



The Voice of OECD Business

BIAC Perspective on a Globally Harmonized Approach for the Evaluation of Endocrine Activity of Chemical Substances

December 2008

1. The chemical industry is committed to providing chemicals that can be manufactured, transported, used and disposed of safely. The industry is also committed to making health, safety, environment and resource conservation critical considerations for all new and existing products and processes. In alignment with Responsible Care®, the member companies of chemical industry associations seek to understand risks posed by products throughout their life-cycle and to have systems in place to manage those risks effectively and appropriately in line with the shared responsibility with downstream users. Consistent with that commitment to Responsible Care®, the chemical industry recognizes the public's concerns that exposure to small levels of man-made and naturally occurring chemicals in the environment may pose a risk to humans and wildlife.

2. The chemical industry will continue to work together with government, the scientific community and other stakeholders to better understand the science related to the endocrine disruption issue. Through ongoing participation in the OECD EDTA, industry has contributed, and will continue to contribute, to the development, standardization and validation of new and revised OECD endocrine screens and tests. Internationally harmonized OECD Test Guidelines based on scientifically validated test methods must form the core foundation for screening substances for the potential to interact with components of the endocrine system and for such substances which have such a potential, for testing these materials to understand the dose-response relationship for production of adverse effects.

3. The chemical industry supports the development of internationally harmonised procedures to prioritise substances efficiently, to screen for endocrine activity, to test for adverse effects and to evaluate substances in the framework of a coherent chemicals policy. These procedures should provide for a tiered, hierarchical scientific framework in which validated screening assays are used to identify substances with endocrine activity and prioritise substances for further, more definitive testing. Definitive testing, using validated, harmonised protocols, is necessary to identify adverse effects caused by alterations to endocrine system function. Using a tiered approach, results from definitive tests must outweigh or supersede results from screening assays in guiding policy and management in both the public and private sectors. Definitive tests are necessary for hazard and risk characterization.

4. For hazard characterization, the chemical industry supports the development of a "weight of evidence" evaluation process that consists of a comprehensive, objective, transparent and balanced interpretation of the totality of scientific evidence regarding hormonal activity and adverse effects that might result from an endocrine mechanism. A defensible hazard characterization for hormonally active chemicals requires not only summarizing toxicological screening and testing data (hazard identification), but also requires an objective weight of evidence evaluation of whether the effects produced are adverse and whether adverse effects are due to a hormonal activity of the chemical.

5. The chemical industry supports the development of a globally harmonized approach concerning the interpretation of endocrine screening data, integration of screening data with definitive testing data and use of such data in prioritization, hazard characterization and risk assessment. Screening assays are to be used to prioritize substances for subsequent definitive testing. The results of definitive tests will provide data on adverse effects and dose response for use in hazard characterization and risk assessment. To perform risk assessments based on screening results would be an inappropriate use of the data. As the OECD finalizes test method validation and adopts endocrine Test Guidelines, there is a pressing need for a globally harmonized evaluative processes and framework because such assays will soon be deployed in the U.S. as part of the EPA's Endocrine Disruptor Screening Program and in Europe as part of REACH. Consistent interpretations of the same data from the same endocrine-specific Test Guidelines and consistent use of such information in hazard assessment are needed to uphold OECD's Mutual Acceptance of Data.

6. The chemical industry supports the development of overarching guidance, applicable to all OECD member countries to ensure uniform and consistent description of the results of endocrine screening assays as well as the appropriate application of a weight of evidence framework for use of screening and definitive testing results in hazard characterization and risk assessment. (For specifics on weight of evidence see Attachment 1 CEFIC EMSG weight of evidence paper).

7. A tiered, hierarchical scientific framework for Endocrine screening and testing provides the most efficient and effective approach to evaluating substances. In this hierarchical framework, OECD validated, internationally harmonized test guidelines are used to screen substances identifying those with the potential to interact with one or more components of the endocrine system, and prioritizing for further definitive testing. Definitive testing, using validated, harmonized protocols, is necessary to identify adverse effects caused by alterations to endocrine system function. Using a tiered approach, results from definitive tests must outweigh or supersede results from screening assays in policy and risk management. Results from *in vitro* and *in vivo* screening assays only indicate that a substance interacts with a component of the endocrine system through one mechanism. Such screening results do not provide evidence on whether that substance will cause adverse biological effects. The *in vitro* and *in vivo* screening assays do not represent either the biological complexity of the intact endocrine system of an organism or evaluate adverse effects. If results from screening assays suggest the potential for endocrine activity, longer term *in vivo* studies are necessary. In this case exposures will be evaluated in the complete and intact endocrine system encompassing critical life stages and processes. In addition, assessments will include measurements of adverse effects, in order to fully characterize

dose response of the potential endocrine mediated adverse health effects upon the endocrine system and in other tissues or organs.

8. Experience has shown that arraying toxicity tests in a tiered framework, provides scientific rigour and flexibility to account for differing chemical toxicities and to address specific concerns associated with existing or anticipated exposures to specific chemicals. In this manner tiered testing focuses efforts to collect data where it is most needed, to promote screening of the greatest number of prioritized chemicals (or classes of chemicals) and, where indicated by specific toxicity results, to indicate which substances pose a particular concern and points to the specific, more complex, test that should be considered. Where scientifically supported, the use of chemical categories can also increase efficiencies in providing knowledge of potential biological activities of chemical substances.

9. The chemical industry agrees that chemicals should continue to be tested for adverse health effects of concern under chemical regulation laws, the OECD high production volume chemical testing programme, REACH and pursuant to product stewardship activities of individual companies even as the development, standardization and validation of harmonized OECD test guidelines of screens and tests for endocrine activity proceeds.

10. The chemical industry has and will continue to contributing funding and expertise to national and international efforts to develop and validate *in vitro* and *in vivo* screening assays necessary to identify and prioritize chemicals that may interact with the endocrine system in humans and wildlife. The chemical industry has been and will continue to participate in the Organisation for Economic Cooperation and Development (OECD) Endocrine Disrupter Testing and Assessment (EDTA) efforts to develop and validate endocrine screens and tests. The chemical industry members urge governments to take advantage of the OECD Task Force on EDTA's on-going efforts to standardise and validate the globally harmonised test guidelines prior to initiating new, routine screening and testing procedures for endocrine disruption.

11. The chemical industry recognizes the benefits of international cooperation among all stakeholders in the validation of screening and testing methods to most efficiently use existing resources, to avoid unnecessary duplication of effort, to minimize the use of laboratory animals, and to help ensure mutual acceptance of data that is essential for international harmonization.

12. The chemical industry agrees that further research and understanding provide a sound basis for more effective public policy. Additional research is necessary to increase our scientific understanding and determine if there are possible adverse effects in humans and wildlife via hormone-mediated processes from exposure to chemicals in the environment. As has been stated in the comprehensive International Union of Pure and Applied Chemistry publication on endocrine disruption, J. Miyamoto and J. Burger (2003), "It is too early to reach firm conclusions about whether human populations are seriously at risk from potential exposures to EAS, and further vigilance is clearly required. However, it is somewhat reassuring that after substantial research in the past decade, there have been no conclusive findings of low level environmental exposures to EAS causing human disease."

