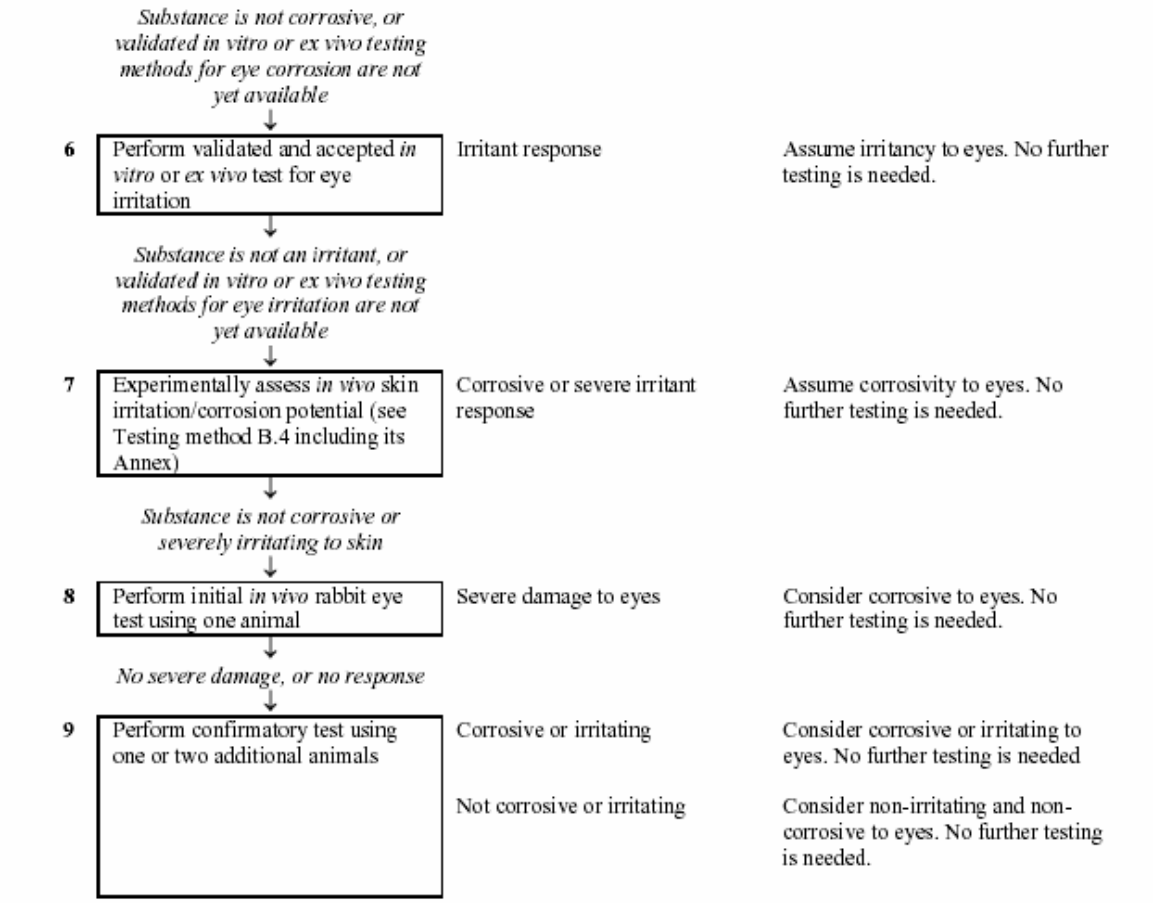
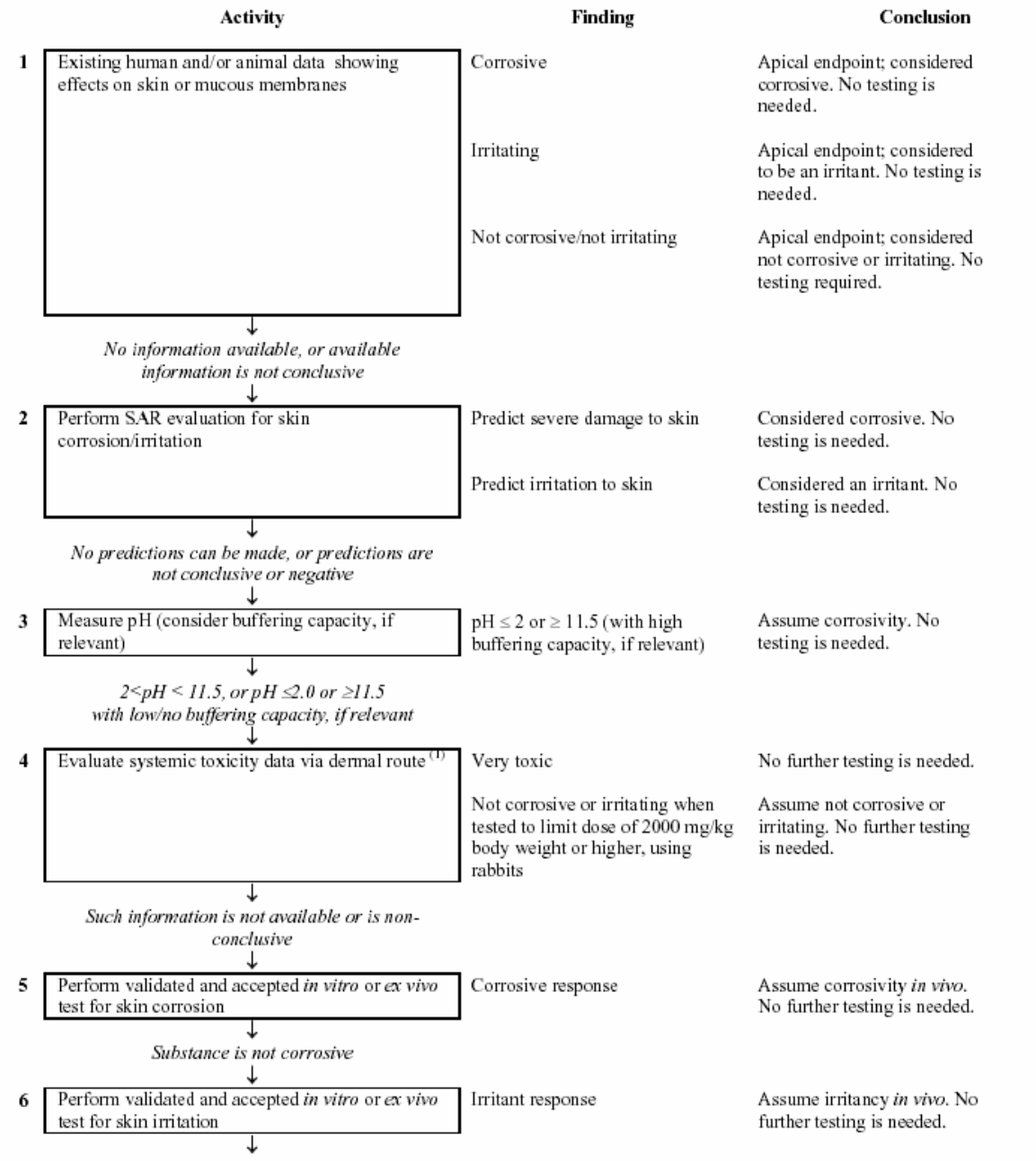
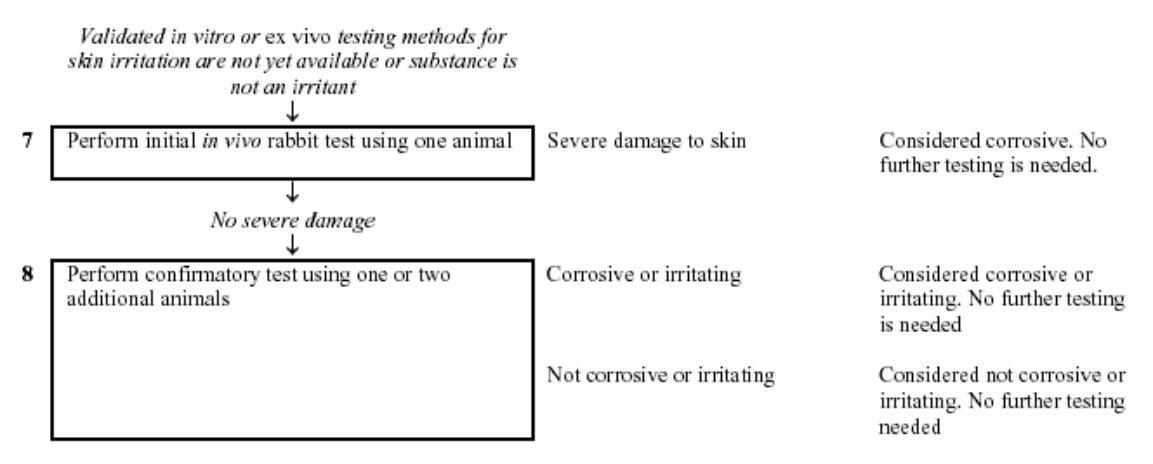

Substance is not corrosive, or validated in vitro or ex vivo testing methods for eye corrosion are not yet available
6
Perform validated and accepted in vitro or ex vivo test for eye irritation
Irritant response
Assume irritancy to eyes. No further testing is needed.
Substance is not an irritant, or validated in vitro or ex vivo testing methods for eye irritation are not yet available
7
Experimentally assess in vivo skin irritation/corrosion potential (see Testing method B.4 including its Annex)
Corrosive or severe irritant response
Assume corrosivity to eyes. No further testing is needed.
Substance is not corrosive or severely irritating to skin
8
Perform initial in vivo rabbit eye test using one animal
Severe damage to eyes
Consider corrosive to eyes. No further testing is needed.
No severe damage, or no response
9
Perform confirmatory test using one or two additional animals
Corrosive or irritating
Consider corrosive or irritating to eyes. No further testing is needed
Not corrosive or irritating
Consider non-irritating and non-corrosive to eyes. No further testing is needed.

