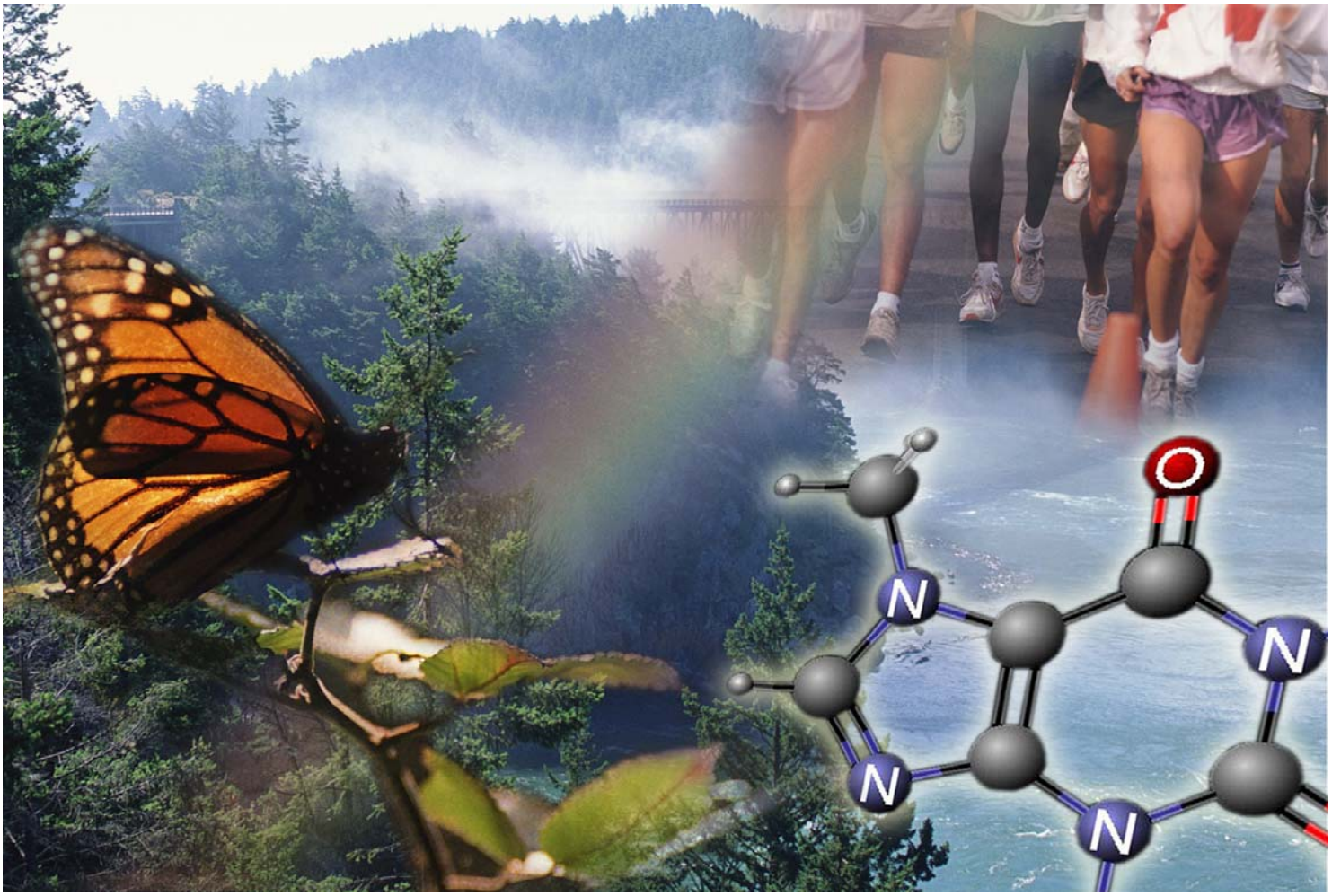
Guidance on information requirements and chemical safety assessment

Chapter R.4: Evaluation of available information



May 2008

LEGAL NOTICE

This document contains guidance on REACH explaining the REACH obligations and how to fulfil them. However, users are reminded that the text of the REACH regulation is the only authentic legal reference and that the information in this document does not constitute legal advice. The European Chemicals Agency does not accept any liability with regard to the contents of this document.

© European Chemicals Agency, 2008
Reproduction is authorised provided the source is acknowledged.

PREFACE

This document describes the information requirements under REACH with regard to substance properties, exposure, use and risk management measures, and the chemical safety assessment. It is part of a series of guidance documents that are aimed to help all stakeholders with their preparation for fulfilling their obligations under the REACH regulation. These documents cover detailed guidance for a range of essential REACH processes as well as for some specific scientific and/or technical methods that industry or authorities need to make use of under REACH.

The guidance documents were drafted and discussed within the REACH Implementation Projects (RIPs) led by the European Commission services, involving stakeholders from Member States, industry and non-governmental organisations. These guidance documents can be obtained via the website of the European Chemicals Agency (http://echa.europa.eu/reach_en.asp). Further guidance documents will be published on this website when they are finalised or updated.

This document relates to the REACH Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006[1]

1 Corrigendum to Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC (OJ L 396, 30.12.2006); amended by Council Regulation (EC) No 1354/2007 of 15 November 2007 adapting Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) by reason of the accession of Bulgaria and Romania (OJ L 304, 22.11.2007, p. 1).