13. The chemical industry, on a global basis, coordinates the research of its members to maximize efficiency and effectiveness and prevent duplication of effort, in part through the Long-range Research Initiative (LRI). Working with academic and research institutions and governments, LRI members in Europe, North America and Asia sponsor basic research on the potential of chemicals to interact with and affect the hormone system and cause adverse effects. LRI members have pledged to conduct research through an open and transparent process at institutions selected through a competitive peer-reviewed process. The results of this research will be made available to the public and acted upon by industry in a timely manner.

14. A tiered, hierarchical scientific framework is needed for screening and testing (Table 1). Validated screening assays are used to identify substances with endocrine activity and prioritise substances for further, more definitive testing. Definitive testing, using validated, harmonized protocols, is necessary to identify adverse effects caused by alterations to endocrine system function. Using a tiered approach, results from definitive tests must outweigh or supersede results from screening assays in guiding policy and management in both the public and private sectors. For discussion, three tiers are suggested.

Table 1. Proposed OECD Endocrine Screening and Testing Hierarchical Framework

STAGE	DESCRIPTION	ASSAYS FOR MAMMALIAN TOXICITY	ASSAYS FOR ECOTOXICITY
<p>OECD Stage 1 Initial Assessment to Set Priorities for Further Evaluation</p>	<p>Available data (for example) Production volume and pattern of use Available exposure information Predicted environmental properties, e.g., fate Toxicological data, especially endocrine-relevant data (i.e., results of histopathology on reproductive organs from repeat dose studies, developmental or reproductive toxicological information.)</p>	<p>All relevant studies</p>	<p>It is presumed that the data review will incorporate both mammalian and ecotoxicological issues.</p>
	<p>Structure activity relationship</p>	<p>It is presumed that structure activity relationships for receptor mediated modes of action will be applicable across mammalian orders.</p>	<p>It is presumed that structure activity relationships will be applicable across vertebrate classes. Invertebrates may have unique receptors.</p>
	<p>Molecular screening results</p>	<p>All relevant studies</p>	<p>All relevant studies</p>

<p>OECD Stage 2 Screening Assays (mode of action)</p> <p>OECD 2002 Framework Level 2-4</p>	<p><i>In vitro</i> assays providing mechanistic information / data OECD 2002 Framework Level 2</p>	<p>Oestrogen and androgen and receptor binding assays Transfected mammalian cell assays (ER and AR and TR) <i>In Vitro</i> Aromatase <i>In Vitro</i> Steroidogenesis</p>	<p>It is presumed that receptor binding will in principle be applicable across vertebrate classes and to any invertebrates expressing similar receptors.</p>
	<p><i>In vivo</i> assays providing mechanistic information / data on single endocrine mechanisms OECD 2002 Framework Level 3</p>	<p>Uterotrophic screening assay (estrogen and anti-oestrogen) Hershberger screening assay (androgen and anti-androgen)</p>	<p>Fish screening assay (vitellogenin and secondary sex characteristics) Frog metamorphosis assay</p>
	<p><i>In vivo</i> assays providing mechanistic information / data on multiple endocrine mechanisms OECD 2002 Framework Level 4</p>	<p>Enhanced TG407* Adult Intact Male assay Male Pubertal assay Female Pubertal assay</p>	<p>OECD Fish Screening Assay (VTG and secondary sex characteristics as mandatory endpoints; other endpoints are optional)</p>

<p>OECD Stage 3 Definitive Testing (evaluation of apical endpoints, adverse effects and dose response for hazard identification and characterization)</p> <p>OECD 2002 Framework Level 5</p>	<p>Reproduction/developmental tests –shorter scope - includes <i>in utero</i> exposure, developmental, and reproductive capacity endpoints</p>	<p>(TG407* (as adopted on October 03,2008) depending on the exposure situation as the method does not include the reproductive phase) Reproductive /developmental screening test (TG 421)</p> <p>Combined repeat dose with reproduction / developmental screening (TG 422)</p> <p>One generation reproductive toxicity (TG 415)</p> <p>Two generation reproductive toxicity (TG 416)</p> <p>[Enhanced one generation reproductive toxicity -if and when a final OECD TG is developed]</p>	<p>Partial and full life cycle assays in fish, birds, amphibians and invertebrates (developmental and reproduction) Fish full life cycle</p>
--	--	---	--

** Remark: Depending on the situation a TG-407 would suffice to establish a NOAEL in selected instances. Such instances could include, for example, substances with low potency, minimal human exposure likely intermediates or substances manufactured in closed system, and limited potential for environmental release. This would serve to focus the more extensive testing only on substances that have high exposure potential.*

15. Proposed OECD Stage 1: Initial Assessment

The purpose of the proposed OECD Stage 1: Initial Assessment (Table 2) is to rapidly assess the universe of substances in order to recognize those substances where scientifically relevant data exist to permit more rapid prioritization and assessment.

The OECD Stage 1: Initial Assessment will permit one to:

- Recognize those substances (a) where pertinent data do not exist and (b) have some likelihood of exhibiting the mode(s) of action, e.g., binding to the oestrogen or androgen receptors, that lead to the particular hazard(s), e.g., reproductive and developmental effects, targeted by the assessment. These substances should then be prioritized.
- Recognize those substances which do not take part or are very unlikely to take part in the mode(s) of action(s) that lead to the particular hazard(s) targeted by the assessment. This reduces the overall effort and size of the assessment program. These substances can be reconsidered should new and compelling information emerge

Table 2. Proposed OECD Stage 1: Initial Assessment to Set Priorities for Further Evaluation

<p>OECD Stage 1 Initial Assessment to Set Priorities for Further Evaluation</p>	<p>Available data (for example) Production volume and pattern of use Available exposure information Predicted environmental properties, e.g., fate Toxicological data, especially endocrine-relevant data (i.e., results of histopathology on reproductive organs from repeat dose studies, developmental or reproductive toxicological information.)</p>	<p>All relevant studies</p>	<p>It is presumed that the data review will incorporate both mammalian and ecotoxicological issues.</p>
	<p>Structure activity relationship</p>	<p>It is presumed that structure activity relationships for receptor mediated modes of action will be applicable across mammalian orders.</p>	<p>It is presumed that structure activity relationships will be applicable across vertebrate classes. Invertebrates may have unique receptors.</p>
	<p>Molecular screening results</p>	<p>All relevant studies</p>	<p>All relevant studies</p>

16. Review of available toxicological information is an important step. Current toxicology testing protocols provide much information about potential endocrine mediated toxicity. Endpoints indicating endocrine-mediated toxicities are present in current testing guidelines or are being developed:

- Endocrine-mediated cancer mechanisms are well known, and endpoints for these mechanisms are included in current cancer bioassays,
- Multigenerational and reproductive studies are designed to examine all relevant windows of sensitivity; *in utero*, lactational and pubertal as well as periods of functional reproductive capacity in adults,
- Work is in progress via the OECD Endocrine Disruptor Testing and Assessment Task Force to review and where appropriate change/introduce screening and testing protocols.