Figure 4. Testing and Evaluation Strategy for Skin Irritation/Corrosion [18].




Validated in vitro or ex vivo *testing methods for skin irritation are not yet available or substance is not an irritant*
7
Perform initial *in vivo* rabbit test using one animal
Severe damage to skin
Considered corrosive. No further testing is needed.
No severe damage
8
Perform confirmatory test using one or two additional animals
Corrosive or irritating
Considered corrosive or irritating. No further testing is needed
Not corrosive or irritating
Considered not corrosive or irritating. No further testing needed

4. Literature-based testing strategies

In the literature, several stepwise hierarchical testing strategies have been proposed to hazard evaluation. They are also based on the sequential application of one or more alternative methods prior to the use of any animal test, with the purpose of reducing and refining the use of animals in toxicity testing without compromising human safety.

Tiered assessment strategies have been applied jointly and separately to both endpoints. There is strong evidence that chemicals that are corrosive to the skin should also be classified as being corrosive to the eye, especially if the assessment is made from knowledge of acidity and alkalinity. In particular, in the EU and OECD classification schemes, chemicals that have been found to be corrosive to the skin are automatically considered to be corrosive to the eye as well (and are therefore labelled as such without animal testing).

A report by Balls *et al.* [20] describes some initiatives aimed at the short-term reduction and refinement of animal use, and the long-term replacement of the Draize test in eye irritation testing. These initiatives included the evaluation of the use of reference standards (benchmark chemicals) in the validation process; an evaluation of tiered testing strategies; further analyses of the data generated during previous validation studies; and research into the mechanistic basis of eye irritation. Special emphasis was placed on the review of stepwise testing strategies. Hierarchical testing schemes proposed in the literature for skin irritation/corrosion [21]-[22], and proposals based on the combined use of a cytotoxicity test and an organotypic test [23]-[25] for eye irritation were cited. The tiered approach to eye irritancy/corrosivity testing provided in the 1987 update of OECD TG 405 [17] was also reviewed.

A study published by Worth *et al.* [26] evaluated the use of a stepwise hierarchical testing strategy consisting of three steps to classify skin corrosives. The effect of applying the three steps, taken individually and in sequence, was assessed by using a set of 60 chemicals. The alternative methods considered included the use of structure-activity relationships (SAR); physicochemical properties (pH measurements and acid/alkali reserve), and *in vitro* tests. Within this strategy, animal tests were only used to confirm negative results (non-corrosive) generated by one or more alternative methods. In the fist step, two-descriptor prediction models (PM) were derived by

binary logistic regression. Separate models were derived for organic acids, organic bases, phenols, electrophiles, and neutral organics; they were selected from a range of possible PMs on the basis of their statistical significance. The most recurrent parameter (appearing in the predictive models of acids, phenols, and electrophiles) was Log*P*, followed by molecular volume (MV) and the melting point (MP), the negative logarithm of the dissociation constant (pK_a), and dipole moment (DM). In the second step, a PM based on the pH was applied to each substance; for pH $\leq$ 2 or pH $\geq$ 11.5, the substance was classified as corrosive. Although the combined use of pH and acid/alkali reserve was evaluated, it turned out not to be relevant. In the final step, two *in vitro* methods were evaluated, i.e. the rat skin transcutaneous electrical resistance (TER) assay and the EPISKIN assay. The predictive ability of individual steps was 95% for SARs, 77% for pH, and around 82-83% for *in vitro* tests (Table 1). To assess the predictive ability of the sequence of three steps for the tiered testing strategy, if a chemical was predicted to be corrosive (C) by one of the alternative methods, a C classification was assigned and there was no further progress through the strategy. Conversely, if a chemical was predicted to be non-corrosive (NC), it entered the next step to check whether the prediction had been a false negative. Progress through the strategy continued until the chemical was either predicted to be C by one of the alternative methods, or until a classification C/NC was assigned on the basis of the results of the rabbit test. The results showed that the sequential application of the three alternative methods for the integrated testing strategy allowed the classification of chemicals as C or NC with sufficient reliability.

Table 1. Prediction models and performance of individual steps in a stepwise testing strategy for skin corrosivity [26].

Step	Applicability	Predictive Model (PM)	% Concordance (Specificity; Sensitivity)
1st step	for Organic Acids	If $1.055\log P + 0.082MP \leq 6.896$, C	95 (86; 100)
	for Organic bases	If $1.926\,pKa - 0.507\log P \geq 15.7$, C	
	for Phenols	If $0.087MV - 1.908\log P \leq 3.634$, C	
	for Electrophiles	If $0.116MV - 1.355DM \leq 5.42$, C	
2nd step	for Skin Corrosion	If $pH \leq 2$ or if $pH \geq 11.5$, C	77 (56; 94)
	combined use of pH and acid/alkali reserve	If pH – acid reserve/6 ≤ 1, or If pH + alkali reserve/12 ≥ 14.5, C	50 (29; 92)
3rd step	a different prediction model for each single *in vitro* endpoint (TER, EPISKIN)		82-83 (85-93; 73-82)
	Prediction Models in the form "*If* condition, *then* predict C" (C: Corrosive)		

Subsequently, Worth *et al.* [27] reported a similar study to [26] to evaluate stepwise testing strategies for eye irritation/corrosion. The approach was also based on the sequential application of three steps involving alternative methods prior to animal testing. As in the previous scheme, if any of the alternative methods predicted the chemical of interest to be toxic, a classification was assigned and testing was stopped. Otherwise, testing continued to the next step. In this way, toxic chemicals could be screened out by alternative methods, so that animal tests conducted in the final step would mainly serve to confirm predictions of non-toxicity.

The three steps were applied both on their own and as a sequence to the training set made up of 60 chemicals [26]. The predictivities of nine *in vitro* tests were examined, and PMs were derived to compare their performances. In the first step, two PMs based on physicochemical properties were derived, one for aliphatic chemicals and the other one for aromatic chemicals. A PM for aliphatic chemicals related dipole moment (DM) and Log*P*; the SAR described a two-dimensional ellipse enclosing irritant chemicals and excluding non-irritant chemicals, and it was based on the observation that irritant aliphatic chemicals form an embedded cluster within the more diffuse cluster of non-irritants. The PM for aromatic chemicals established ranges of Log*P* values to classify I from NI. In the second step, pH ranges to differentiate C

from NC, and I from NI (comprised within the previous range) were set. The performance of individual steps was 74% concordance for SARs, and 47% for pH. Also the combined use of pH acid/alkali reserve was evaluated, resulting in the non-significant concordance of 45%. In the third step, the single endpoints of *in vitro* tests with lowest false positive rates were evaluated, obtaining comparable results in all the cases (between 64 and 79%), as shown in Table 2.

Table 2. Prediction models and performance of the individual steps of a tiered testing strategy for eye irritation [27].

Step	Applicability	Predictive Model (PM)	% Concordance (Specificity; Sensitivity)
1st step	for Aliphatics (Eye Irritation)	If $\frac{(\log P - 1.408)^2}{2.002^2} + \frac{(DM - 1.576)^2}{0.215^2} \leq 1$, I	74 (71; 80)
	for Aromatics (Eye Irritation)	If $-1.09 \leq \log P \leq 2.72$, I	
2nd step	for Corrosion	If $pH \leq 2$ or if $pH \geq 11.5$, C	47 (35; 81)
	for Eye Irritation	If $pH \leq 3.14$ or if $pH \geq 9.35$, I	
	combined use of pH and acid/alkali reserve	If pH – acid reserve/6 ≤ 1, or If pH + alkali reserve/6 ≥ 13, I	45 (33; 80)
3rd step	a different prediction model for each single *in vitro* endpoint		64-79 (74-100; 0-81)
Prediction Models in the form "If condition, then predict C / I" (C: Corrosive / I: Irritant)			

Tiered testing strategies for skin corrosion have been developed and assessed by Worth and Cronin [28]-[31]. The evaluation of a two-step strategy, based on the sequential use of pH measurements and *in vitro* data, indicated that the combined use of these data improved the ability to predict corrosion potential [28]. Subsequently, a three-step strategy was reported, based on the sequential use of QSARs, pH measurements and *in vitro* data [26]. A separate study by Worth confirmed the usefulness of pH as a predictor of skin corrosion potential, and provided a new prediction model (PM) for identifying corrosive chemicals by a pH-dependent mechanism [30].

Tiered assessment schemes for the prediction of skin irritation and corrosion have been designed and evaluated by Worth [32]. The first step of the process

consisted of classifying corrosivity based on melting point (MP) and molecular weight (MW). This was followed by the use of a classification model (CM) based on pH (Step 2). Subsequently, *in vitro* data from the EPISKIN assay were used (Step 3). If the compound was ultimately classified as non-corrosive in the third step, a similar iterative process was performed to identify skin irritants. Ultimately, if the compound was predicted non-irritant, the *in vivo* Draize skin test was applied. Thus, the purpose of such a scheme was to classify chemicals and to confirm negative classifications with the use of animal tests.