Convention for citing the REACH regulation

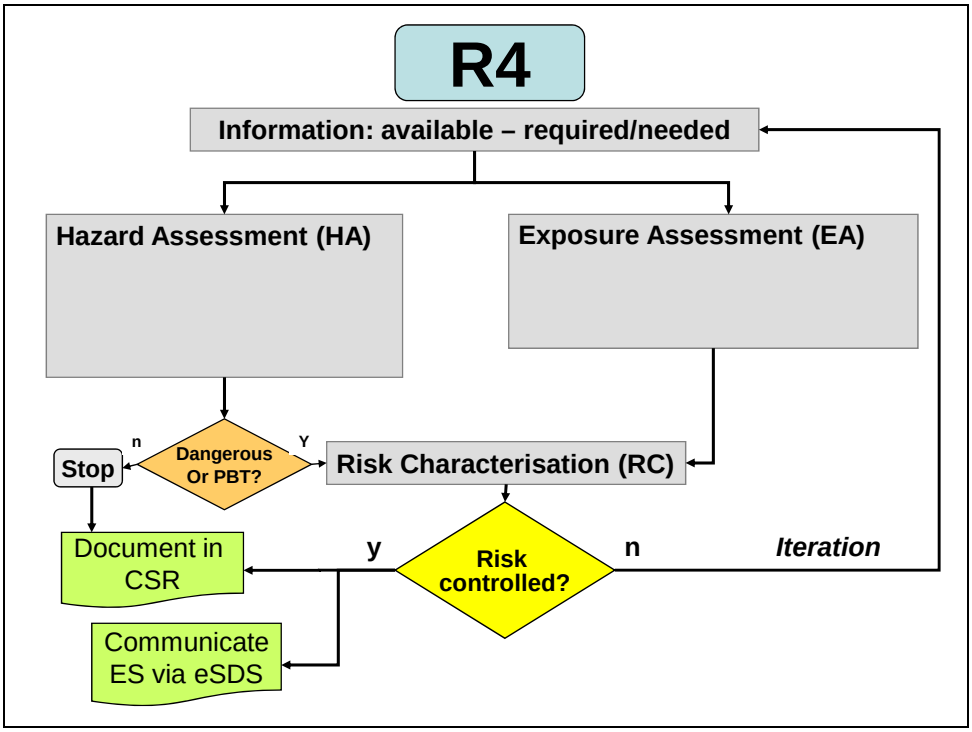
Where the REACH regulation is cited literally, this is indicated by text in italics between quotes.

Table of Terms and Abbreviations

See Chapter R.20

Pathfinder

The figure below indicates the location of Chapter R.4 within the Guidance Document



CONTENTS

R.4 EVALUATION OF AVAILABLE INFORMATION 7

R.4.1 Relevance of information 8

R.4.2 Reliability of information 9

R.4.3 Adequacy of information 11
R.4.3.1 Non-human data 11
R.4.3.1.1 In vitro data 12
R.4.3.2 Non-testing data 16
R.4.3.2.1 (Q)SAR data 16
R.4.3.2.2 Data obtained by grouping approaches 18
R.4.3.3 Human data 20

R.4.4 Evaluation and Integration of all available Information including Weight of Evidence 21

R.4.5 References 22

TABLES

Table R.4-1: The criteria for validation derived from the OECD GD 34 13
Table R.4-2: The criteria for suitability assessment according to the ECVAM criteria for entering the pre-validation study, (Curren *et al*, 1995). 15

R.4 EVALUATION OF AVAILABLE INFORMATION

This chapter aims to provide guidance on how to evaluate available information gathered in the context of REACH Annex VI-XI. The information should be evaluated for its *completeness* and *quality* for the purpose of REACH to assess whether:

1. it fulfils the specific requirements triggered by tonnage as described in REACH Annex VII-X, including application of REACH Annex XI.
2. it is appropriate for hazard classification and risk assessment, including CMR, PBT and vPvB assessment.

Practically, this assessment is usually performed by an evidence-based approach to determine whether the information requirements are already met by the available information. If this is not the case, the information gaps should be defined and appropriate action(s) taken to address these.

The evaluation of *data quality* includes assessment of *adequacy* of the information for hazard/risk assessment and C&L purposes (see above) and furthermore the two basic elements of *relevance* and *reliability*. These terms were defined by Klimisch *et al* (1997) as follows (see also OECD, 2005a):

> *"**Relevance** - covering the extent to which data and tests are appropriate for a particular hazard identification or risk characterisation.*
>
> ***Reliability** - evaluating the inherent quality of a test report or publication relating to preferably standardised methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings. Reliability of data is closely linked to the reliability of the test method used to generate the data (see Section R.4.2).*
>
> ***Adequacy** - defining the usefulness of data for hazard/risk assessment purposes. Where there is more than one study for each endpoint, 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."*

The terms *relevance* and *reliability* are also used in the context of test methodology (see OECD GD 34 (OECD, 2005b)). The knowledge of how a study was carried out and consequently its relevance and reliability, is a prerequisite for the subsequent evaluation of information.

The *completeness* of the information refers to the conclusion on the comparison between the available information and the information that is required under the REACH registered for the tonnage level of the substance.