17. For priority setting purposes, the totality of scientific evidence regarding hormonal activity and adverse effects that might result from an endocrine mechanism should be considered. This should include: assessment of existing regulatory actions; existing data on adverse effects; the presence or absence of response in different taxa; the nature and magnitude of positive responses in relation to the relevance & reliability of the assays; dose response (or lack thereof); repeatability; relative potency; and coherence of responses across assays in relation to the postulated mode of action. In cases where the weight of evidence indicates no, or at most very low, hormonal potency and there is little or no likelihood of release to the environment or potential for exposure, then such substances should be given a very low priority for further investigations.

18. Structure activity relationships ((Q)SARs) and molecular screening methods, where the model predictions can be validated, would permit the rapid assessment of specific, targeted mode(s) of action. For example, a relevant property for an oestrogenic mode of action is to bind to an oestrogen receptor. The volume and nature of the oestrogen-binding site determines a substance's ability to bind to the receptor and the actual binding affinity. (Q)SARs and molecular screening assays can then assist in the prioritization of substances with predictions of binding affinity and can also indicate a large number of substances where binding to an oestrogen receptor is highly unlikely.

19. Prioritization should include an integrated evaluation of several attributes such as:

- the possible potency of the substance as indicated by initial structural activity assessments;
- production volumes;
- the amount of environmental releases and the environmental levels indicated by monitoring;
- environmental properties such as long-range transport, persistence and bioaccumulation;

-
- use as related to the number of persons exposed, the relevance of the route of exposure, the intensity and duration of exposure, and possibility for disproportionate exposure in subpopulations; and
 - weight of evidence evaluation of existing data.

20. Proposed OECD Stage 2: Screening.

The purpose of the Proposed OECD Stage 2 Screening is to efficiently and effectively develop information as to whether a substance has the potential to interact with one or more components of the endocrine system. Most often OECD Stage 2: Screening assays are based on a particular mode of action especially leading to relevant biological responses. A positive mechanistic response warrants further evaluation and, if necessary, testing to characterize relevant hazard(s), if any, connected with the mode of action. 21. The proposed OECD Stage 2: Screening (Table 3) is subdivided into *in vitro* assays and *in vivo* assays as these assays are not equivalent. Both economy and animal welfare considerations argue that *in vitro* assays be employed both to accommodate a significant number of compounds and to avoid undue delays in assessing substances. For flexibility, the option to proceed directly to *in vivo* assays should not be precluded, nor should *in vitro* assay results be required if *in vivo* results are available. *In vivo* assays incorporate i) substance-specific complexities that cannot be obtained from *in vitro* assays, including absorption, distribution, metabolism and excretion, and ii) reflect the complex and dynamic homeostasis and operation of the intact endocrine system. Therefore, *in vivo* results would supersede *in vitro* results as noted by the Veldhoven workshop (SETAC Europe 1997). Examples of *in vitro* assays for an oestrogen mode of action include: oestrogen receptor-binding assays, yeast strains containing a hormone receptor and a hormone responsive reporter gene, or mammalian cell lines naturally expressing the oestrogen receptor that are transfected with a reporter gene controlled by an oestrogen receptor.¹ An example of an *in vivo* screening assay is the uterotrophic assay for an oestrogen mode of action. The assay is based on the growth and weight increase response of the uterine target organ, can be conducted in 3 days, and uses a minimum number of animals. The response of the uterine target-organ is highly relevant for the oestrogen mode of action.

¹ Mammalian cell lines transfected with both a receptor and a reporter gene may have patent protection and their use therefore may be restricted.

Table 3. Proposed OECD Stage 2: Screening Assays to Evaluate the Potential of Substances to Interact with One or More Components of the Endocrine System

LEVEL	DESCRIPTION	ASSAYS FOR MAMMALIAN TOXICITY	ASSAYS FOR ECOTOXICITY
Stage 2: Screening (mode of action)	<i>In vitro</i> assays providing mechanistic information / data OECD 2002 Framework Level 2	Oestrogen and androgen and receptor binding assays Transfected mammalian cell assays (ER and AR and TR) <i>In Vitro</i> Aromatase <i>In Vitro</i> Steroidogenesis	It is presumed that receptor binding will in principle be applicable across vertebrate classes and to any invertebrates expressing similar receptors.
	<i>In vivo</i> assays providing mechanistic information / data on single endocrine mechanisms OECD 2002 Framework Level 3	Uterotrophic screening assay (estrogen and anti-oestrogen) Hershberger screening assay (androgen and anti-androgen)	Fish screening assay (vitellogenin and secondary sex characteristics) Frog metamorphosis assay
	<i>In vivo</i> assays providing mechanistic information / data on multiple endocrine mechanisms OECD 2002 Framework Level 4	TG407* Adult Intact Male assay Male Pubertal assay Female Pubertal assay	

* Remark: Depending on the situation a TG-407 would suffice to establish a NOAEL in selected instances. Such instances could include, for example, substances with low potency, minimal human exposure likely intermediates or substances manufactured in closed system, and limited potential for environmental release. This would serve to focus the more extensive testing only on substances that have high exposure potential.

22. The principles for the consideration of assays into the screening level of the tiered framework are:

- Individual assays should be selected and the protocols designed to provide the necessary information in a resource efficient manner. That is, in addition to having the capacity to address the anticipated number of substances, the assays themselves should be simple, rapid, and economical.
- Highly specialized and technically demanding techniques should be avoided unless essential. The assay technique(s) should be within the competence and qualifications of the necessary laboratory resources.
- As the results from assays comprising the proposed OECD Stage 2: Screening provide only mechanistic information and not evidence for adverse effects, such results should not be used for classification or regulation. In accordance with the Weybridge definitions (EC 1996), OECD Stage 2: Screening stage results do not indicate that a compound is an 'endocrine disruptor.'
- The TG407 has some utility as both a screening assay and definitive test. The TG407 has played a major role in the OECD SIDS battery as one of the leading repeat dose toxicity studies for commodity chemicals. The validation of the TG407 showed that only substances with strong or moderate endocrine activity were detected, especially as for weakly active substances, the overall study outcome was governed by general toxicity. In this respect, the TG 407 is a method evaluating some endocrine specific endpoints as well as apical endpoints, adverse effects and dose response, and can therefore be used in certain instances for hazard identification and characterization. As such, the TG407 would fall within the proposed OECD Stage3: Definitive Testing. However, as described in the newly issued TG407 OECD Test Guideline, there are some strengths and limitations inherent in employing the TG407 as either a screening assay in the proposed OECD Stage 2: Screening step or as a test for adverse effects in the proposed OECD Stage3: Definitive Test step. In a comparison of TG407 to higher tier studies, with regard to the observed overall (i.e., irrespective of the type of toxicity) No Adverse Effect Levels, the difference in sensitivity is usually not more than a factor of 10 (Gelbke *et al.*, 2007). Depending on the exposure situation a TG 407 may suffice for the proposed OECD Stage3: Definitive Testing for human health hazard characterization in selected instances. Such instances could include, for example, substances with low potency, confined intermediates, minimal human exposure, and limited potential for environmental release. Reliance on the TG407 in such cases for human health hazard characterization would serve to focus the more extensive testing only on substances that have high exposure potential

23. The screening assays comprising Stage 2: Screening focus on detecting estrogenic, androgenic and thyroid modes of action. To effectively screen chemicals with unknown endocrine activity, evaluation would consist of laboratory studies of the chemical in the complete battery of assays that comprise the proposed Stage 2: Screening step. If information from existing assays or functionally equivalent test methods is available that

allows one to reach a scientifically sound conclusion regarding the activity of the substance with respect to one or more of the modes of action encompassed in the proposed Stage 2: Screening step, then those assays for such modes of action would not necessarily need to be run for such a substance.