In a subsequent study, Worth *et al.* [28] evaluated the uncertainty associated with the predictive abilities of two-group classification models (CM), expressed in terms of Cooper statistics. Standard and percentile bootstrap resampling techniques were used to judge whether predicted classifications were significantly better than the predictions made by a different CM, or whether the performance of a CM exceeded predefined performance criteria in a statistically significant way. This method was illustrated by constructing 95% confidence intervals (CI) for the Cooper statistics of four alternative skin corrosivity tests (TER, EPISKIN, Skin^2, and CORROSITEX), as well as two-step sequences in which each *in vitro* test was used in combination with a physicochemical test for skin corrosion based on pH measurements. The PMs were applied to a dataset of 60 chemicals already published [26]. Cooper statistics were used to determine whether the four two-step sequences, with sensitivities greater than or equal to 70%, were significantly more predictive than the four stand-alone *in vitro* tests. This study showed that the performances of the TER, EPISKIN, Skin^2 and CORROSITEX tests in combination with the pH test were better than the individual performances of the *in vitro* tests (Table 3).

Table 3. Classification results of the individual *in vitro* tests and the *in vitro* tests in sequence with the pH test.

	Bootstrap mean estimates					
In vitro test	Concordance		Sensitivity		Specificity	
TER	78	83*	87	96*	71	73*
EPISKIN	81	88*	83	96*	80	82*
Skin2	75	83*	45	70*	100	94*
CORROSITEX	73	74*	71	72*	76	78*

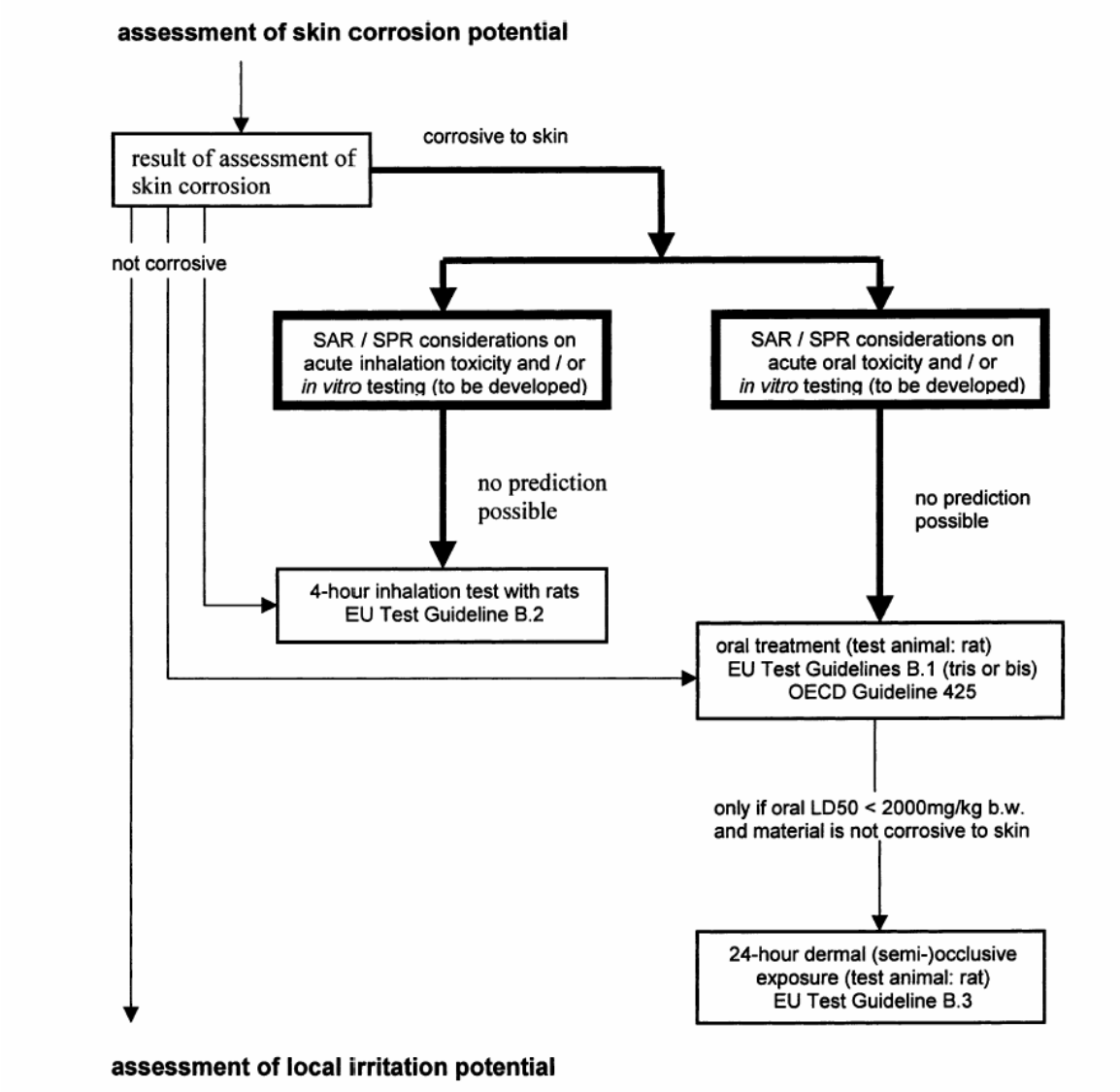
* Use of the *in vitro* test in sequence with the pH test

The bootstrap resampling technique was subsequently applied to assess the variability of Draize tissue scores to estimate acute dermal and ocular effects [29]. This technique was used to estimate biological variability arising from the use of different animals, and temporal variability arising from different time-points. The estimates of variability where then used to determine the extent to which Draize skin and eye test tissue scores could be predicted. A dataset of 143 ECETOC chemicals was used for the variability of Draize skin scores, while for eye scores the dataset consisted of 92 ECETOC chemicals. The results indicated that the variability in Draize skin scores was such that no model for predicting PII could be expected to have r^2>0.57; in contrast, the variability in Draize eye scores was such that no model for predicting MMAS could be expected to have r^2> 0.81.

Gerner and Schlede [33] reviewed the introduction of *in vitro* data into local irritation/corrosion testing strategies; these tiered testing strategies combine *in vitro* tests and SARs. They reported several strategies, from the more general assessment of acute toxic hazard (Figure 5), to the classification of the skin corrosion potential (Figure 6), and the local irritation potential (Figure 7).

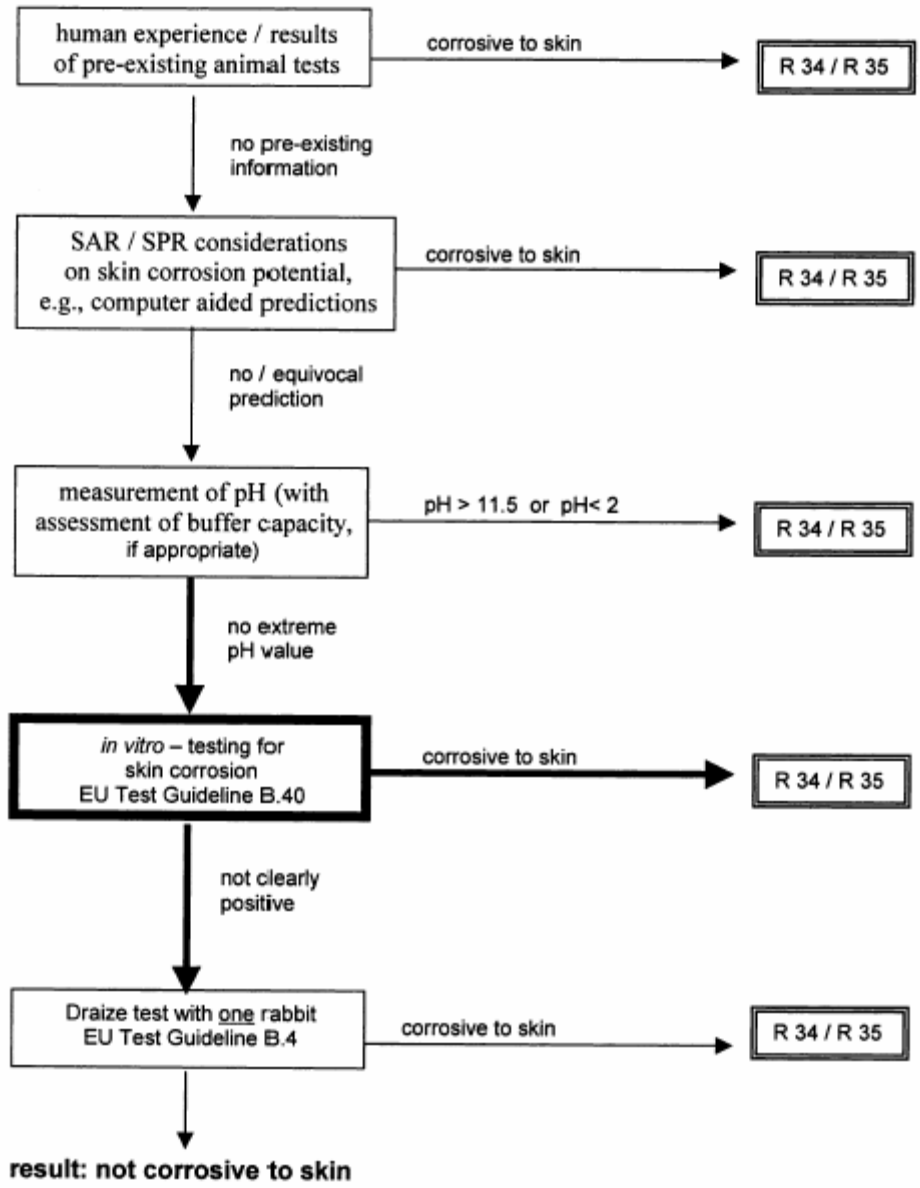
The assessment and classification of the acute toxicity of a chemical observed after swallowing of the substance, after inhalation of its gases, vapours or aerosols and/or after skin contact was deduced from the results of standardised testing with rodents. It was assumed that if corrosive effects in contact with skin were observed, corrosivity in the stomach or in lungs by oral ingestion or inhalation could be predicted. This observation enabled a differentiation between dangerous substances because of their universal corrosive properties and substances exhibiting corrosion exclusively in contact with skin. Those chemicals should be tested for oral and inhalation toxicity irrespective of their corrosive properties, because they could cause systemic effects after oral or inhalation exposure not related to their corrosive dermal properties. The aim of this strategy was to avoid or reduce acute dermal toxicity testing of skin corrosive chemicals.