Available information on the individual substance should be evaluated in relation to the level of certainty and accuracy needed to meet the regulatory requirement under REACH; it should be considered whether generation of new data would impact such regulatory decision making. In other words, all information has to be *adequate for the purpose*.

A *Weight of Evidence* approach, mentioned in Annex XI Section1.2 of REACH, integrates available information from guideline tests, non-guideline tests, and other types of information which may justify adaptation of the standard testing regime.

R.4.1 Relevance of information

In order to evaluate the relevance of the available data the following aspects could for example be considered:

- Was the substance tested representative for the substance as being registered?
- Has the appropriate species been studied?
- Is the route of exposure relevant for the population?
- Were appropriate doses/concentrations tested?
- Were the critical parameters influencing the endpoint considered adequately?

Human data is in principle the most relevant source of information on human toxicity. Since there may be limitations with regard to the reliability of these studies, they are normally considered together with animal, *in vitro* and other information in order to be able to reach a conclusion about the relevance of the effects to humans.

The evaluation of the relevance for humans of data from studies in laboratory animals is aided by use of information (when available) on the toxicokinetics of the substance in both humans and the animals species used in the toxicity tests, even when such information is relatively limited. Further guidance on the value and use of toxicokinetics is given in Section R.7.12.

Normally, for human health hazard assessment, a *no* or *lowest observed (adverse) effect level* (NO(A)EL, LO(A)EL), or a *benchmark dose* (BMD) for adverse effects in laboratory animals are extrapolated to an exposure level (DNEL) below which it is assumed that adverse effects are unlikely to occur in humans exposed to the substance. For substances evoking effects that have no definable threshold, e.g. genotoxic carcinogens, it may not be possible to identify an exposure level without effects; in such cases, extrapolation is made to an exposure level that represents a risk level of very low concern for humans (DMEL). For more guidance on the derivation and application of these indices in the chemical safety assessment, see Chapter R.8.

For environmental compartments such as surface water, sediment and soil, a *predicted no effect concentration* (PNEC) is obtained by extrapolation based on the lowest *no observed effect concentration* (NOEC) or *effect concentration* causing marginal effects (ECx) by application of assessment factors. For more guidance on the derivation and application of these indices in the chemical safety assessment, see Chapter R.10.

When data are available, dose-response relationships in the animal studies (or the severity of the effect, when only a single dose has been tested) are also assessed as a part of the risk assessment process. Both aspects are taken into account at the risk characterisation stage when a judgement is made of whether adverse effects in humans or the environment would occur at a particular level of exposure.

Where the data suggest that an effect might be species specific, i.e. that the effects observed in the studies of one species are not likely to occur in a different species, specifically humans, clear, well-documented evidence is necessary (e.g. light hydrocarbon-induced nephropathy in the kidney of male rats) to justify the conclusion that a particular effect is not expected to occur in humans exposed to the substance.

In general, the results of *in vitro* tests provide supplementary information which may be used *inter alia* to facilitate the interpretation of the relevance of animal data for humans, or to gain a better understanding of the mechanism of action of a substance. Depending on the type of *in vitro* data and its predictivity for effects *in vivo*, such data may be also used as an alternative to test data on laboratory animals or as an important part of the basis for deciding whether such tests may be warranted.

R.4.2 Reliability of information

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 substance properly (in terms of purity, physical characteristics, etc.), the use of crude techniques/procedures that have since become refined, and the fact that certain endpoint information, now recognised as being important, may have not been recorded or measured. Moreover, other reasons could be poor reporting of information and poor quality assurance.

Klimisch *et al* (1997) developed a scoring system to assess the reliability of data, particularly from toxicological and ecotoxicological studies, that may be extended to physico-chemical and environmental fate and behaviour studies:

> **1 = reliable without restrictions:** *"studies or data [...] generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline [...] or in which all parameters described are closely related/comparable to a guideline method."*
>
> **2 = reliable with restrictions:** *"studies or data [...] (mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable."*
>
> **3 = not reliable:** *"studies or data [...] in which there were interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g. unphysiological pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for assessment and which is not convincing for an expert judgment."*
>
> **4 = not assignable:** *"studies or data [...] which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.)."*

The use of such scoring tools, e.g. the mentioned *Klimisch codes,* allows ranking the information, and organising it for further review. This implies focussing on the most relevant ones, taking into account the endpoint being measured or estimated. The evaluation of the reliability is performed considering certain formal criteria using international standards as references. The scoring of information, e.g. according to Klimisch codes, should not exclude all unreliable data from further consideration by expert judgement because of possible pertinence of these data related to the evaluated endpoints.

In general, some types of data that are not reliable (i.e. those where insufficient documentation exist for making an assessment) and data for which it is not possible to assign reliability, may only be used as supporting data.

For many existing substances, at least some of the available information could have been generated prior to the requirements of GLP and the standardisation of testing methods. While such information may still be usable for REACH purposes, both the data and the methodology used must be evaluated in order to determine their reliability. Such an evaluation needs evidence based decision making following established criteria and must be transparent to justify the use of a particular data set. For some substances, information may be available from tests conducted according to the methods included in EU Annex V of Directive 67/548/EEC or 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. REACH Article 13.3 states that any new tests should be "*conducted in accordance with the test methods laid down in a Commission Regulation or… other international test methods recognised by the Commission or the Agency as being appropriate…. Information on intrinsic properties may also be generated using other test methods provided they meet the conditions set out in Annex XI.*"

Furthermore, new ecotoxicological and toxicological tests shall be carried out in compliance with the principles of GLP (see Directive 2004/10/EC) or equivalent international standards. This does not apply to tests for physico-chemical properties.