24. A weight of evidence process needs to be implemented in order to integrate results across the complement of assays that comprise the proposed OECD Stage 2: Screening step. Substances which are positive based on overall consideration of the weight of evidence in OECD Stage 2: Screening are considered to be high priority candidates for further evaluation in definitive tests (OECD Stage 3: Definitive Testing). However, there is not an automatic triggering. Instead, a weight of evidence approach (see Attachment 1, CEFIC EMSG Weight of Evidence paper) should be used to evaluate results of screening battery. Prior to initiating additional work, it is appropriate to consider the potential for human exposure and potential for entrance into the environment. The weight of evidence evaluation should include: the presence or absence of response in different taxa; the nature and magnitude of positive responses in relation to the relevance and reliability of the assays; dose response (or lack thereof); relative potency; and coherence of responses across assays in relation to the postulated mode of action. In addition, as indicated above, in cases where the weight of evidence (derived from consideration of results of validated screening assays) indicates at most very low potency and there is little or no likelihood of release to the environment or potential for exposure, then such substances should be given a very low priority for further investigation in definitive tests. Existing information and data from standard toxicity studies (e.g., Repeat Dose Study (TG 407 and TG 408), Developmental Toxicity Study (TG 414; TG 421/422), Chronic Toxicity Study (TG 452 and 453)) should be reviewed as part of the weight of evidence evaluation. Results from these studies can provide important information on dose response and adverse effects on endpoints of potential concern. The term 'potential endocrine disruptor' could be easily misinterpreted, and generally use of this term should be avoided. From a scientific perspective, it is important to determine the overall weight of evidence of the performance of a substance in the screening assays/battery, as described above.

25. Proposed OECD Stage 3: Definitive Testing

The purposes of the proposed OECD Stage 3: Definitive Testing are to accurately and effectively identify and characterize the hazard(s) from chemicals. The tests comprising the proposed OECD Stage 3: Definitive Testing has been specifically developed to assess reproductive and developmental toxicity. The results obtained should be assessed relative to target site toxicity, and the endpoints evaluated include both endocrine and non-endocrine toxic endpoints. The array of tests included in Table 4 should be viewed as a matrix of available options, and not as a sequential list of assays and tests. It would not be necessary to conduct all tests, but instead, from this matrix, the appropriate test could be selected. In the interests of flexibility and minimizing animal and resources, for example the enhanced repeat dose study and the shorter-scope reproduction/developmental tests would not be required in cases where a longer scope test is already available or planned.

Table 4. Proposed OECD Stage 3: Definitive Testing for Adverse Effects and Dose Response (for use in hazard identification and risk assessments)

LEVEL	DESCRIPTION	ASSAYS FOR MAMMALIAN TOXICITY	ASSAYS FOR ECOTOXICITY
Stage 3: Definitive Testing (evaluation of apical endpoints, adverse effects and dose response for hazard identification and characterization) OECD 2002 Framework Level 5	Reproduction/developmental tests – shorter scope - includes <i>in utero</i> exposure, developmental, and reproductive capacity endpoints	(TG407* depending on the exposure situation as the method does not include the reproductive phase) Reproductive /developmental screening test (TG 421) Combined repeat dose with reproduction / developmental screening (TG 422) One generation reproductive toxicity (TG 415) Two generation reproductive toxicity (TG 416) [Enhanced one generation reproductive toxicity -if and when a final OECD TG is developed]	Partial and full life cycle assays in fish, birds, amphibians and invertebrates (developmental and reproduction)

* Remark: Depending on the situation a TG-407 would suffice to establish a NOAEL in selected instances. Such instances could include, for example, substances with low potency, minimal human exposure likely intermediates or substances manufactured in closed system, and limited potential for environmental release. This would serve to focus the more extensive testing only on substances that have high exposure potential.

26. A defensible hazard characterization for hormonally active chemicals requires not only summarizing toxicological screening and testing data (hazard identification), but also requires an objective evaluation of whether the effects produced are adverse and whether adverse effects are due to a hormonal activity of the chemical. Hazard characterization is based on overall consideration of the weight of evidence. This includes consideration of the proposed OECD Stage 2 Screens and proposed OECD Stage 3 Definitive Tests, and results from standard toxicity studies (e.g., Repeat Dose Study (TG 407/408), Reproduction/Developmental Toxicity Study (TG 414/421/422), Chronic Toxicity Study (TG 452/453) etc). Such an evaluation provides the context for considering adverse effects across organ systems and modes of action and for comparing NOAELs. However, hazard identification is insufficient to characterize risks.

27. Risk characterization requires integration of scientific data and knowledge of hazard, dose-response and exposure, as well as an evaluation of the foundations of the hazard data, inferences drawn from the data, and inherent uncertainties. In cases of low potency and low or negligible actual and potential exposures, test methods such as the Repeat Dose Study (TG 407) or the Reproduction/Developmental Toxicity Screening Tests (TG 421/422) could be used to provide dose-response data of effects on apical endpoints. This would serve to focus the more extensive testing only on substances that have high production volume and the highest potential for human and ecological exposures. In all cases, results from Stage 3: Definitive Testing outweighs or supersedes results from Stage 2: Screening.

28. In developing the OECD framework for evaluating endocrine activity, all have recognized the need to employ the OECD must use standardized, validated and internationally harmonized test methods, and this has been the foundation of the OECD EDTA and WNT work over the past 10 years. In considering the screens and tests for evaluating endocrine activity, the OECD EDTA determined that certain types of studies were not standardized and lacked adequate data for a validation determination. Hence, the OECD EDTA formed the validation management groups to coordinate the necessary laboratory studies to achieve standardized and validated test methods. Considerable progress has been made in this regard, due to the efforts of industry, government, academia and laboratories.