Figure 5. Tiered testing and assessment strategy for acute toxicity of chemicals [33].



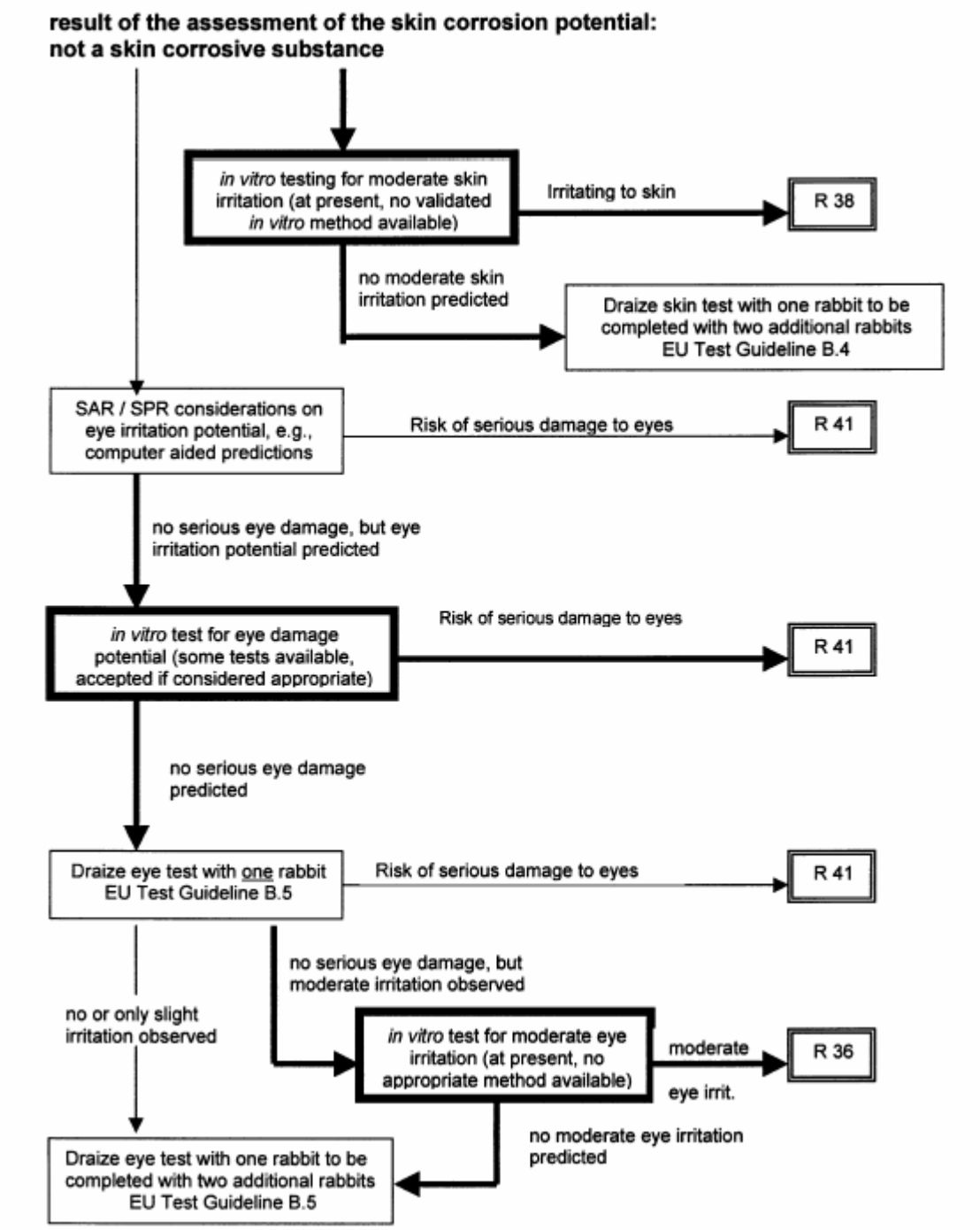
Since skin corrosivity is considered as a crucial effect, the assessment of the skin corrosion potential of a chemical should be performed prior to any animal testing according to international test guidelines. Thus, development and validation of *in vitro* tests for the replacement of the Draize skin test by non-animal alternatives have been intensively explored. These efforts resulted in the European Test Guideline B.40 Skin Corrosion [34] which was adopted by the EU Member States in 2000. The OECD developed a testing and assessment strategy in order to provide guidance on how to base hazard classification on data obtained with *in vitro* or *ex vivo* methods.

Figure 6. Assessment of skin corrosion potential of chemicals (classification according to EU regulations) [33].



The testing and assessment strategy for the classification of skin and eye irritation/corrosion within the GHS was also reported [35]. It was demonstrated that, for a proper assessment and classification of local irritation caused by a single contact with skin or eyes, three different kinds of *in vitro* or *ex vivo* data could replace irritation testing with rabbits: testing for moderate skin irritation; testing for serious eye damage; and testing for moderate eye irritation.

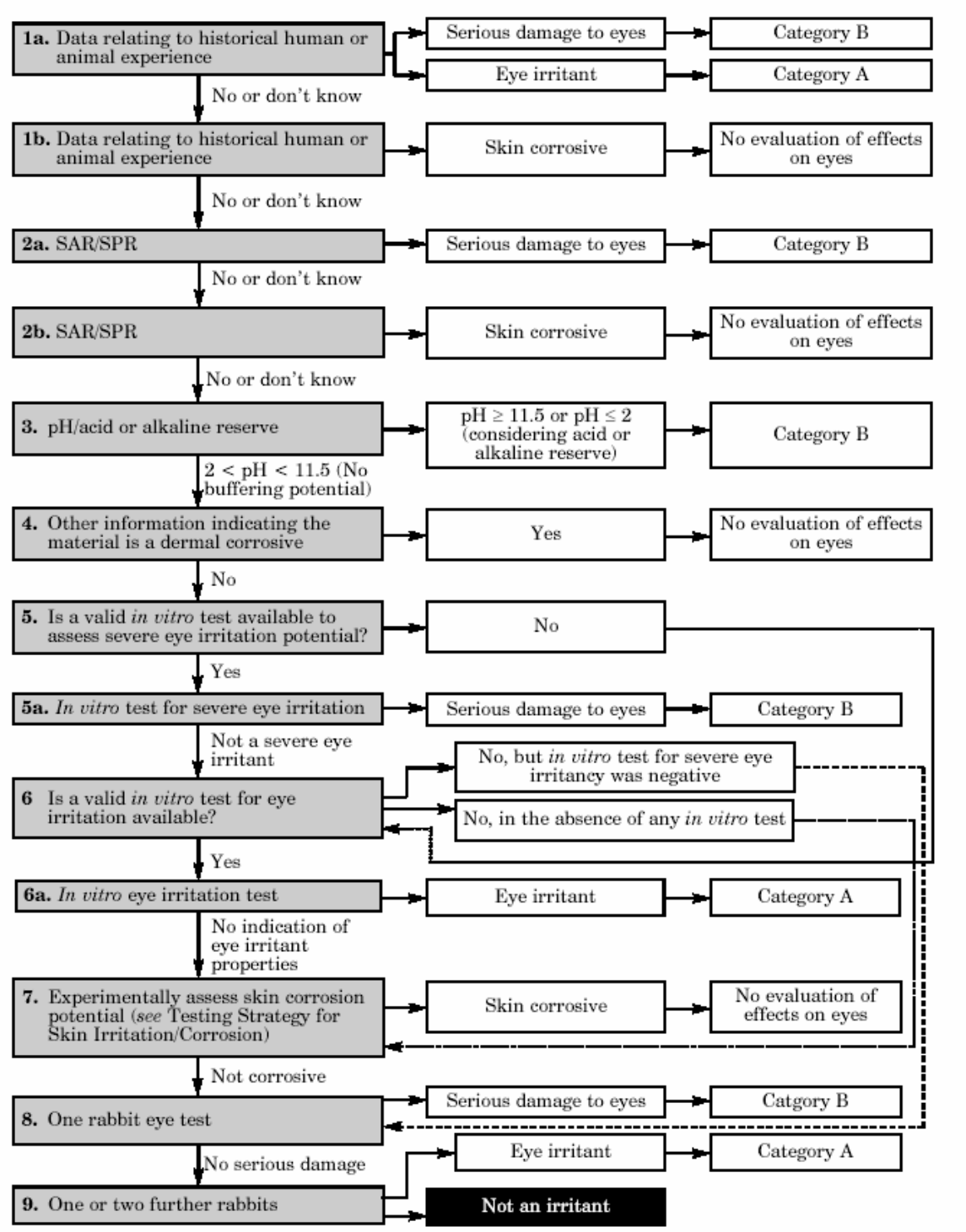
Figure 7. Tiered testing and assessment strategy for local irritation potential of chemicals (classification according to EU regulations) [33].