The following are key points that an assessor should consider when evaluating data reliability:

- The proven ability of the laboratory to perform the test method
- The purity/impurities and origin of the test substance, as well as the reference substances, must be reported;
- The availability of the raw data from the study
- There must be an adequate description of the study e.g. a complete test report, or a sufficiently detailed description of the test procedure, which must be in accordance with generally accepted scientific standards. In these cases, the information may be considered reliable;
- When the test procedure used to generate the test data is found to differ significantly from that described by the recognised test method or generally accepted scientific standards, or the reliability of the data cannot be established fully, the assessor must decide if and how the information can be used, e.g. as supporting information where a reliable study already exists.

- The following factors, inter alia, can be used to support the view that these data may be acceptable for use in meeting the requirements of REACH:
  - there are other studies or calculations available on the substance, and the data under consideration are consistent with them,
  - other studies are available, for example on isomers with similar structure activity profile, homologues, relevant precursors, breakdown products or other chemical analogues, and the data under consideration are consistent with them,
  - an approximate value is sufficient for taking a decision on the endpoint of interest for the conclusion required by REACH;
- Where critical supporting information is not reported (e.g. species tested, substance identity and dosing procedure) the test data should be considered to be unreliable for the purposes of REACH.

In principle, the same criteria apply to test data reported in the published literature; the extent of the information provided will provide the basis for deciding upon the reliability of the data reported. In general, publications in peer-reviewed journals are preferable to those which are not. High-quality reviews, summaries or abstract publications may be used as supporting information.

R.4.3 Adequacy of information

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 by the registrant about (a) whether the substance meets the criteria for classification, (b) whether it is a potential PBT/vPvB and (c) whether appropriate DNEL/PNEC values can be derived for risk assessment purposes. The evaluation of adequacy of test results and documentation for the intended purpose is particularly important for substances under REACH where there may be (a number of) test results available for each effect, but where some or all of them have not been carried out according to current standards. Where there is more than one study for each endpoint, 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.

The type of information that may be available consists of non-testing data, (the latter refer to (Q)SAR predictions or data on structurally-related substances, obtained by grouping approaches), *in vitro* data, data on living organisms, including data on laboratory animals, on humans or other data on (parts of) ecosystems.

R.4.3.1 Non-human data

The guidance given above on the evaluation of the adequacy (relevance and reliability) of information relates predominantly to information generated in tests on physico-chemical properties, animal studies, plant and micro-organism studies. Some specific guidance is given below for data generated *in vitro* systems.

R.4.3.1.1 In vitro data

When considering the adequacy of *in vitro* information it is important to distinguish between the suitability of the methodology per se and the adequacy of data that have been produced by such methods.

Use of *in vitro* methods within REACH

Suitable in vitro test methods are at least those that are sufficiently well developed according to internationally agreed test development criteria, e.g. fulfilling the ECVAM criteria for entry of the method into the pre-validation process (see details in Table R.4-2). In the frame of the Community Action Plan on the Protection and Welfare of Animals, a reference laboratory (CORRELATE) has been established at JRC-IHCP-ECVAM that assesses proposed *in vitro* test methods with regard to suitability and validation for the intended purpose.

At present the following two categories of *in vitro* methods are referred to within REACH as suitable:

- validated methods (e.g. *in vitro* tests for skin corrosion and *in vitro* genotoxicity tests, e.g. Ames salmonella typhimurium mutagenicity test) and
- those *in vitro* tests that meet the internationally agreed pre-validation criteria (e.g. meeting the ECVAM criteria of entering the pre-validation process).

There are clear definitions on what constitutes a fully validated *in vitro* assay. These criteria are detailed in OECD GD 34 (OECD, 2005b; see details in Table R.4-1) and were initially established by ECVAM and ECB and later refined by ECVAM (Hartung *et al*, 2004).

Use of adequate information derived from *in vitro* methods

Adequate information from *in vitro* studies can be used in two ways: first, the existing information from a validated and accepted *in vitro* test may fully or partly replace animal testing, and second, information derived from suitable *in vitro* methods can be used for adapting the standard testing regime as set out in Annex XI.

Information from validated in vitro tests may fully or partly replace an animal test

Article 25 (1) of the REACH Regulation states that testing on vertebrate animals shall only be performed as a last resort. Once scientifically validated according to internationally agreed validation principles (OECD GD 34 (OECD; 2005b)) *in vitro* test may fully or partly replace an *in vivo* test depending on the purpose for which the test method was validated and adopted. One of the main criteria for acceptance is the adequacy of the information generated using such a test(s) for the purpose of classification and labelling and/or risk assessment.

Information derived from suitable in vitro methods

Annex XI Section 1.4 opens the way for the use of results of *in vitro* methods that have not yet been scientifically validated but are identified as being *suitable*, meaning that the methods are sufficiently well developed according to internationally agreed test development criteria e.g. the ECVAM criteria for entry of the method into the pre-validation process (see Table R.4-2 and Section R.5.2.1.4 for a discussion of the use of *in vitro* testing to adapt the standard testing regime).