29. Research laboratory studies using novel test methods, non-standardized, and not yet validated methods and / or non-standard test species generally lack the quality criteria that typically encompassed in studies employed using OECD Test Guidelines and Good Laboratory Practices, and thus such novel methods or non-standard test species must be evaluated with due caution for regulatory purposes. As basic research of endocrine mechanisms advance, new and novel scientific methods have, and will continue to be reported. These new and novel types of studies are significantly different from laboratory studies using standardized and validated techniques. For example, they may lack appropriate documentation for reliability of the test method performance or unambiguous interpretation of the relevance of the endpoints evaluated. Thus, such novel research studies should not be used for regulatory action, but should trigger further research and/or method validation efforts.

30. Validation of a test method is a required prerequisite for it to be considered for regulatory use. For a new or revised test method to be considered validated for regulatory purposes it should meet the criteria specified by OECD GD 34 (2005) and/or ICCVAM (1997)). For validation, the extent of a test method's variability and reproducibility within and across laboratories must have been demonstrated, and sufficient data provided to permit assessment of the method's range and limitation of application.

31. The issues of GLP, quality assurance and quality control have important international dimensions, as recognized by OECD. Under Mutual Acceptance of Data, regulatory authorities in OECD member countries can rely on safety test data developed abroad, thus eliminating duplicative testing, and this serves to both enhance laboratory animal welfare and to create testing efficiencies while at the same time assuring high quality scientific information is developed for regulatory and product stewardship purposes. Screening and testing of substances using validated test methods yields greater confidence in results compared to studies performed using non-validated methods. The requirement across OECD member countries of mutual acceptance of data is based on the foundation that evaluations of chemicals are based on test data of sufficient quality, rigour and reproducibility.

32. As scientific methods advance, regulatory bodies will continue to be challenged to review and interpret new and novel methods and studies with non-standard species. The responsible standard when dealing with results from such studies, whose findings may be especially influential, is the standard which is embodied in the scientific method – hypothesis formulation, hypothesis testing through experimentation, and independent replication. The OECD EDTA Framework should endorse and integrate this responsible standard concept into the OECD globally harmonized approach for evaluation of the endocrine activity of chemical substances

33. New and novel methods, and studies with non-standard species, may provide important scientific information. Recognizing this fact, we suggest the following course of action for addressing such study reports. The first order task should be a thorough review of the study report. This would entail independent review of the methods and procedures and review of the data. The study documentation should be sufficient to permit independent verification/calculation of results, and the methods should be sufficiently described to permit replication of the study procedures by other laboratories. Following this review, it is recommended that an additional laboratory replicate the study. Then, if the findings are shown to be reproducible, then two courses of action would be advised:

A) The new or novel test method could be subjected to standardization and validation within the OECD TG program (EDTA) or within a similar formal program sponsored by a national government or recognized scientific organization (e.g., ISO, ASTM), provided that such a method is viewed as necessary to augment or replace one or more existing test methods in the screening and testing battery. International harmonization should also be an objective, to promote mutual acceptance of data and to optimise utilization of laboratory animals. To the extent possible, in a manner similar to that of the OECD EDTA, it is beneficial for lead organizations to undertake a certain amount of standardization and pre-validation work, and for information to be

shared across sectors (government, industry, academic and other scientific researchers). Organization & conduct of the formal test method validation effort should be conducted within the existing processes of the relevant competent authorities; or

B) The substance of concern could be evaluated in one of the wide variety of existing validated test method using standardized OECD TG methods and species (or similarly validated scientific methods, for example those promulgated by ISO, ASTM, US EPA). Results of this study would then be evaluated within the tiered hierarchical OECD EDTA Framework. In general, this would be the preferred course of action.

34. While review of the totality of relevant scientific information is needed for robust regulatory decision making, in regulatory reviews the weight afforded to results from standardized and validated test methods and laboratory studies conducted under GLP generally exceeds the weight given to novel and non-validated, non-GLP studies. Until such time as either course A or B in paragraph 33 is completed, it would be imprudent to initiate extreme risk management actions based upon a report of a new, novel or non-validated test method. Experience has shown that it is inappropriate to automatically accept novel and new research reports for regulatory purposes. At times, even peer review by scientific journals won't suffice to screen out flawed studies, a fact that should counsel against overly enthusiastic acceptance of any single study – especially one whose results are both extraordinary and groundbreaking. If the results of single studies seem especially important special care should be taken to verify the quality of such studies. It is also important to consider the scope of peer review. For the vast majority of scientific journal publications, peer reviewers in fact do not examine the detailed laboratory study records. Instead, this system relies on the submitting author's summarization of experimental procedures and results. If the results of single studies seem especially important special care should be taken to verify the quality of such studies.

REFERENCES

J. Miyamoto and J. Burger (2003) Implications of endocrine active substances for humans and wildlife: Executive summary. Pure and Applied Chemistry Vol. 75, Issue 11-12, xv-xxiii. (<http://www.icsu-scope.org/projects/complete/endocrineexecsum.htm>)

Veldhoven workshop. SETAC 1997 SETAC (1997) SETAC-Europe/OECD/EC Expert Workshop on Endocrine Modulators and Wildlife: Assessment and Testing (EMWAT). Veldhoven, The Netherlands, 10-13 April 1997. Editors L Tattersfield, P Matthiessen, P Campbell, N Grandy and R Lange. SETAC –Europe Publications, Brussels

Weybridge EC 1996. European Workshop On The Impact Of Endocrine Disrupters On Human Health And Wildlife. 2-4 December 1996 Weybridge, UK. Report Of Proceedings [REPORT EUR 17549]

EMSG. Towards the establishment of a weight of evidence approach to prioritizing action in relation to endocrine disruption. Endocrine Modulation Steering Group, LRI-CEFIC: Brussels. (2000); Attachment 1.

H-P. Gelbke, A. Hofman, J. Owens and A. Freyberger. The enhancement of the subacute repeat dose toxicity test OECD TG 407 for the detection of endocrine active chemicals: comparison with toxicity tests of longer duration. *Archives of Toxicology* 81:227-250, 2007

OECD GD 34 (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (http://www.oecd.org/document/30/0,3343,en_2649_34377_1916638_1_1_1_1,00.html)

ICCVAM (1997). Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (NIH 97-3981).

http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf

Attachment 1. CEFIC EMSG Paper: Towards the Establishment of a Weight of Evidence Approach to Prioritizing Action in Relation to Endocrine Disruption

SUMMARY

Large uncertainty still remains about actual risk posed by endocrine disrupting substances and the scientific assessments that should be included when developing risk management options. The position of CEFIC is that a weight of evidence approach using the most credible scientific data should be used in making decisions.

A number of publications describe how to evaluate data quality and their use in hazard and risk assessment (e.g. Ref. 1 & 2). This document builds on this earlier work and summarises the key elements of a procedure to evaluate the balance of scientific evidence in relation to the potential of a substance to cause adverse effects through disruption of the endocrine system. It addresses the issues of data relevance, quality and significance - using a weight of evidence approach to indicate whether, and what action needs to be taken in order to assess the hazards and risks of a substance. It has been developed specifically to enhance the prioritisation process and output of the EU DG Environment project: "Towards a Priority List of Substances for Further Evaluation of their Role in Endocrine Disruption."