A subsequent paper by the same author [36] published structural alerts for the classification and labelling of eye irritation/corrosion hazards according to international classification criteria. Physicochemical limit values for prediction of the absence of any eye irritation potential relevant for human health were also published. These detailed testing and assessment strategies are included in the annex to the

current OECD TG 405 for eye irritation/corrosion testing and the GHS [35], as shown in Figure 8.

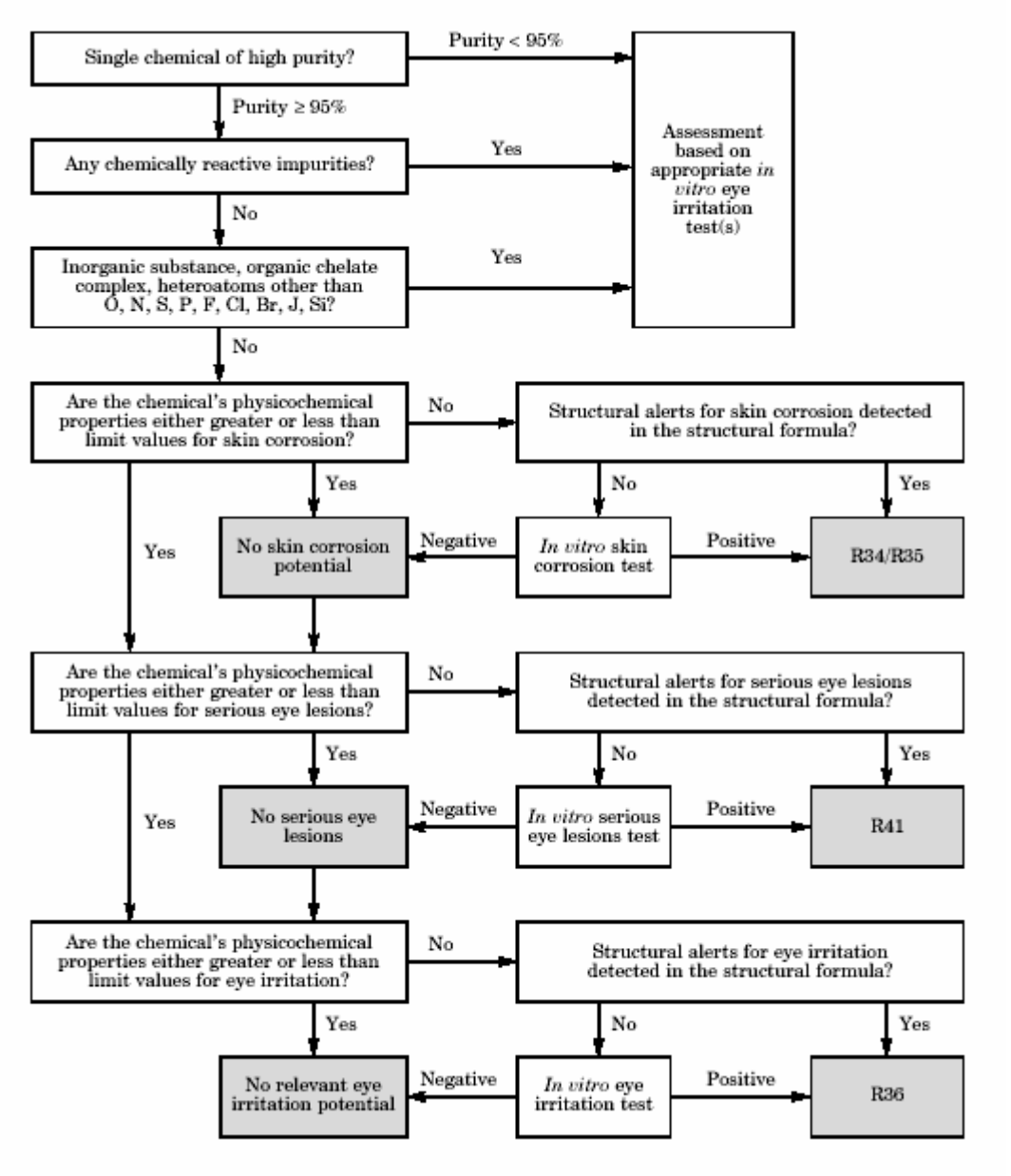
Figure 8. Testing and evaluation strategy for eye irritation/corrosion proposed by the OECD [36].



Testing and assessment strategies composed of structural alerts and *in vitro* tests to be used within the applicability domains defined by physicochemical limit

values and used to classify and label chemicals on the basis of their hazardous properties, as well as to assist in the selection of experimental test methods for the assessment of chemicals are shown in Figure 9.

Figure 9. Testing and assessment strategies for the prediction of eye irritation/corrosion by using physicochemical limit values, structural alerts and the results of specific *in vitro* tests, as implemented within computerised expert systems for the classification of eye hazards [36].



A recent paper of Walker *et al.* [37] reports the so-called Skin Irritation Corrosion Rules Estimation Tool (SICRET) that was developed to estimate whether chemicals are likely to cause skin irritation or skin corrosion. SICRET is a tiered approach that uses physicochemical property limits, structural alerts and *in vitro* tests

to identify and classify chemicals that cause skin irritation or skin corrosion without animal testing. SICRET uses physicochemical property limits to identify chemicals with no skin corrosion or skin irritation potential; if the exclusion physicochemical rules do not identify the chemicals with no skin corrosion or skin irritation potential, then the structural alerts are used to identify chemicals with skin corrosion or skin irritation potential. If a chemical does not contain structural alerts that indicate it has skin corrosion or skin irritation potential, then *in vitro* skin corrosion or skin irritation testing is evaluated. If the *in vitro* test is positive, then the data are included in feedback loops for development of new structural alerts to identify chemicals with skin corrosion or skin irritation potential. If the *in vitro* test is negative then the data are included in feedback loops for development of new physicochemical property limits to identify chemicals with no skin corrosion or skin irritation potential.

SICRET is a tiered approach that it has been proposed to complement the current OECD skin corrosion and skin irritation testing strategy as described in TG 404 for acute dermal irritation/corrosion. A significant difference between SICRET and the strategy described in the OECD TG 404 is that the former only uses information from *in vitro* tests, whereas the latter also uses *in vivo* tests information. Although the *in vitro* corrosion testing proposed by SICRET has been adopted by the OECD member states, *in vitro* skin irritation tests have not been yet validated. After the external validation of the physicochemical property limits and the structural alerts, SICRET software is planned to be coded to allow users determining whether a chemical is likely to cause either skin irritation or skin corrosion by providing a probability.

5. Proposed Integrated Testing Strategy for REACH

In the current regulation panorama, one of the objectives within the REACH Implementation Project (RIP) 3.3 was to create and test an Integrated Testing Strategy (ITS) for irritation and corrosion [38]-[40]. This strategy takes into account all data sources, including: non-testing information, *in vivo* and *in vitro* testing information, field and human data. The aim of the developed ITS is also to enable hazard assessment and classification of a chemical substance via a stepwise procedure that is cost efficient and scientifically sound whilst taking into account animal welfare concerns by reference to all existing data before considering *in vivo* testing. Earlier references dealing with the use of testing and waiving strategies in the context of REACH can be found at [41]-[43].

More recently, a series of papers on detailed suggestions for applying non-animal methods to each of the major toxicity endpoints in REACH have been published [3]-[4], [44]-[45]. Some of them review the status of alternative approaches to animal testing, and systems for the safety testing and risk assessment of chemicals [44]. Others present individual, decision-tree strategies for the eleven major toxicity endpoints of the REACH system, including human health effects and ecotoxicity [45].