Table R.4-1: The criteria for validation derived from the OECD GD 34

Concerned items	Decision criteria to be considered
Rationale for the test method	Clear statement of: - scientific basis - regulatory purpose - need for the test method
Relationship between the test method's endpoint and (biological) phenomenon of interest	Description of the scientific relevance of the measured effects Mechanistic (biological) or empirical (correlative) relationship to the specific type of effect or toxicity of interest
Detailed protocol for the test method	Detailed protocol and SOP including: - description of materials - what is measured - how it is measured - how data will be analysed - decision criteria for evaluation - criteria for acceptable test performance
Test method performance using reference substances (accuracy assessment)	Sufficient number of reference substances measured in coded procedure Reference data and reference results for reference substances established
Performance evaluation	Performance evaluation in relation to: - relevant information from the species of concern - existing relevant toxicity testing data
Intra- and Inter-laboratory reproducibility	Data available on - Repeatability and reproducibility - Robustness (variability)
Relevance	- Demonstration of the predictive capacity of the method - Precise definition of the applicability domain
Test method data quality	Evidence that all data supporting the validity are gained under quality conditions, e.g. GLP, GCCP
Data availability	- All raw data should be available for expert review - Detailed method protocol public available

Information from in vitro test may provide mechanistic insight

Information from advanced *in vitro* assays may provide valuable information that aid and inform the risk assessment process. For example, with the growth of new technologies such as toxicogenomics, new possibilities are emerging that allow designer cell lines to assess specific mode of action (molecular pathways) of the potential toxicity of a substance or substance class. Such information is likely to be increasingly important in the future.

Adequacy of information from in vitro testing

The assessment of alternative testing data (to decide whether and how they can be used) in terms of adequacy for fulfilling the information requirements of REACH will follow the general criteria already discussed, e.g. applied quality measures, i.e. how they take into account the relevance, reliability and completeness of the information with regard to the regulatory decision to be taken. This includes how well the study is reported, how well the test substance is characterised and to what extent the information requirements have been met for the endpoint under consideration.

Table R.4-2: The criteria for suitability assessment according to the ECVAM criteria for entering the pre-validation study, (Curren *et al*, 1995).

Concerned items	**Decision criteria to be considered**
Purpose and proposed use	- Description of intended purpose and scientific basis - Fit of intended purpose with intended use - Position of the method in the context of regulatory testing and/or 3Rs
Evidence of the need for the test in comparison with other *in vivo/in vitro* test, state of the art	Complete and concise presentation of state of the art, human data, *in vivo*, non-testing and *in vitro* data Weighed judgment about the contribution of the proposed test method compared to state of the art, including weaknesses and limitations e.g.: improved reliability: accuracy, sensitivity, specificity, robustness, defined performance e.g. improved relevance: predictive capacity, applicability domain
Addressed endpoint described	- Demonstration of relevance for the *in vivo* situation - Description of data analysis and interpretation
Availability of a written procedure detailed enough to allow performance in another laboratory	Method protocol: - complete and readable - feasible and transferable - SOP standardised with respect to selected model and measurement performance
Reference substances, test materials and related results	- Description of reference substances, test materials and controls - Selection, identity, use in the measurement process including calibration and data interpretation
Data derived from the test using an appropriate set of test materials	- Data gained by measuring above reference substances or test materials - Test performance evaluation
Development of method according to GLP and GCCP conditions	Statement about data quality
Summary of how method has been derived and the biological basis for its relevance	- List of any additional documentation, which contributes to the above items - Statement about intellectual property rights and search for existence of any protection of intellectual property rights

R.4.3.2 Non-testing data

Non-testing data refers to data obtained by applying computational methods, such as SARs and QSARs (collectively referred to as (Q)SARs) as well as data obtained by grouping approaches (analogue and chemical category approaches).

R.4.3.2.1 (Q)SAR data

According to Article 13 (1) of REACH, information on intrinsic properties of substances may be generated by means other than tests, provided that the conditions set out in Annex XI are met. In particular for human toxicity, information shall be generated whenever possible by means other than vertebrate animal tests, through the use of alternative methods, for example, *in vitro* methods or qualitative or quantitative structure-activity relationship models or from information from structurally related substances (grouping or read-across) [see also REACH Article 25 (1)].

REACH Annex XI allows for the results of (Q)SARs to be used instead of testing when the following conditions are met:

- results are derived from a (Q)SAR model whose scientific validity has been established,
- the substance falls within the applicability domain of the (Q)SAR model,
- results are adequate for the purpose of classification and labelling and/or risk assessment, and,
- adequate and reliable documentation of the applied method is provided.

REACH Annex XI also indicate that the Agency in collaboration with the Commission, Member States and interested parties shall develop and provide guidance in assessing which (Q)SARs will meet these conditions and provide examples. In the meantime, a database has been developed to provide information on QSAR models and their validity (JRC QSAR Model Database (QMDB) http://qsardb.jrc.it). In addition to replacing the need for testing, (Q)SAR results may also in some cases indicate the need for further testing.

To apply the conditions of REACH Annex XI, it is important to distinguish between the validity of the (Q)SAR model, and the reliability and adequacy of an individual (Q)SAR estimate, and the appropriateness of the documentation associated with models and their predictions (see Section R.6.1 for detailed explanation).