The procedure includes:

- A data collection step that covers a search of the published literature and an extension of the search into unpublished literature, particularly to gather data used for regulatory purposes.
- An evaluation step that considers:
 1. What endpoint has been measured and the relevance of that endpoint to the effects of potential endocrine disruption mechanisms.
 2. The repeatability, reliability and quality of a particular study and its protocol, together with the extent of peer review.
 3. The significance (or 'weight') of the data based on the assessments under 1 & 2 above.
 4. Whether there is sufficient coherence of the data to draw conclusions (balance of the 'weight of evidence')
 5. What further evidence is required, including a prioritised action identification step leading to risk assessment in accordance with the existing, or any future coherent chemicals regulatory framework.

1. INTRODUCTION

The European Commission has completed the initial stages of a project through DG Environment to prepare a “Priority List of Substances for Further Evaluation of their Role in Endocrine Disruption.” This exercise required the evaluation of toxicological data in order to achieve a prioritisation rating, but the Chemical Industry believes that the approach taken to create the initial list was too superficial to add meaningfully to the debate and that the list may be misinterpreted.

The process to develop the list used an “evidence of suspicion” approach in which the presumption of endocrine toxicity may be based upon as little as a single data point. Studies showing consistently that there is no evidence of endocrine toxicity have been ignored, irrespective of quality, since they do not add to the strength of suspicion.

Furthermore, the ‘List’ adds nothing to the debate because it fails to identify and incorporate the priority actions required to assess ED hazards and risks properly. It also fails to present a strategy for assessing all other substances for which there is little or no data to judge the ED hazard. It is merely another list of often poorly founded suspicions that, because of its apparent ‘official’ status and pseudo-scientific analysis, may be misinterpreted as a ‘Definitive List of Endocrine Disrupters’. Failure to take all available data into consideration could well lead to economically damaging de-selection of products without protecting human health or the environment.

Despite these well founded concerns, the European Chemical Industry accepts the political desire to develop a list of priority substances for further evaluation of their role in endocrine disruption but believes that the interests of both public and environmental health would be better served if the ‘List’ was to be more ‘action’ orientated and based on sound scientific principles.

Through this document, CEFIC offers an approach that weighs both the relevance and reliability of evidence in the balance using scientifically based criteria to identify the priority actions required for each substance under suspicion and indeed, any other substance that may come under review - truly providing a priority action plan towards risk assessment within the processes laid down under existing community regulations. We have termed this approach as the ‘Weight of Evidence Approach’ and it fits between an initial step to define a group of chemicals requiring evaluation and later steps to fill data gaps and undertake risk assessment. This is shown graphically in Figure 1 on the next page.

At a European Union meeting in Weybridge (Ref. 3), the following definitions were agreed:

“An endocrine disrupter is an exogenous substance that causes adverse health effects in an intact organism, or its progeny, consequent to changes in endocrine function.”

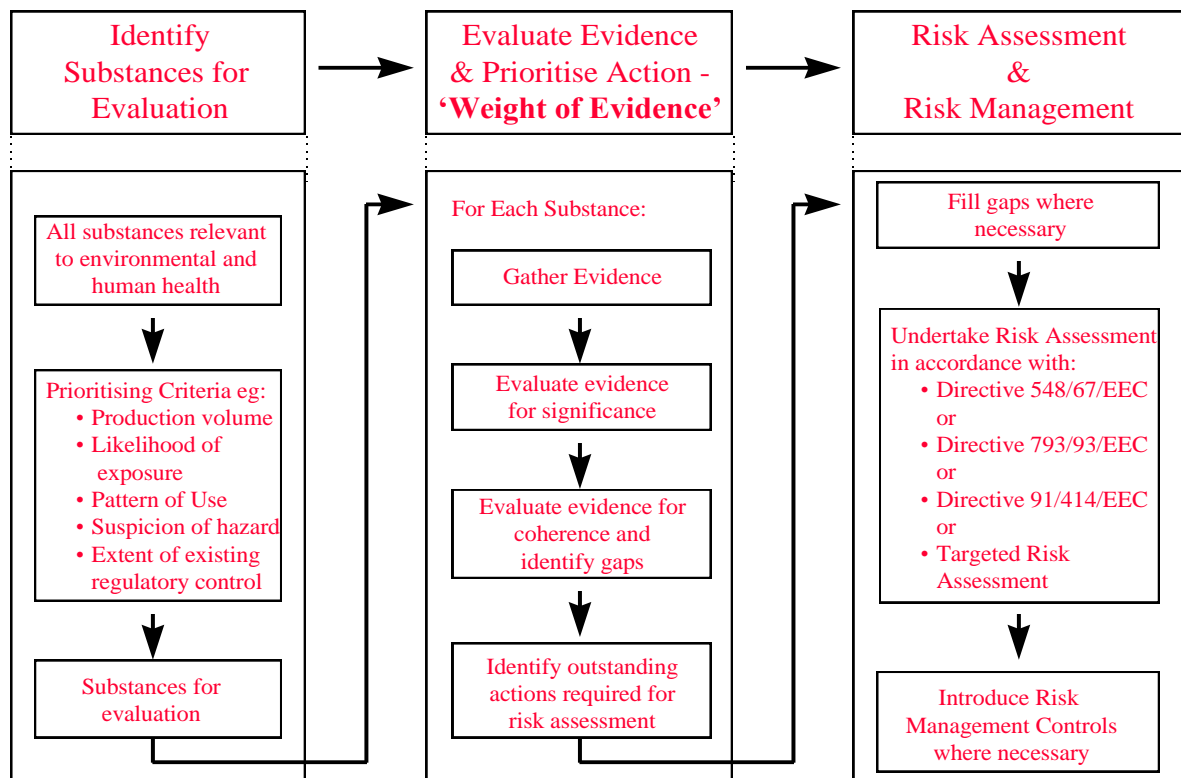
“A potential endocrine disrupter is a substance that possesses properties that might be expected to lead to endocrine disruption in an intact organism.”

These definitions have gained wide acceptance in the international arena, have been adopted by the International Programme for Chemical Safety (IPCS) and consequently, have been used as the basis for this set of proposals.

In agreement with the aim of the European Commission project, the procedure developed works towards prioritising actions required for the further assessment of substances in relation to their potential endocrine disrupting activity. In this context, the phrase ‘consequent to’ is interpreted to mean demonstration of a causal link between mechanistic activity and adverse health effects.

It specifically addresses six potential mechanisms - agonistic and antagonistic effects on the oestrogen, androgen and thyroid systems. However, where relevant, it also makes provision for reporting non-endocrine adverse effects so that risks from other sources are not ignored.

Figure 1: The Role of A Weight of Evidence Approach



2. WEIGHT OF EVIDENCE APPROACH

The following approach was designed to assist in the conduct of a weight of evidence review of available toxicological data in order to enable the identification and prioritisation of chemicals for further assessment in relation to endocrine related activity. It consists of the two basic tasks shown below:

2.1 Collecting the data.

2.2 An evaluation step that considers:

2.21 What endpoint has been measured and the relevance of that endpoint to the effects of potential endocrine disruption mechanisms (**Data Relevance**).