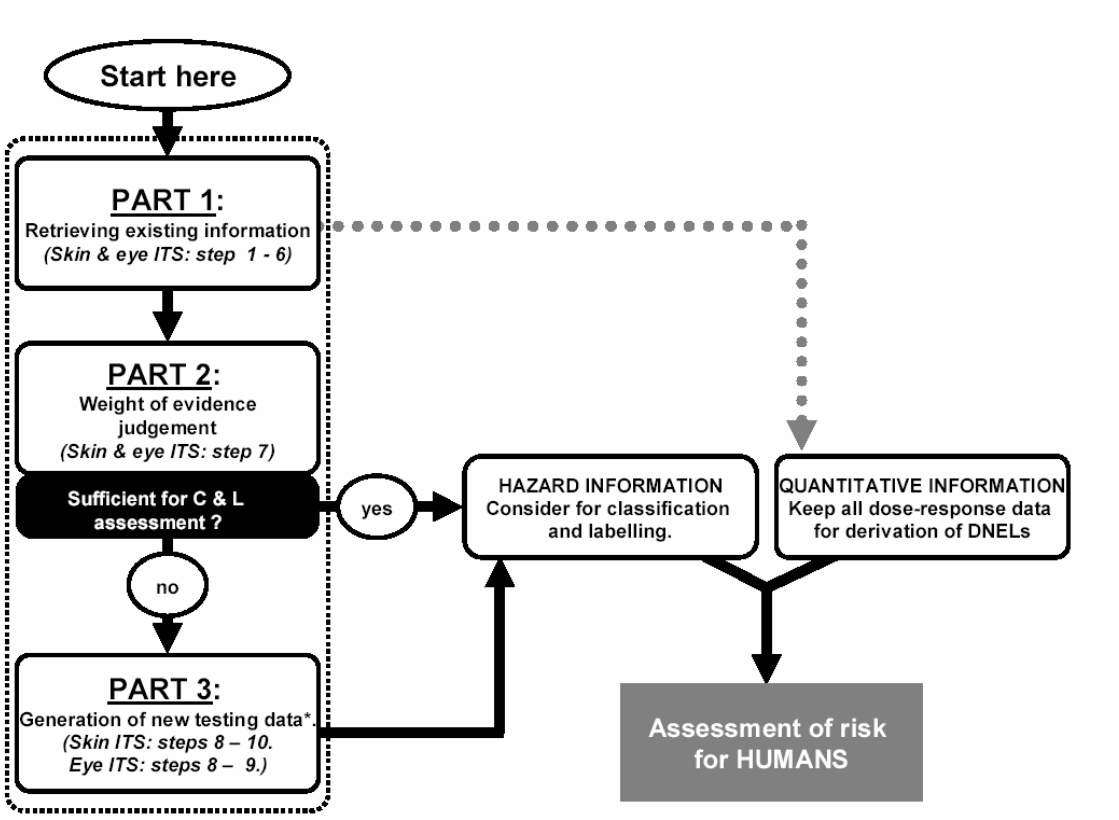
According to the REACH proposal, there are data and testing requirements for skin and eye irritation/corrosion for all substances produced in the EU or imported at levels greater that a tonne per year. Before testing, all relevant physicochemical and toxicological information e.g. acid or alkaline reactions, human and animal data, *in vitro* test data and (Q)SAR analysis, should be assessed. If these data are not available or they are inadequate for hazard and risk assessment, an *in vitro* skin corrosion study is normally required. Where the substance is corrosive in the *in vitro* study, it should be classified accordingly and no further testing for irritation conducted. However, if the substance is not corrosive in this study an *in vitro* test for skin irritation and normally an *in vitro* eye irritation study should be undertaken. If there are positive *in vitro* results from these studies the substance should be classified as being irritating to skin and eyes. When a level of 10 tonnes per year is exceeded, *in vivo* skin and eye irritation tests are normally required, unless the substance is already classified in which case the corresponding *in vivo* testing need not be done. In the scoping study of RIP 3.3, two similar sequential test strategies for skin and eye irritation/corrosion

were proposed for substances with no or very few data [39]. These ITS, similar to the sequential testing strategy proposed by B.4 [18] and B.5 [19] for skin and eye irritation/corrosion, respectively, are recommended for assessment and classification of the corrosive and irritating properties of substances. For existing substances with insufficient data, this strategy can be used to decide which additional data are needed.

A risk assessment of the irritating potential of a substance is normally made in a qualitative way when the substance has been classified as being irritating/corrosive to skin. If the substance is not classified for skin irritation/corrosion, no risk assessment for this endpoint is performed, regardless of the exposure. Therefore, classification is a key determinant in this strategy.

Both ITS for skin and eye irritation/corrosion include three parts (Figure 10) retrieval of existing information, 2) Weight of Evidence (WoE) analysis and judgement of existing data and 3) generation of new information by testing if necessary. In the information retrieval part, existing and available information from the literature and databases is gathered and considered in a stepwise process. At the end of this part all information collected is analysed using a WoE approach (step 7), which establishes whether *in vitro* or *in vivo* tests should be conducted. It is recommended that the strategy is followed to step 6 in all cases and thereafter the WoE analysis is performed. Before the WoE analysis in step 7, no new *in vitro* or *in vivo* tests should be conducted, but the assessment should be based on the existing data. In the information generation part, new information on the irritation potential of substances is created by means of testing. Prior to perform any new *in vivo* test, the use of *in vitro* methods should be fully exploited. Both the second and third parts may either lead to a decision on classification and labelling or an informed decision that there is no necessity to classify/label.

Figure 10. Overview of the ITS for dermal and ocular corrosion/irritation [40].



Detailed information and guidance on the various steps, addressing skin and eye effects separately, is provided in the RIP 3.3 scoping study report [39], and in the RIP3.3 phase 2 report [40]. ITS for assessing the skin irritation/corrosion and eye irritation potential of substances are displayed in Figure 11 and Figure 12, respectively.

Figure 11. Integrated testing strategy (ITS) for assessing the skin corrosion and skin irritation potential of substances [40].

Step	Information	Conclusion
Existing data on physical-chemical properties		
1a	Is the substance spontaneously flammable) in contact with air(pyrophoric) or water at room temperature? → ↓ NO ↓	YES: No testing required. **No need to proceed.**
1b	Is the substance an organic hydro peroxide or an organic peroxide? → ↓ NO ↓	YES: 1) Consider to classify as ■ corrosive (R34; "causes burns") if the substance is a hydro peroxide or ■ irritating as R38 ("Irritating to skin") if the substance is a peroxide. OR 2) Provide evidence for the contrary **Proceed to next step**
1c	Is the pH of the substance lower than 2 or higher than 11.5? [a] → ↓ NO ↓	YES: Consider to classify as corrosive. Where classification is based upon consideration of pH alone (see step 7!), R35 should be applied. **Proceed to next step**
1d	Are there other physical or chemical properties that indicate that the substance is irritating/corrosive? → ↓ NO ↓	YES: Use this information for WoE analysis (step 7). **Proceed to next step**
Existing human data		
2	Are there adequate existing human data[b] which provide evidence that the substance is an irritant or corrosive → ↓ NO ↓	YES: Consider to classify accordingly. **Proceed to next step**
Existing animal data from irritation/corrosivity studies		
3	Are there data from existing studies *on irritation and corrosion* in laboratory animals, which provide sound conclusive evidence that the substance is a corrosive, irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (either R35 or R34 or R38 or no classification). **Proceed to next step**
Existing data from general toxicity studies via the dermal route and from sensitization studies		
4a	Is the substance acutely toxic (LD_{50}≤400 mg/kg bw) or very toxic (LD_{50}≤50 mg/kg bw) via the dermal route? [c] → ↓ NO ↓	YES: The substance will be classified for its acute dermal toxicity. **Proceed to next step**
4b	Has the substance proven to be a corrosive, irritant or non-irritant in a suitable acute dermal toxicity test? [d] → ↓ NO ↓	YES: If test conditions are consistent with OECD 404, consider to classify accordingly (R35 or R34 or R38 or no classification). **Proceed to next step**
4c	Has the substance proven to be a corrosive or an irritant in sensitisation studies or after repeated exposure? [e] → ↓ NO ↓	YES: This information cannot be used for considering a concrete classification conclusion but must be used exclusively within the integrated WoE judgement. **Proceed to next step**
Existing (Q)SAR data and read across		
5a	Are there structurally related substances (suitable "read across" or grouping), which are classified as corrosive (R34, R35) on the skin, or do suitable QSAR methods indicate corrosion potential of the substance? [f] → ↓ NO ↓	YES: Consider to classify as R35 **Proceed to next step**

5b	Are there structurally related substances (suitable "read across" or grouping), which are classified as irritant on the skin (R38), or do suitable (Q)SAR methods indicate irritating potential of the substance? [f] → ↓ NO ↓	YES: Consider to classify as R38. **Proceed to next step**
Existing in vitro data		
6a	Has the substance demonstrated corrosive properties in an OECD adopted *in vitro* test? → ↓ NO [g] ↓	YES: Consider to classify as corrosive. If discrimination between R34 and R35 is not possible, R35 must be chosen. **Proceed to next step**
6b	Are there acceptable data from a validated *in vitro* test (adopted by OECD or not), which provide evidence that the substance is an irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (R38 or no classification). **Proceed to next step**
6c	Are there data from a non-validated *in vitro* test, which provide sound conclusive evidence that the substance is an irritant[h]? → ↓ NO ↓	Yes: Consider to classify as R38, **Proceed to next step**
Weight of evidence analysis		
7	Taking all existing and relevant data (steps 1-6) into account, is there sufficient information to make a decision of whether classification/labelling is necessary, and – if so – how to classify and label? → ↓ NO ↓	YES: Classify accordingly (R35 or R34 or R38 or no classification)
New in vitro/ex vivo tests for corrosivity (Annex VII)		
8	Does the substance demonstrate corrosive properties in an OECD adopted *in vitro* or *ex vivo* tests for skin corrosion? → ↓ NO [g] ↓	YES: Classify R34 or R35. If discrimination between R34 and R35 is not possible, R35 must be chosen.
New in vitro/ex vivo tests for irritation (Annex VII)		
9a	Does the substance demonstrate irritating or non-irritating properties in validated *in vitro* tests (adopted by OECD or not) for skin irritation? → ↓ NO ↓	YES: Classify accordingly
9b	Does the substance demonstrate irritating properties in a non-validated *in vitro* test for skin irritation[h]? → ↓ NO ↓	YES: Classify accordingly.
New in vivo test for irritation (Annex VIII) [i]		
10	Does the substance demonstrate irritancy in an OECD adopted *in vivo* test? → ↓ NO ↓ No classification	YES: Classify accordingly.