The extent to which valid (Q)SARs are available for the different REACH endpoints is variable and is an evolving situation, as an increasing number of models are being characterised and documented according to the OECD validation principles described below. Information on the status of (Q)SARs for specific endpoints is given in Chapter R.7.

Valid (Q)SARs should be assessed for their applicability to the substance of interest, to determine the reliability of the QSAR estimate, and for their relevance to the regulatory purpose, to determine the adequacy of the (Q)SAR estimate. The adequacy of a (Q)SAR estimate (see Section R.6.1.5.4) takes into account the relevance and reliability of the (Q)SAR model and its prediction for the substance of interest as well as completeness of the information generated by the model.

A valid (Q)SAR is a model that has been characterised and documented according to the internationally agreed OECD Principles for the validation of (Q)SAR models. According to these principles, a (Q)SAR model that is proposed for regulatory use should be associated with a defined endpoint (principle 1), an unambiguous algorithm to ensure transparency in the model algorithm (principle 2), a defined domain of applicability (principle 3), and appropriate measures of internal performance and predictivity (principle 4). If possible, a mechanistic interpretation should also be provided, to add to the confidence in the model (principle 5).

Taken together, these five principles form the basis of a conceptual framework for characterising (Q)SAR models.

Preliminary guidance on how to characterise (Q)SARs according to the OECD validation principles is provided in this document (see Section R.6.1) This report was subsequently adopted, with minor revisions, by the OECD Member Countries and the Commission, as an OECD GD (OECD, 2007).

Whether the prediction from a scientifically valid QSAR model is reliable depends, *inter alia*, on whether the substance is within the applicability domain (see also Section R.6.1.5.3). Consideration of the applicability domain may include: 1) descriptor domain - do the descriptor values of the chemical fall within defined ranges; 2) structural fragment domain - does the chemical contain fragments that are not represented in the model training set; 3) mechanistic domain - does the chemical of interest act according to the same mode or mechanism of action as other chemicals for which the model is applicable; and 4) metabolic domain - does the chemical of interest undergo transformation or metabolism, and how does this affect reliance on the prediction for the parent compound.

The QSAR Model Reporting Format (QMRF) has been developed to provide a means of documenting (Q)SAR model characteristics in a transparent and consistent manner, in accordance with the OECD validation principles. Further information on QMRFs is given in Section R.6.1.9. In particular, the JRC QSAR Model Database (JRC QMDB) is being developed as a repository of quality-reviewed information on QSAR models and their validity. In this database, QSAR models will be linked with their corresponding QMRFs. Before developing a QMRF, the registrant should check whether it is already included in the JRC QMDB or other suitable source (e.g. OECD QSAR Toolbox[2]). If the appropriate QMRF for a given model is not already available, it will be necessary to develop one by applying the five validation principles and documenting the results. Since the general format of the QMRF is already defined, it is sufficient to fill this in with the appropriate information on the model. The ECB has developed a QMRF editor as a tool to facilitate the generation of new QMRFs.

To be used as a replacement for experimental data, it is necessary, but not sufficient, for a (Q)SAR model to be valid. The (Q)SAR model should also be shown to be applicable to the substance of interest, to determine whether the model estimate is reliable for the intended purpose. Whereas the (Q)SAR model should be reported in the form of a QMRF, individual model predictions should be documented according to the (Q)SAR Prediction Reporting Format (QPRF). Further information on QPRFs is given in Section R.6.1.10, and in the JRC QMDB.

QMRFs and QPRFs are important tools for documenting and reporting information on (Q)SARs and their estimates, respectively. It should be noted that these reporting formats are likely to evolve as experience is gained.

[2] http://www.oecd.org/dataoecd/33/41/37850114.pdf

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 QSAR validity and QSAR 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 (firstly by industry and then by the authorities working via an Agency committee). See Section R.6.1 for more detailed guidance.

(Q)SAR predictions may be gathered from databases (in which the predictions have already been generated and documented) or generated *de novo* through the applicable of available models. In the latter case, specialised expertise may be required.

Up to date information on QSAR models, QMRF, QPRF, editors, and examples is available in the JRC QMDB http://qsardb.jrc.it.

R.4.3.2.2 Data obtained by grouping approaches

Conclusions about the likely properties of a substance can also be based on the knowledge of the properties of one or more similar substances, by applying *grouping methods*. More details of such methods are provided in Section R.6.2.

REACH Annex XI contains under 1.5 a section on the use of grouping of substances and read-across approaches.

In this guidance, the terms *category approach* and *analogue approach* are used to describe techniques for grouping chemicals, whilst the term *read-across* is reserved for a technique of filling data gaps in either approach. The term *analogue approach* is sometimes used when the grouping is based on a very limited number of chemicals. A chemical category is a group of chemicals whose physico-chemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristic). In principle, more members are generally present in a chemical category, enabling the detection of trends across endpoints.

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

Grouping approaches avoid the need to test all members of the group for all endpoints of interest, thereby reducing costs and animal testing. Additional benefits are described in Section R.6.2.

The assessment of chemicals by using a category approach differs from the approach of assessing them on an individual basis, since the effects of the individual chemicals within a category are assessed on the basis of the evaluation of the category as a whole, rather than based on measured data for any one particular substance alone.