2.22 The repeatability, reliability and quality of a particular study and its protocol, together with the extent of peer review (**Study Repeatability**).

2.23 The significance (or 'weight') of a data set based on the assessments under 2.21 & 2.22 above (**Data Significance**).

2.24 Whether there is sufficient coherence of the data to draw conclusions (balance of the 'weight of evidence'), what further evidence is required to take action and what that action should be. (**Coherence, Gaps and Framework for Further Action**).

Expert judgement is required at each stage and it is important to record the basis of decisions to aid transparency (See Section 3).

It should be emphasised that none of these proposals are new. Such an approach is well accepted and documented in peer reviewed journals (e.g. Refs. 1, 2 and 4).

2.1 Data Collection

In order to ensure that as many data as possible are included in the assessment, an extensive search of all relevant databases is required. This should capture any data available in SIDS, IUCLID and other relevant databases, as well as in the published literature. Criteria for the search and organisation of the search results should be based on expert judgement, and developed on a case-by-case basis, details of which should be recorded. (See Section 3). The literature search should, as a minimum, include those commercially available databases listed in Appendix 1. Consideration should also be given to seeking unpublished literature, particularly data used for regulatory purposes - but only where the quality can be assessed under Section 2.23.

2.2 Data Evaluation

2.21 Data Relevance

There are various assays, measures and toxicological endpoints that are claimed by different sources to be relevant to the assessment of endocrine disruption. In reality, this field is still in an early stage of development and there is uncertainty

regarding the significance of many of the findings, especially the relevance of *in vitro* assays and short-term screening assays to toxicological effects.

A weight of evidence approach should be able to differentiate between various toxicological endpoints in relation to their relevance to mechanistic evidence and observed effects. For the approach described here, endpoint relevance has been weighted to enable a hierarchy which can differentiate between:

- Observed adverse health effects with mechanistic support to establish causal linkage
- Observed adverse health effects with limited understanding of mechanism
- Biomarker of exposure
- Mechanistic potential with no observed effect

Substances should only be considered endocrine disruptors if they cause “adverse health effects in an intact organism, or its progeny, consequent to changes in endocrine function” (Ref. 3). Hence, it is inappropriate to assess a substance as an endocrine disruptor on the basis of mechanistic *in vitro* assays alone and the approach has been designed to reflect this.

Similarly, many current testing criteria exist for the *in vivo* determination of adverse effects on reproduction and/or development without providing evidence of mechanistic cause. Under these circumstances, a negative result may be sufficient to demonstrate that a substance is not an endocrine disruptor, but a positive result may need further testing to distinguish the mechanistic cause. Nonetheless, those with a financial interest in the substance may feel that it is more prudent and efficient to proceed directly towards risk assessment - rather than undertake additional testing - on the conservative assumption of an endocrine cause.

For non-standard protocol endpoints, the assessment of endpoint relevance would in many instances be a subjective decision which should be based on sound expert judgement. If such a judgement proves impossible, then the data should be treated as being of ‘low significance’ (See Section 2.23) until such time that additional research is able to clarify the relevance to risk for species known to be exposed to the substance in question.

In all instances, the relevance rating would need to be clearly documented with appropriate justification. Adverse effects identified but thought to be of non-endocrine origin should be reported for further assessment by the relevant Competent Authorities.

Assessing Relevance of *In vivo* Data

The most relevant data for reaching an evaluation of endocrine toxicity is found from repeat dose toxicity and/or reproductive toxicity studies which include measurements and observations associated with endocrine toxicity. Other types

of *in vivo* studies, including screening assays such as the uterotrophic and Hersberger assays do provide relevant information, and data from such studies should be included in the any weight of evidence review. It must be remembered that, positive results in screening assays are not conclusive evidence of adverse health effects and are of lower relevance than the repeat dose studies in making a judgement about endocrine disruption. Nevertheless *in vivo* screening assays do serve a useful purpose by indicating potential for harm and should be regarded as “indicative studies” leading to actions as indicated in Figure 2.

A summary of the relevance of *in vivo* studies is shown in Table 1

Table 1: Relevance of In Vivo Assays	
High relevance	<ul style="list-style-type: none"> <input type="checkbox"/> endpoint(s) in a multi-generation test or other repeat dose toxicology test that is specifically controlled by the endocrine system, or <input type="checkbox"/> parallel dose-responsive changes in hormone levels in the presence of consequent toxicological effects (mammalian only) <input type="checkbox"/> negative data from a short term/screening assay specifically controlled by the endocrine system
Medium relevance	<ul style="list-style-type: none"> <input type="checkbox"/> endpoint in a multi-generation test, or other repeat dose standard toxicology test, which may be influenced by the endocrine system, but is also known to be affected by other factors, e.g. water quality, environmental stress, toxicity etc.; or <input type="checkbox"/> endpoint in a short-term/screening assay specifically controlled by the endocrine system; or <input type="checkbox"/> changes in hormone levels in the absence of any toxicological effects (mammalian only)
Low relevance	<ul style="list-style-type: none"> <input type="checkbox"/> evidence indicates that the endpoint is not controlled by the endocrine system. Positive results of adverse effects should be reported for further risk assessment.

Assessing the Relevance of *In vitro* Data

The purpose of *in vitro* testing is basically to identify intrinsic endocrine modulation potential and determine potency relative to a reference hormone. For example, “can a substance bind to a receptor?” and “what amount is required to produce an equivalent response to a natural hormone such as oestrogen?”. As the predictive ability of *in vitro* tests to detect effects in animals is, at best, uncertain it must be recognised that results from *in vivo* assays are more relevant for judging whether or not a substance will cause endocrine toxicity.

Despite such limitations, *in vitro* tests can be reliable for detecting potential endocrine modulating activity *per se* and therefore are a useful tool in the overall context of endocrine toxicity testing.

A number of *in vitro* screening systems are available which involve the interaction of chemicals with vertebrate steroid receptors. Although the number of *in vitro* assays for taxa other than mammals is limited, receptors, such as for oestrogen, androgen, and thyroid, and their essential roles are conserved across vertebrates. The endocrine systems of invertebrates are poorly understood. The role of oestrogen and other vertebrate hormones, if any, in invertebrates is unclear, and will not be further discussed here.

It is recommended that the data review incorporates all available *in vitro* data, and that for the purposes of assessing the relevance of *in vitro* endpoints, attention should be focused on both a hierarchy of information and the quality of the particular measurement system:

- whether the assay is designed to indicate simple receptor binding potential or the more indicative receptor binding coupled with transcriptional activation.
- whether the assay is a cellular or subcellular assay, which would be indicative of whether or not the endocrine receptor was likely to be exposed to metabolites of the parent compound.
- whether the assay examines relevant endocrine parameters such as steroid metabolism.

On the basis of the above discussion, a hierarchy of *in vitro* endpoint relevance is proposed in Table 2.