Figure 12. Integrated testing strategy (ITS) for assessing the eye irritation potential of substances [40].

Step	Information	Conclusion
Conclusion of the information strategy on skin irritation/corrosion		
0a	Is the substance classified as a skin corrosive? → ↓ NO ↓	YES: when assigned R34 or R35, the risk of severe damage to eyes is considered implicit. **No need to proceed.**
Existing data on physical-chemical properties		
1a	Is the substance spontaneously flammable in contact with air(pyrophoric) or water at room temperature? → ↓ NO ↓	YES: no testing required **No need to proceed**
1b	Is the substance an organic hydro peroxide or an organic peroxide? → ↓ NO ↓	YES: Consider to classify for ■ severe irritancy (hydro peroxide) using R41 ("risk of serious damage to eyes") or ■ for irritation (peroxide) using R36 ("irritating to eyes"). **Proceed to next step**
1c	Is the pH of the substance lower than 2 or higher than 11.5? [a] → ↓ NO ↓	YES: when assigned R35, the risk of severe damage to eyes is considered implicit. **No need to proceed**
1d	Are there other physical or chemical properties that indicate that the substance is irritating to the eye [b]? → ↓ NO ↓	YES: Use this information for WoE analysis (step 7). **Proceed to next step**
Existing human data		
2	Are there adequate existing human data [c] which provide evidence that the substance is irritating to the eye? → ↓ NO ↓	YES: Consider to classify (R41 or R36), or use for WoE analysis (step 7). **Proceed to next step**
Existing animal data from eye irritation studies		
3	Are there data from existing studies *on eye irritation* in laboratory animals, which provide sound conclusive evidence that the substance is an eye irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (R41 or R36 or no classification). **Proceed to next step**
Existing data on acute dermal toxicity		
4	Is the substance acutely toxic (LD_{50}≤400 mg/kg bw) or very toxic (LD_{50}≤50 mg/kg bw) via the dermal route? [d] → ↓ NO ↓	YES: The substance will be classified for its acute dermal toxicity. **Proceed to next step**

Existing (Q)SAR data and read across		
5	Are there structurally related substances (suitable "read across" or grouping), which are classified as irritating to the eye, or do valid QSAR methods indicate eye irritation of the substance? [e] → ↓ NO ↓	YES: Consider to classify accordingly (R41 or R36). If discrimination between R41 and R36 is not possible, R41 must be chosen. **Proceed to next step**
Existing in vitro data		
6a	Are there data from a validated *in vitro* test (adopted by OECD or not), which provide evidence that the substance is an eye irritant or non-irritant? → ↓ NO ↓	YES: Consider to classify accordingly (R36, R41 or no classification). If discrimination between R41 and R36 is not possible, R41 must be chosen. **Proceed to next step**
6b	Are there acceptable data from a non-validated *in vitro* test, which provide evidence that the substance is an irritant to the eye [f]? → ↓ NO ↓	YES: Consider to classify R41, **Proceed to next step**
Weight of evidence analysis		
7	Taking all existing and relevant data (steps 1-6) into account, is there sufficient information to make a decision of whether classification/labelling is necessary, and – if so – how to classify and label? → ↓ NO ↓	YES: Classify for accordingly (R36, R41 or no classification).
New in vitro/ex vivo tests for eye irritation (Annex VII)		
8a	Does the substance demonstrate irritating or non-irritating properties in validated *in vitro* or *ex vivo* tests (adopted by OECD or not) for eye irritation? → ↓ NO ↓	YES: Classify accordingly (R36, R41 or no classification). If discrimination between R41 and R36 is not possible, R41 must be chosen.
8b	Does the substance demonstrate severe irritating properties in acceptable non-validated *in vitro* or *ex vivo* tests for eye irritation (at present only IRE, ICE, BCOP and HET-CAM) [f]? → ↓ NO ↓	YES: Classify R41
New in vivo test for eye irritation (Annex VIII)		
9	Does the substance demonstrate irritancy in an OECD adopted *in vivo* test? → ↓ NO ↓ No classification	YES: Classify accordingly.

6. Conclusions

The further development and validation of testing strategies and *in silico* approaches is necessary. In particular, a considerable effort to evaluate and promote the use of valid (Q)SARs is being carried out by the European Chemicals Bureau (ECB) [46]-[47]. This includes the need to present information on the characteristics of the models in a transparent way. At present, QSAR Model Reporting Formats (QMRFs) have been developed for several models that predict skin/eye irritation/corrosion, for this purpose [48].

The most effective approach to build testing systems is to integrate all appropriate information to make a Weight of Evidence (WoE)-based assessment of the chemical hazard and risk [8]. Integrated Testing Strategies (ITS) combine all possible sources of information from (Q)SARs, expert systems, read-across and other grouping approaches, and test methods (especially *in vitro* tests). Considerable work have been carried out within the context of the REACH-Implementation Projects (RIP 3.3) [40] to develop further existing tiered approaches for the assessment of skin and eye irritation/corrosion potential. At the same time, the conceptual framework for integrating different components of ITS and weighing their data needs to be further investigated.

7. References

[1] Snodin, D.J. (2002). An EU perspective on the use of *in vitro* methods in regulatory pharmaceutical toxicology. *Toxicology Letters* **127**, 161-168.

[2] Pugsley, M.K. (2004) An overview of some pharmacological methods used in safety pharmacology studies. *Proceedings of the Western Pharmacology Society* **47***,* 18-22.

[3] Grindon, C., Combes, R., Cronin, M.T.D., Roberts, D.W., Garrod, J.F. (2006). Integrated Testing Strategies for Use in the EU REACH System. *Alternatives to Laboratory Animals* **34***,* 407-427.

[4] Combes, R., Barratt, M., Balls, M. (2006). An Overall Strategy for the Testing of Chemicals for Human Hazard and Risk Assessment under the EU REACH System. *Alternatives to Laboratory Animals* **34 (Suppl.1)***,* 15-27.

[5] Combes, R., Balls, M. (2005). Intelligent testing strategies for chemicals testing - a case of more haste, less speed? *Alternatives to Laboratory Animals* **33**, 289-297.

[6] Kroeger, M. (2006). How omics technologies can contribute to the '3R' principles by introducing new strategies in animal testing. *Trends in Biotechnology* **24(8)**, 343-346.

[7] Hengstler, J.G., Foth, H., Kahl, R., Kramer, P.-J., Lilienblum, W., Schulz, T., Schweinfurth, H. (2006). The REACH concept and its impact on toxicological sciences. *Toxicology* **220**, 232-239.

[8] Balls, M., Amcoff, P., Bremer, S., Casati, S., Coecke, S., Clothier, R., Combes, R., Corvi, R., Curren, R., Eskes, C., Fentem, J., Gribaldo, L., Halder, M., Hartung, T., Hoffmann, S., Schechtman, L., Scott, L., Spielmann, H., Stokes, W., Tice, R., Wagner, D., Zuang, V. (2006). The Principles of Weight of Evidence Validation of Test Methods and Testing Strategies. The Report and Recommendations of ECVAM Workshop 58. *Alternatives to Laboratory Animals* **34**, 1–19.

[9] Hansson, S.O., Rudén. C. (2007). Towards a Theory of Tiered Testing. *Regulatory Toxicology and Pharmacology* **48,** 35-44.

[10] Hoffmann, S., Hartung, T. (2006). Toward an evidence-based toxicology. *Human & experimental toxicology* **25**, 497-513.

[11] Organisation for Economic Co-operation and Development (1999). Series on Testing and Assessment No.14. *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries.* ENV/JM/MONO(99)4. Available at: http://www.olis.oecd.org/olis/1999doc.nsf/LinkTo/env-jm-mono(99)4

[12] Gerner, I., Schlede, E., Spielmann, H. (1993). Regulatory perspectives on alternatives to the Draize eye test in the European Community. *Reports of the Working Groups of the "IRAG Workshop on Alternatives to Ocular Irritation"*, Washington D.C., 1993.

[13] Organisation for Economic Co-operation and Development (1981). Guideline for Testing of Chemicals No. 405: *Acute Eye Irritation/Corrosion.* OECD, Paris, France, adopted on 12 May 1981, revised on 24 February 1987. Updated Guideline, adopted on 24 April 2002.

[14] Organisation for Economic Co-operation and Development (2001). OECD Attachment to the Test Guideline 405: A Sequential Testing Strategy for Eye Irritation and Corrosion. OECD, Paris, France.

[15] Organisation for Economic Co-operation and Development (1998). Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances, as endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, November 1998. Available at: http://www.oecd.org/ehs/Class/HCL6.htm

[16] Walker A.P. (1990). A proposal for the harmonisation of classification schemes for chemicals. Contract No. B6680/90/6065 for the Commission of the European Communities, DG XI-A-2.

[17] Organisation for Economic Co-operation and Development (1996). Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. OECD, Paris, France. Held in Solna, Sweden, 22-24 January 1996, 62pp. Available at: http://www.oecd.org/ehs/test/background.htm

[18] European Economic Community (1992). Annex V-Methods for the determination of toxicity. B.4. Acute toxicity (Skin Irritation). *Official Journal of European Communities* **L383 A 35**.

[19] European Economic Community (1992). Annex V-Methods for the determination of toxicity. B.5. Acute toxicity (Eye Irritation). *Official Journal of European Communities* **L383 A 35**.

