

TECHNICAL GUIDANCE DOCUMENT FOR PREPARING THE CHEMICAL SAFETY ASSESSMENT

Chapter R.4: Evaluation of available information

**“Technical Guidance Documents in support of the New EU Chemicals Legislation (REACH) –
V: Development of a Technical Guidance Document for preparing the Chemical Safety
Assessment (REACH Implementation Project 3.2-2)”**

Service Contract Number CCR.IHCP.C432365.X0



Disclaimer: "The present views expressed are those developed by the RIP 3.2-2 contractor for discussion at the relevant Stakeholder Expert Group meetings and may not in any circumstances be regarded as a final position nor document"

CONTENTS

R.4 EVALUATION OF AVAILABLE INFORMATION	2
R.4.1 Relevance of information	3
R.4.2 Reliability of information	4
R.4.3 Adequacy of information	6
R.4.3.1 Non-human data	6
R.4.3.2 Non-testing data	10
R.4.3.2.4 (Q)SAR data	10
R.4.3.2.5 Data obtained by grouping approaches	12
R.4.3.3 Human data	13
R.4.4 Evaluation and Integration of all available Information including Weight of Evidence	15
R.4.5 References for chapter 4	16

TABLES

Table R.4-1: The criteria for validation derived from the OECD GD 34	8
Table B.4-2: The criteria for suitability assessment according to the ECVAM criteria for entering the pre-validation study, (Curren <i>et al</i> , 1995)	9

1 **R.4 EVALUATION OF AVAILABLE INFORMATION**

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

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

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

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

15 *“Relevance - covering the extent to which data and tests are appropriate for a particular*
16 *hazard identification or risk characterisation.*

17 *Reliability - evaluating the inherent quality of a test report or publication relating to*
18 *preferably standardised methodology and the way the experimental procedure and results*
19 *are described to give evidence of the clarity and plausibility of the findings. Reliability of*
20 *data is closely linked to the reliability of the test method used to generate the data (see*
21 *section R.4.2).*

22 *Adequacy - defining the usefulness of data for hazard/risk assessment purposes. Where*
23 *there is more than one study for each endpoint, the greatest weight is attached to the studies*
24 *that are the most relevant and reliable. For each endpoint, robust summaries need to be*
25 *prepared for the key studies.”*

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

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

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

36 A *Weight of Evidence* approach, mentioned in Annex XI section 1.2 of REACH, integrates available
37 information from guideline tests, non-guideline tests, and other types of information which may
38 justify adaptation of the standard testing regime.

39 **R.4.1 Relevance of information**

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

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

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

51 The evaluation of the relevance for humans of data from studies in laboratory animals is aided by
52 use of information (when available) on the toxicokinetics of the substance in both humans and the
53 animals species used in the toxicity tests, even when such information is relatively limited. Further
54 guidance on the value and use of toxicokinetics is given in section R.6.3.

55 Normally, for human health assessment, a *no* or *lowest observed (adverse) effect level* (NO(A)EL,
56 LO(A)EL), or a *benchmark threshold dose* for adverse effects in laboratory animals are extrapolated
57 to an exposure level (DNEL) below which it is assumed that adverse effects are unlikely to occur in
58 humans exposed to the substance. For substances evoking effects that have no definable threshold,
59 e.g. genotoxic carcinogens, it may not be possible to identify an exposure level without effects; in
60 such cases, extrapolation is made to an exposure level that represents a risk level of very low
61 concern for humans.

62 For environmental compartments such as surface water, sediment and soil, a *predicted no effect*
63 *concentration* (PNEC) is obtained by extrapolation based on the lowest *no observed effect*
64 *concentration* (NOEC) or *effect concentration* causing marginal effects (EC_x) by application of
65 assessment factors.

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

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

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

80 laboratory animals or as an important part of the basis for deciding whether such tests may be
81 warranted.