The category approach has been applied successfully under the EU classification system, where all *similar* substances (sometimes identifying all the individual substances, sometimes leaving them as a generic group) are expected to have the same property as the substance[3]. Categories have also been developed in the context of the OECD HPV Chemicals Programme (http://www.oecd.org/document/21/0,3343,en_2649_34379_1939669_1_1_1_1,00.html Within a chemical category, data gaps may be filled by applying one or more of three general approaches: a) read-across; b) trend analysis (i.e. use of internal models, purposefully developed from the underlying data of the category); and c) use of external models (e.g. QSARs, Quantitative Activity-Activity Relationships (QAARs) and expert systems that were not specifically developed in the context of the category).

Read-across is a technique for data gap filling in which information for one or more *source* chemicals is used to make a prediction for a *target* chemical, which is considered to be *similar* in some way. Read-across can be used to fill data gaps in the context of both the analogue approach and the wider category approach.

The chemical category approach is, by its very nature, a *Weight of Evidence* approach, since it integrates estimated and experimental data, and involves expert judgement. The category approach also provides a means of strategic testing. The biggest challenge in this approach lays in defining the category itself (its underlying rationale/mechanistic basis) and in particular its boundaries.

The wider category approach is considered to be more robust than simple analogue approaches, which are more limited, ad-hoc ways of comparing small numbers of substances. As the number of possible chemicals being grouped into a category increases, the potential for developing hypotheses for specific endpoints and making generalisations about the trends within the category will also increase, and hence increase the robustness of the evaluation.

When applying the category approach, the robustness of the overall category is assessed, rather than the reliability for an individual substance (since in some cases, individual substances may display exceptional behaviour). Thus, the adequacy (relevance and reliability) of the approach needs to be assessed for individual substances of interest.

Grouping approaches can be used directly to fulfil information requirements in REACH, provided a number of conditions are met. Although REACH makes no explicit reference to the need for validation for grouping approaches, it will be necessary for the industry registrant making use of a grouping method to provide a scientific justification and to demonstrate that the grouping approach used is adequate for the regulatory purpose (classification and labelling and/or risk assessment). Guidance on how to demonstrate the adequacy of grouping approaches is provided in Section R.6.2.4.1. Furthermore, appropriate documentation of the grouping approach must be provided in the form of a suitable reporting format, as also described in Section R.6.2.6.

[3] Under EU legislation, these *categories* are the *group entries* in Annex I of Directive 67/548/EEC.

R.4.3.3 Human data

The evaluation and use of information derived from studies in humans usually requires more elaborate and in-depth critical assessment of the reliability than animal data (WHO, 1983). Four major types of human data may be submitted (1) analytical epidemiology studies on exposed populations, (2) descriptive or correlation epidemiology studies, (3) case reports and (4) in very rare, justified cases controlled studies in human volunteers.

Analytical epidemiology studies (1) are useful for identifying a relationship between human exposure and effects such as biological effect markers, early signs of chronic effects, disease occurrence, or mortality and may provide the best data for risk assessment. Study designs include:

- **Case-control (case-referent) studies**, where a group of individuals with (cases) and without (controls/referents) a particular effect are identified and compared to determine differences in exposure in the recent or more distant past;
- **Cohort studies**, where groups of variously exposed and *non-exposed* individuals are identified and differences between the groups in effect occurrence over time are studied;
- **Cross-sectional studies**, where a population (e.g. a workforce) is studied, so that morbidity at a given point in time can be assessed in relation to concurrent exposure.

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. 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. In general, cohort studies provide stronger evidence than case-control studies, because exposure is assessed independently of the health status or outcome of the subjects in the study. Other characteristics that support a causal association are presence of a dose-response association, a consistent relationship in time and (biological) plausibility.

Criteria for assessing the adequacy of epidemiology studies include the proper selection and characterisation of the case and control groups (in case-control studies), adequate characterisation of exposure, sufficient length of follow-up for disease occurrence (in cohort studies), valid ascertainment of effect, proper consideration of biases and confounding factors. Assessment of adequacy of the studies should be conducted by epidemiologists by training.

Due to both uncertainties in epidemiological studies and true variability in the association between exposure and health outcomes within and among human populations, the available body of epidemiological evidence should be systematically reviewed and, if possible, combined. A *Weight of Evidence* approach is essential for risk assessment based on epidemiological data to (a) assess (sources of) heterogeneity across the studies and (b) increase statistical stability of the risk estimates. The best option to combine and summarise epidemiological data is a pooled analysis of the original data sets of the contributing studies. A meta-analysis based on published study results is a good, but somewhat more restricted alternative.

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

Descriptive epidemiology studies (2) examine differences in disease rates among human populations in relation to age, gender, race, and differences in temporal or environmental conditions. These studies are useful for identifying areas for further research but are not very useful for risk assessment. Typically these studies can only identify patterns or trends in disease occurrence over time or in different geographical locations but cannot ascertain the causal agent or degree of human exposure.

Case reports (3) describe a particular health condition in an individual or a group of individuals who were exposed to a substance. They may be particularly relevant when they demonstrate effects which cannot be observed in experimental animal studies. In many such studies, information is lacking on critical aspects such as substance identity and purity, exposure, health status of the persons exposed and even the symptoms reported; thorough assessment of the reliability and relevance of case reports is therefore necessary. Case reports also trigger analytical studies.