[20] Balls, M., Berg, N., Bruner, L.H., Curren, R.D., de Silva, O., Earl, L.K., Esdaile, D,J., Fentem, J.H., Liebsch, M., Ohno, Y., Prinsen, M.K., Spielmann, H., Worth, A.P. (1999). Eye Irritation Testing: The Way Forward. The Report and Recommendations of ECVAM Workshop 34. *Alternatives to Laboratory Animals* **27**, 53-77.

[21] Botham, P., Chamberlain, M., Barratt, M. D., Curren, R.D., Esdaile, D.J., Gardner, J.R., Gordon, V.C., Hildebrand, B., Lewis, R.D., Liebsch, M., Logemann, P., Osborne, R., Ponec, M., Régnier, J.-F., Steiling, W., Walker, A.P., Balls, M. (1995). A

Prevalidation Study on *In vitro* Skin Corrosivity Testing. The Report and Recommendations of ECVAM Workshop 6. *Alternatives to Laboratory Animals* **23***,* 219-255.

[22] Basketter, D.A., Whittle, E., Chamberlain, M. (1994). Identification of irritation and corrosion hazards to skin, an alternative strategy to animal testing. *Food and Chemical Toxicology* **32**, 539-542.

[23] Lewis, R.W., McCall, J.C., Botham, P.A. (1994). Use of an *in vitro* test battery as a prescreen in the assessment of ocular irritancy. *Toxicology in Vitro* **8**, 75-79.

[24] Spielmann, H., Liebsch, M., Kalweit, S., Moldenhauer, F., Wirnsberger, T., Holzhutter, H-G., Schneider, B., Glaser, S., Gerner, I., Pape, W.J.W., Kreiling, R., Krauser, K. & Miltenburger, H.G., Steiling, W., Luepke, N.P., Muller, N., Kreuzer, H., Murmann, P., Spengler, J., Bertram-Neis, E., Siegemund, B. & Wiebel, F.J. (1996). Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, the MET-CAM test and the 3T3-NRU cytotoxicity test. *Alternatives to Laboratory Animals* **24**, 741-858.

[25] de Silva, O., Cottin, M., Dami, N., Roguet, R., Catroux, P., Toufic, A., Sicard, C., Dossou, K.G., Gerner, l., Schlede, E., Spielmann, H., Gupta, K.C., Hill, R.N. (1997) .Evaluation of eye irritation potential: statistical analysis and tier testing strategies. *Food and Chemical Toxicology* **35**, 159-164.

[26] Worth, A.P., Fentem, J., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Liebsch, M. (1998). An Evaluation of the Proposed OECD Testing Strategy for Skin Corrosion. *Alternatives to Laboratory Animals* **26**, 709-720.

[27] Worth, A.P., Fentem, J.H. (1999). A General Approach for Evaluating Stepwise Testing Strategies. *Alternatives to Laboratory Animals* **27**, 161-177.

[28] Worth, A.P., Cronin, M.T.D. (2001). The Use of Bootstrap Resampling to Assess the Uncertainty of Cooper Statistics. *Alternatives to Laboratory Animals* **29**, 447-459.

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

[30] Worth, A.P., Cronin, M.T.D. (2001). The use of pH measurements to predict the potential of chemicals to cause acute dermal and ocular toxicity. *Toxicology* **169**, 119-131.

[31] Worth, A.P., Cronin, M.T.D. (2000). Embedded cluster modelling: a novel quantitative structure-activity relationship for generating elliptic models of biological activity. In, *Progress in the Reduction, Refinement and Replacement of Animal*

Experimentation. Balls, M., van Zeller, A.-M., Halder, M.E. (Eds.) Elsevier Science, Amsterdam, The Netherlands, pp. 479-491.

[32] Worth, A.P. (2000). The integrated use of physicochemical and *in vitro* data for predicting chemical toxicity. PhD thesis. Liverpool John Moores University, Liverpool, UK.

[33] Gerner, I., Schlede, E. (2002). Introduction of *in vitro* data into local irritation/corrosion testing strategies by means of SAR considerations: assessment of chemicals. *Toxicology Letters* **127**, 169-175.

[34] Commission of the European Communities (2000). Directive 2000/33/EC of 25 April 2000 Adapting to Technical Progress for the 27th time the Council Directive 67/548/EEC on the Approximation of Laws, Regulations and Administrative Provisions Relating to the Classification, Packaging and Labelling of Dangerous Substances. Annex 1. B.40. Skin corrosion. *Official Journal of European Communities* **L 136**.

[35] Organisation for Economic Co-operation and Development (1998). Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures, ENV/JM/HCL(2001)1, OECD Paris, France. Available at: http://www.oecd.org/ehs/Class/hcl/finaw.pdf

[36] Gerner, I., Liebsch, M., Spielmann, H. (2005). Assessment of the Eye Irritating Properties of Chemicals by Applying Alternatives to the Draize Rabbit Eye Test: The Use of QSARs and *In vitro* Tests for the Classification of Eye Irritation. *Alternatives to Laboratory Animals* **33**, 215-237.

[37] Walker, J.D., Gerner, I., Hulzebos, E., Schlegel, K. (2005). The Skin Irritation Corrosion Rules Estimation Tool (SICRET). *QSAR & Combinatorial Science* **24**, 378-384.

[38] ECB (2005). Three point three – A Project for the Information Requirements of REACH. Lot 3: Scoping study on the Development of a Technical Guidance Document on Information requirements on Intrinsic Properties of Substances (RIP 3.3-1). Final TAPIR Report RIP 3.3-1. Submitted to European Chemical Bureau, Joint Research Centre, European Commission under Contract No. – 22553-2004-12 F1SC ISP BE. Available at: http://ecb.jrc.it/reach/rip/

[39] ECB (2005). TAPIR Final report appendix 12. Report of Endpoint Working Group 4 "Irritation / Corrosion". REACH Implementation Project 3.3-1.

[40] RIP 3.3-EWG2 Draft Report for the Human Health Endpoints of Skin Corrosion / Irritation, Eye Irritation, Respiratory Irritation. Version 4, 2 February 2007.

[41] Sauer, U.G. (2001). White paper on the future chemical policy in the European Union - implications for animal welfare. *Alternativen zu Tierexperimenten* **18(4)**, 281-284.

[42] Sauer, U. (2002). The new EU chemicals policy-Discussions on details relevant for animal welfare. *Alternativen zu Tierexperimenten* **19 Suppl 1**, 26-29.

[43] Sauer, U.G. (2004). The new EU chemicals policy: Challenges and chances for animal-free test methods. *Alternatives to Laboratory Animals* **32 (Suppl. 1A)**, 343-348.

[44] Grindon, C., Combes, R., Cronin, M.T.D., Roberts, D.W., Garrod, J. (2006). A review of the status of alternative approaches to animal testing and the development of integrated testing strategies for assessing the toxicity of chemicals under REACH - a summary of a DEFRA-funded project conducted by Liverpool John Moores University and FRAME. *Alternatives to Laboratory Animals* **34 (Suppl. 1)**, 149-158.

[45] Grindon, C., Combes, R., Cronin, M.T.D., Roberts, D.W., Garrod, J. (2006). Integrated decision-tree testing strategies for environmental toxicity with respect to the requirements of the EU REACH legislation. *ATLA, Alternatives to Laboratory Animals* **34(6)**, 651-64.

[46] Organisation for Economic Co-operation and Development (2004). *The Report From The Expert Group On (Quantitative) Structure-Activity Relationships [(Q)Sars] On The Principles For The Validation Of (Q)SARs*. 2nd Meeting of the ad hoc Expert Group on QSARs, OECD Headquarters, Paris, 20-21 September, 2004.

[47] Worth, A.P., Bassan, A., Gallegos, A., Netzeva, T.I., Patlewicz, G., Pavan, M., Tsakovska, I., Vracko, M. (2005). The Characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance. JRC Technical Report. EUR 21866 EN, 95pp. Ispra, Italy: European Commission, Joint Research Centre. Available at: http://ecb.jrc.it/qsar/publications/

[48] QSAR Reporting Formats can be found in the webpage: http://www.ecb.jrc.it/QSAR/

European Commission

EUR 22881 EN – DG Joint Research Centre, IHCP
Title: Testing Strategies for the Prediction of Skin and Eye Irritation and Corrosion for Regulatory Purposes
Authors: Ana Gallegos Saliner & Andrew P Worth
Luxembourg: Office for Official Publications of the European Communities
2007 – 39 pp. – 21 x 29.7 cm
EUR - Scientific and Technical Research series; ISSN 1018-5593

Abstract

This report reviews the use of stepwise testing approaches for the prediction of skin and eye irritation and corrosion in a regulatory context. It is published as a companion report to the *Review of Literature-Based Models for Skin and Eye Irritation and Corrosion*, an ECB report which reviewed the state-of-the-art of in silico and in vitro dermal and ocular irritation and corrosion human health hazard endpoints. In the former review, the focus was placed on reviewing alternative *in silico* approaches to assess acute local toxic effects, such as QSARs, SARs, chemical categories, and read-across and analogue approaches. Special emphasis was placed on literature-based (Q)SAR models for skin and eye irritation and corrosion and expert systems. In the present review, the emphasis is on different schemes (testing strategies) that have been conceived for the integrated use of different approaches, including *in silico*, *in vitro* and *in vivo* methods.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

Publications Office
Publications.eu.int