82 **R.4.2 Reliability of information**

83 The quality of the study, the method, the reporting of the results, and the conclusions that are
84 drawn, must be evaluated carefully. Reasons why existing study data may vary in quality include
85 the use of outdated test guidelines, the failure to characterise the test substance properly (in terms of
86 purity, physical characteristics, etc.), the use of crude techniques/procedures that have since become
87 refined, and the fact that certain endpoint information, now recognised as being important, may
88 have not been recorded or measured. Moreover, other reasons could be poor reporting of
89 information and poor quality assurance.

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

93 **1 = reliable without restrictions:** “*studies or data [...] generated according to generally*
94 *valid and/or internationally accepted testing guidelines (preferably performed according to*
95 *GLP) or in which the test parameters documented are based on a specific (national) testing*
96 *guideline [...] or in which all parameters described are closely related/comparable to a*
97 *guideline method.*”

98 **2 = reliable with restrictions:** “*studies or data [...] (mostly not performed according to*
99 *GLP), in which the test parameters documented do not totally comply with the specific*
100 *testing guideline, but are sufficient to accept the data or in which investigations are*
101 *described which cannot be subsumed under a testing guideline, but which are nevertheless*
102 *well documented and scientifically acceptable.*”

103 **3 = not reliable:** “*studies or data [...] in which there were interferences between the*
104 *measuring system and the test substance or in which organisms/test systems were used*
105 *which are not relevant in relation to the exposure (e.g. unphysiological pathways of*
106 *application) or which were carried out or generated according to a method which is not*
107 *acceptable, the documentation of which is not sufficient for assessment and which is not*
108 *convincing for an expert judgment.*”

109 **4 = not assignable:** “*studies or data [...] which do not give sufficient experimental details*
110 *and which are only listed in short abstracts or secondary literature (books, reviews, etc.).*”

111 The use of such scoring tools, e.g. the mentioned *Klimisch codes*, allows ranking the information,
112 and organising it for further review. This implies focussing on the most relevant ones, taking into
113 account the endpoint being measured or estimated. The evaluation of the reliability is performed
114 considering certain formal criteria using international standards as references. The scoring of
115 information, e.g. according to *Klimisch codes*, should not exclude all unreliable data from further
116 consideration by expert judgement because of possible pertinence of these data related to the
117 evaluated endpoints. In general, some types of data that are not reliable (i.e. those where insufficient
118 documentation exist for making an assessment) and data for which it is not possible to assign
119 reliability, may only be used as supporting data.

120 For many existing substances, at least some of the available information could have been generated
121 prior to the requirements of GLP and the standardisation of testing methods. While such
122 information may still be usable for REACH purposes, both the data and the methodology used must
123 be evaluated in order to determine their reliability. Such an evaluation needs evidence based

124 decision making following established criteria and must be transparent to justify the use of a
125 particular data set. For some substances, information may be available from tests conducted
126 according to the methods included in EU Annex V of Directive 67/548/EEC or to OECD Test
127 Guidelines (or other standards like CEN, ISO, ASTM, OSPAR methods, national standard
128 methods), and in compliance with the principles of GLP or equivalent standards. REACH Article
129 13.3 states that any new tests should be “*conducted in accordance with the test methods laid down
130 in a Commission Regulation or... other international test methods recognised by the Commission or
131 the Agency as being appropriate.... Information on intrinsic properties may also be generated
132 using other test methods provided they meet the conditions set out in Annex XI.*”

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

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

- 137 - The proven ability of the laboratory to perform the test method
- 138 - The purity/impurities and origin of the test substance, as well as the reference substances, must
139 be reported;
- 140 - The availability of the raw data from the study
- 141 - There must be an adequate description of the study e.g. a complete test report, or a sufficiently
142 detailed description of the test procedure, which must be in accordance with generally accepted
143 scientific standards. In these cases, the information may be considered reliable;
- 144 - When the test procedure used to generate the test data is found to differ significantly from that
145 described by the recognised test method or generally accepted scientific standards, or the
146 reliability of the data cannot be established fully, the assessor must decide if and how the
147 information can be used, e.g. as supporting information where a reliable study already exists.
148 The following factors, inter alia, can be used to support the view that these data may be
149 acceptable for use in meeting the requirements of REACH:
 - 150 o there are other studies or calculations available on the substance, and the data under
151 consideration are consistent with them,
 - 152 o other studies are available, for example on isomers with similar structure activity profile,
153 homologues, relevant precursors, breakdown products or other chemical analogues, and the
154 data under consideration are consistent with them,
 - 155 o an approximate value is sufficient for taking a decision on the endpoint of interest for the
156 conclusion required by REACH;
- 157 - Where critical supporting information is not reported (e.g. species tested, substance identity and
158 dosing procedure) the test data should be considered to be unreliable for the purposes of
159 REACH.