When they are already available, well-conducted controlled human exposure studies (4) in volunteers, including low exposure toxicokinetics studies, can also be used in risk assessment. However, few human experimental toxicity studies are available due to the practical and ethical considerations involved in deliberate exposure of individuals. Such studies, e.g. studies carried out for the authorisation of a medical product, have to be conducted in line with the World Medical Association Declaration of Helsinki, which describes the general ethical principles for medical research involving human subjects (World Medical Association, 2000).

Criteria for a well-designed experimental study include the use of a double-blind study design, inclusion of a randomised control group, sufficient duration of exposure and an adequate number of subjects to detect an effect. A meta-analysis of available similar, even small, studies is a good option.

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

R.4.4 Evaluation and Integration of all available Information including Weight of Evidence

Within the REACH legislation, the so-called *Weight of Evidence* (WoE) approach is a component of the decision-making procedure on substance properties and thus an important part of the chemical safety assessment.

The term WoE does neither constitute a scientifically well-defined term nor an agreed formalised concept characterised by defined tools and procedures (Weed, 2005). Nevertheless, from daily life everybody is familiar with the essence of *Weight of Evidence* reasoning and its basic mechanism may be regarded as a matter of commons sense.

An evidence based 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. To this end, a value needs to be assigned to each piece of information. These weights/values can be assigned either in an *objective* way by using a formalized procedure 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 and reliability and adequacy for the purpose have to be considered.

Examples of tools to identify the quality include the Klimisch scores (for toxicological studies, see also Section R.4.2), Hills criteria for evaluation of epidemiological data in Hill (1965), ranking of chemicals on their endocrine potential (Calabrese *et al*, 1997), evaluation of ecologic risk (Menzie *et al*, 1996).

An evidence based approach may imply formalised decision schemes where explicit rules for weighing information elements have been established. After having assessed/ranked the quality of the individual components the next step should be the integrating, comparing and putting together all information pieces with their relative values or weights and drawing a conclusion. This often includes expert judgement.

In the GHS, an evidence based approach is given a prominence for classification. All available information that can contribute to the determination of classification for an endpoint is considered together. Included is information such as epidemiological data and case reports in humans, and specific studies along with the sub-chronic, chronic and special study results in animals that provide relevant information, etc.

In REACH there will also be cases where data from sources other than tests specifically addressing an endpoint can provide valuable information. In addition, it is reasonable to expect that there will be cases where several pieces of *inadequate* data on a given REACH endpoint may exist. For example there may be several repeated dose studies available on a chemical, none of which would be acceptable by itself due to some deficiency (e.g. small group sizes, insufficient number of dose groups, insufficient parameters, etc). Collectively, however, the different studies show effects in the same target organ at approximately the same dose and time. If a rationale is given to show that such data adequately describe the REACH endpoint of concern, further information on that particular endpoint may not be necessary.

The way the *Weight of Evidence* is implemented is case-dependent. It is influenced by the relation between the amount of information needed and the importance of the decision to be taken and also by the likelihood of, and consequences for, the decision based on that information being wrong. It is important to document and communicate how the evidence based approach was used in a reliable, robust and transparent manner.

R.4.5 References

Calabrese, E. J., Baldwin, L. A.,Kostecki, P.T.,&Potter,T.L. (1997). A toxicologically based weight-of-evidence methodology for the relative ranking of chemicals of endocrine disruption potential. *Regulatory Toxicology and Pharmacology*, **26**, 36–40.

Curren, Southee, Spielmann, Liebsch, Fentem and Balls M. (1995) - The Role of Prevalidation in the Development, Validation and Acceptance of Alternative Methods - ECVAM Prevalidation Task Force Report 1. *ATLA,* **23**, 211-217.

Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. (2004) - A Modular approach to the ECVAM principles on test validity. *Altern Lab Anim,* **32,** 5, 467-472.

Hill, A. B. (1965).The environment and disease: Association or causation? *Journal of the Royal Society of Medicine*, **58**, 295–300.

Klimisch H., Andreae M. and Tillmann U. (1997) - A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharm*, **25**, 1-5.

Menzie,C., Henning,M. H., Cura, J., Finkelstein, K., Gentile, J., Maughan, J., Mitchell, D., Petron, S., Potocki, B., Svirsky, S., & Tyler, P. (1996). Special report of the Massachusetts weight-of-evidence workgroup: A weight-of-evidence approach for evaluating ecological risks. *Human Ecological Risk Assessment,* **2**, 2, 277–304.

OECD (2005a) – Manual for Investigation of HPV Chemicals. Chapter 3 – Data Evaluation. Organization for Economic Co-operation and Development.

OECD (2005b) - Guidance Document n°34 on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Organization for Economic Co-operation and Development.

OECD (2007) – Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models. OECD Series on Testing and Assessment No. 69. ENV/JM/MONO(2007)2. Organisation for Economic Cooperation and Development, Paris, France. 154 pp.
http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html.

Weed D. (2005) - Weight of evidence: a review of concept and methods. *Risk Anal*, **25,** 6, 1545-1557.

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.

World Medical Association (2000). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Adopted by the 18th World Medical Association (WMA) General Assembly Helsinki, Finland, June 1964 and amended by the 29th WMA General Assembly, Tokyo, Japan, October 1975, 35th WMA General Assembly, Venice, Italy, October 1983, 41st WMA General Assembly, Hong Kong, September 1989, 48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996 and 52nd WMA General Assembly,

Edinburgh, Scotland, October 2000.