160 In principle, the same criteria apply to test data reported in the published literature; the extent of the
161 information provided will provide the basis for deciding upon the reliability of the data reported. In

162 general, publications in peer-reviewed journals are preferable to those which are not. High-quality
163 reviews, summaries or abstract publications may be used as supporting information.

164 **R.4.3 Adequacy of information**

165 Adequacy defines the usefulness of information for the purpose of hazard and risk assessment, in
166 other words whether the available information allows clear decision-making by the registrant about
167 (a) whether the substance meets the criteria for classification, (b) whether it is a potential PBT/vPvB
168 and (c) whether appropriate DNEL/PNEC values can be derived for risk assessment purposes. The
169 evaluation of adequacy of test results and documentation for the intended purpose is particularly
170 important for substances under REACH where there may be (a number of) test results available for
171 each effect, but where some or all of them have not been carried out according to current standards.
172 Where there is more than one study for each endpoint, the greatest weight is attached to the studies
173 that are the most relevant and reliable. For each endpoint, robust summaries need to be prepared for
174 the key studies. Sound scientific judgement is an important principle in considering the adequacy of
175 information and determining the key study.

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

180 **R.4.3.1 Non-human data**

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

185 **IN VITRO DATA**

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

189 **USE OF *IN VITRO* METHODS WITHIN REACH**

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

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

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

201 the ECVAM criteria of entering the pre-validation process).
202 There are clear definitions on what constitutes a fully validated *in vitro* assay. These criteria are
203 detailed in OECD GD 34 (see details in Table R.4-2) and were initially established by ECVAM and
204 ECB and later refined by ECVAM (Hartung *et al*, 2004).

205 **USE OF ADEQUATE INFORMATION DERIVED FROM *IN VITRO* METHODS**

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

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

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

217 ***Information derived from suitable in vitro methods***

218 Annex XI section 1.4 opens the way for the use of results of *in vitro* methods that have not yet been
219 scientifically validated but are identified as being *suitable*, meaning that the methods are
220 sufficiently well developed according to internationally agreed test development criteria e.g. the
221 ECVAM criteria for entry of the method into the pre-validation process (see Table R.4-2 and
222 section R.5.1.1.4 for a discussion of the use of *in vitro* testing to adapt the standard testing regime).

223

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

224

225 INFORMATION FROM IN VITRO TEST MAY PROVIDE MECHANISTIC INSIGHT

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

231 ADEQUACY OF INFORMATION FROM *IN VITRO* TESTING

232 The assessment of alternative testing data (to decide whether and how they can be used) in terms of
 233 adequacy for fulfilling the information requirements of REACH will follow the general criteria
 234 already discussed, e.g. applied quality measures, i.e. how they take into account the relevance,
 235 reliability and completeness of the information with regard to the regulatory decision to be taken.

236 This includes how well the study is reported, how well the test substance is characterised and to
 237 what extent the information requirements have been met for the endpoint under consideration.

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

Concerned items	Decision criteria to be considered
Purpose and proposed use	<ul style="list-style-type: none"> - 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 <i>in vivo/in vitro</i> test, state of the art	<p>Complete and concise presentation of state of the art, human data, <i>in vivo</i>, non-testing and <i>in vitro</i> data</p> <p>Weighed judgment about the contribution of the proposed test method compared to state of the art, including weaknesses and limitations</p> <p>e.g.: improved reliability: accuracy, sensitivity, specificity, robustness, defined performance</p> <p>e.g. improved relevance: predictive capacity, applicability domain</p>
Addressed endpoint described	<ul style="list-style-type: none"> - Demonstration of relevance for the <i>in vivo</i> situation - Description of data analysis and interpretation
Availability of a written procedure detailed enough to allow performance in another laboratory	<p>Method protocol:</p> <ul style="list-style-type: none"> - complete and readable - feasible and transferable - SOP standardised with respect to selected model and measurement performance
Reference substances, test materials and related results	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - 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

240

241 **R.4.3.2 Non-testing data**

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

245 **R.4.3.2.4 (Q)SAR data**

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

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

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

259 REACH Annex XI also indicate that the Agency in collaboration with the Commission, Member
260 States and interested parties shall develop and provide guidance in assessing which (Q)SARs will
261 meet these conditions and provide examples. In the meantime, the ECB has started the development
262 a QSAR Inventory to provide information on QSAR models and their validity
263 (<http://ecb.jrc.it/QSAR>). In addition to replacing the need for testing, (Q)SAR results may also in
264 some cases indicate the need for further testing.

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

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

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

278 A valid (Q)SAR is a model that has been characterised and documented according to the
279 internationally agreed OECD Principles for the validation of (Q)SAR models. According to these
280 principles, a (Q)SAR model that is proposed for regulatory use should be associated with a defined
281 endpoint (principle 1), an unambiguous algorithm to ensure transparency in the model algorithm

282 (principle 2), a defined domain of applicability (principle 3), and appropriate measures of internal
283 performance and predictivity (principle 4). If possible, a mechanistic interpretation should also be
284 provided, to add to the confidence in the model (principle 5).

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

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

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

299 The QSAR Model Reporting Format (QMRF) has been developed to provide a means of
300 documenting (Q)SAR model characteristics in a transparent and consistent manner, in accordance
301 with the OECD validation principles. Further information on QMRFs is given in section R.6.1.10.
302 In particular, the ECB Inventory of QSAR models is being developed as a repository of quality-
303 reviewed information on QSAR models and their validity. In this database, QSAR models will be
304 linked with their corresponding QMRFs. Before developing a QMRF, the registrant should check
305 whether it is already included in the ECB QSAR Inventory or other suitable source (e.g. OECD
306 QSAR Toolbox¹). If the appropriate QMRF for a given model is not already available, it will be
307 necessary to develop one by applying the five validation principles and documenting the results.
308 Since the general format of the QMRF is already defined, it is sufficient to fill this in with the
309 appropriate information on the model. The ECB has developed a QMRF editor as a tool to facilitate
310 the generation of new QMRFs.

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

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

320 The information in the QMRF and QPRF should be used when assessing whether a prediction is
321 adequate for the purpose of classification and labelling and/or risk assessment. The assessment will
322 also need to take into account the regulatory context. This means that the assessments of QSAR
323 validity and QSAR estimate reliability need to be supplemented with an assessment of the relevance
324 of the prediction for the regulatory purposes, which includes an assessment of *completeness*, i.e.

¹ <http://www.oecd.org/dataoecd/33/41/37850114.pdf>

325 whether the information is sufficient to make the regulatory decision, and if not, what additional
326 (experimental) information is needed. The decision will be taken on a case-by-case basis (firstly by
327 industry and then by the authorities working via an Agency committee). See section R.6.1 for more
328 detailed guidance.

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

332 Up to date information on QSAR models, QMRF, QPRF, editors, and examples is available on the
333 ECB website <http://ecb.jrc.it/qsar>.

334 **R.4.3.2.5 Data obtained by grouping approaches**

335 Conclusions about the likely properties of a substance can also be based on the knowledge of the
336 properties of one or more similar substances, by applying *grouping methods*. More details of such
337 methods are provided in section R.6.2.1, illustrated in figure R.4-7 and figure R.4-8.

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

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

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

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

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

356 The category approach has been applied successfully under the EU classification system, where all
357 *similar* substances (sometimes identifying all the individual substances, sometimes leaving them as
358 a generic group) are expected to have the same property as the substance². Categories have also
359 been developed in the context of the OECD HPV Chemicals Programme ([http://cs3-hq.oecd.org/
360 scripts/hpv](http://cs3-hq.oecd.org/scripts/hpv)).

361 Within a chemical category, data gaps may be filled by applying one or more of three general
362 approaches: a) read-across; b) trend analysis (i.e. use of internal models, purposefully developed
363 from the underlying data of the category); and c) use of external models (e.g. QSARs, Quantitative

² Under EU legislation, these *categories* are the *group entries* in Annex I of Directive 67/548/EEC.

364 Activity-Activity Relationships (QAARs) and expert systems that were not specifically developed
365 in the context of the category).

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

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

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

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

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

391 **R.4.3.3 Human data**

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

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

- 400 - **Case-control (case-referent) studies**, where a group of individuals with (cases) and without
401 (controls/referents) a particular effect are identified and compared to determine differences in
402 exposure in the recent or more distant past;
- 403 - **Cohort studies**, where groups of variously exposed and *non-exposed* individuals are identified
404 and differences between the groups in effect occurrence over time are studied;

405 - **Cross-sectional studies**, where a population (e.g. a workforce) is studied, so that morbidity at a
406 given point in time can be assessed in relation to concurrent exposure.

407 The strength of the epidemiological evidence for specific health effects depends, among other
408 things, on the type of analyses and on the magnitude and specificity of the response. Confidence in
409 the findings is increased when comparable results are obtained in several independent studies on
410 populations exposed to the same agent under different conditions. In general, cohort studies provide
411 stronger evidence than case-control studies, because exposure is assessed independently of the
412 health status or outcome of the subjects in the study. Other characteristics that support a causal
413 association are presence of a dose-response association, a consistent relationship in time and
414 (biological) plausibility.

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

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

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

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

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

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

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

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

455 **R.4.4 Evaluation and Integration of all available Information including Weight of** 456 **Evidence**

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

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

464 An evidence based approach involves an assessment of the relative values/weights of different
465 pieces of the available information that has been retrieved and gathered in previous steps. To this
466 end, a value needs to be assigned to each piece of information. These weights/values can be
467 assigned either in an *objective* way by using a formalized procedure or by using expert judgement.
468 The weight given to the available evidence will be influenced by factors such as the quality of the
469 data, consistency of results/data, nature and severity of effects, relevance of the information for the
470 given regulatory endpoint. In all cases the relevance and reliability and adequacy for the purpose
471 have to be considered. Examples of tools to identify the quality include the Klimisch scores (for
472 toxicological studies, see also section R.4.2), Hills criteria for evaluation of epidemiological data in
473 Hill (1965), ranking of chemicals on their endocrine potential (Calabrese *et al*, 1997), evaluation of
474 ecologic risk (Menzie *et al*, 1996).

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

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

485 In REACH there will also be cases where data from sources other than tests specifically addressing
486 an endpoint can provide valuable information. In addition, it is reasonable to expect that there will
487 be cases where several pieces of *inadequate* data on a given REACH endpoint may exist. For
488 example there may be several repeated dose studies available on a chemical, none of which would
489 be acceptable by itself due to some deficiency (e.g. small group sizes, insufficient number of dose
490 groups, insufficient parameters, etc). Collectively, however, the different studies show effects in the
491 same target organ at approximately the same dose and time. If a rationale is given to show that such

492 data adequately describe the REACH endpoint of concern, further information on that particular
493 endpoint may not be necessary.

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

499 **R.4.5 References for chapter 4**

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

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

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

508 Hill, A. B. (1965). The environment and disease: Association or causation? *Journal of the Royal Society of Medicine*, **58**,
509 295–300.

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

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

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

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

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

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

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

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

531 Edinburgh, Scotland, October 2000.

532 Worth A. and Balls M. (2001) - The role of ECVAM in promoting the regulatory acceptance of alternative methods in
533 the European Union. European Centre for the Validation of Alternative Methods. *Altern Lab Anim*, **29**, 5, 525-535.