High relevance	<input type="checkbox"/> endpoint is based upon receptor binding potential coupled with transcriptional activation, whole cell or subcellular assay; or <input type="checkbox"/> receptor binding potential in a whole cell assay <input type="checkbox"/> assessment of steroid metabolism in a whole cell assay
Medium relevance	<input type="checkbox"/> endpoint is based on receptor binding activity in a subcellular assay, or <input type="checkbox"/> endpoint is based on cell growth or other endpoint not a direct measurement of receptor mediated activity <input type="checkbox"/> endpoint of steroid metabolism in a subcellular assay
Low relevance	<input type="checkbox"/> not applicable

It should be noted that the hierarchy is solely for the relevance of the endpoint, and is not indicative of the final weighting applied to the result. The weight of the evidence procedure is described in Section 2.23.

2.22 Study Repeatability

An assessment of study repeatability takes into account:

- The extent to which protocols have been validated and the limits within which conclusions can be drawn
- The extent to which the toxicological endpoints are understood
- The extent of the historical database and the confidence that this provides
- Basic experimental design - adequacy controls; suitability of concentration range
- Exposure data - purity of test material, verification of exposure concentrations
- Test species - suitability, general health, environmental conditions
- Analysis of results - statistical validity of observed effects
- Transparency of the study report

It is essentially, an assessment of the confidence one might have in being able to repeat the study and reproduce the results.

Traditionally, toxicity work has been evaluated against compliance with internationally recognised and validated standard protocols (e.g. ASTM, ISO, OECD). Such studies can be repeated with a high level of confidence. Evaluation of protocols for the determination of endocrine disruption is difficult, since standard protocols are not currently available for this specific area. Nonetheless, many of these standard tests shed light on the adverse effects likely to result from endocrine disruption and their results can be relied upon to provide useful evidence.

Other, perhaps more novel protocols may produce endocrine-specific information, but their reliability needs careful evaluation. Proposed criteria for reported data are listed in Appendix 2, and have been selected as criteria which are indicative of work which has been undertaken to a good standard of scientific practice.

It is proposed that tests carried out in accordance with these criteria form a suitable basis for assessing substances, when combined with a weighting based on the relevance of the endpoint, as previously described in Section 2.21

On this basis, it is proposed that the hierarchy for study repeatability should be ranked as follows in Table 3:

Table 3: Hierarchy of Repeatability

High Confidence of repeatability	All criteria for the experimental design and conditions, and for reporting transparency are met <ul style="list-style-type: none"> • <i>full details of experimental method available and these indicate that studies have been carried out to an acceptable standard</i>
Medium Confidence of repeatability	The main criteria for the experimental design and conditions, and for reporting transparency (see bolded points of attached Appendix 2) are met <ul style="list-style-type: none"> • <i>some details of the experimental method are available which indicate that studies have been carried out to an acceptable standard</i>
Low Confidence of repeatability	Insufficient information is available for the experimental design and conditions to determine reporting reliability
Unreliable	Analysis of the experimental design and conditions indicate that the study may be unreliable or not reported transparently.

2.23 Data Significance

The final task in establishing the ‘weight’ that should be ascribed to any set of data takes into account both the ‘Relevance’ and ‘Repeatability’ of the data as evaluated in Sections 2.21 & 2.22. In effect, the ‘weight’ is measured as the level of significance that can be ascribed to a data set in reaching conclusions about endocrine disruption.

As discussed above *in vivo* data from repeat dose/long term animal studies are the most important in hazard assessment. While *in vitro* information and data from *in vivo* screening studies are useful in making judgments about the presumption of hazard they are not currently linked directly to, or are predictive of adverse/toxicological effects associated with endocrine disruption.

For these reasons *in vivo* data from repeat and chronic* studies examining functional endpoints such as growth, reproduction and development during critical life-stages are considered more significant in assessing the potential for adverse effects and making risk management decision than *in vitro* data. The latter can only provide information about one or two steps in a chain of

events that may, or may not lead to health problems. At best, such results can be taken as being only 'Indicative'.

* In this paper, the term 'chronic' is used for all studies of 28 days exposure or longer and reproductive investigations.

This paper proposes 4 levels of significance that might be ascribed to a data set:

- High Significance
- Indicative Significance
- Low Significance
- Unusable

These terms are used to calibrate the **level of significance** that can be placed on *in vivo* and *in vitro* data as described below:

Assessing The significance of *In Vivo* Evidence

Table 4 displays the evaluation of 'Significance' for *in vivo* data and is based on the following basic principles:

- As tests for chronic effects are the most relevant, if the effects are of High Relevance, studies of Medium and High Reliability should be considered as of High Significance.
- As the overall significance of screening tests is lower than chronic tests, *in vivo* screening endpoints of High Relevance from studies of Medium and High Reliability should be considered as only of Indicative Significance.
- If the effects from a chronic study are of Medium Relevance, studies of Medium and High Reliability should also be considered as only of Indicative Significance.
- Screening studies of only Medium Relevance, but of Medium and High Reliability should be considered as of Low Significance and used merely as supporting information.
- Data from studies considered as of Low Reliability or as Unreliable should be considered as Unusable.

Table 4: *In vivo* Data Significance

		Study Repeatability			
Endpoint Relevance		Unreliable	Low	Medium	High
Chronic	High	<i>Unusable</i>	<i>Unusable</i>	<i>High</i>	<i>High</i>
test	Medium	<i>Unusable</i>	<i>Unusable</i>	<i>Indicative</i>	<i>Indicative</i>
Relevance	Low	<i>Unusable</i>	<i>Unusable</i>	<i>Positive effects to be reported</i>	<i>Positive effects to be reported</i>
Screening test	High	<i>Unusable</i>	<i>Unusable</i>	<i>Indicative</i>	<i>Indicative</i>
Relevance	Medium	<i>Unusable</i>	<i>Unusable</i>	<i>Low</i>	<i>Low</i>

Assessing The Significance of *In Vitro* Evidence

Table 5 is based on the following basic principles for assessing the significance of *in vitro* studies;

- No *in vitro* study can be considered as being of High Significance. At best it can be only 'Indicative' of mechanistic potential. However, a negative result of 'Indicative Significance' would be sufficient to be definitive.
- Only studies meeting both a high reliability and a high relevance should be assessed as being of 'Indicative significance'.
- Studies with a medium reliability and a high relevance, or vice versa should be assigned a 'Low significance' - for support purposes only.
- Data from unreliable studies or those with low reliability should be considered as unusable.

Table 5: *In vitro* Data Significance

Endpoint Relevance		Study Reliability			
		Unreliable	Low	Medium	High
Receptor	High	<i>Unusable</i>	<i>Unusable</i>	<i>Low</i>	<i>Indicative</i>
Relevance	Medium	<i>Unusable</i>	<i>Unusable</i>	<i>Unusable</i>	<i>Low</i>
Metabolic	High	<i>Unusable</i>	<i>Unusable</i>	<i>Low</i>	<i>Indicative</i>
Relevance	Medium	<i>Unusable</i>	<i>Unusable</i>	<i>Unusable</i>	<i>Low</i>

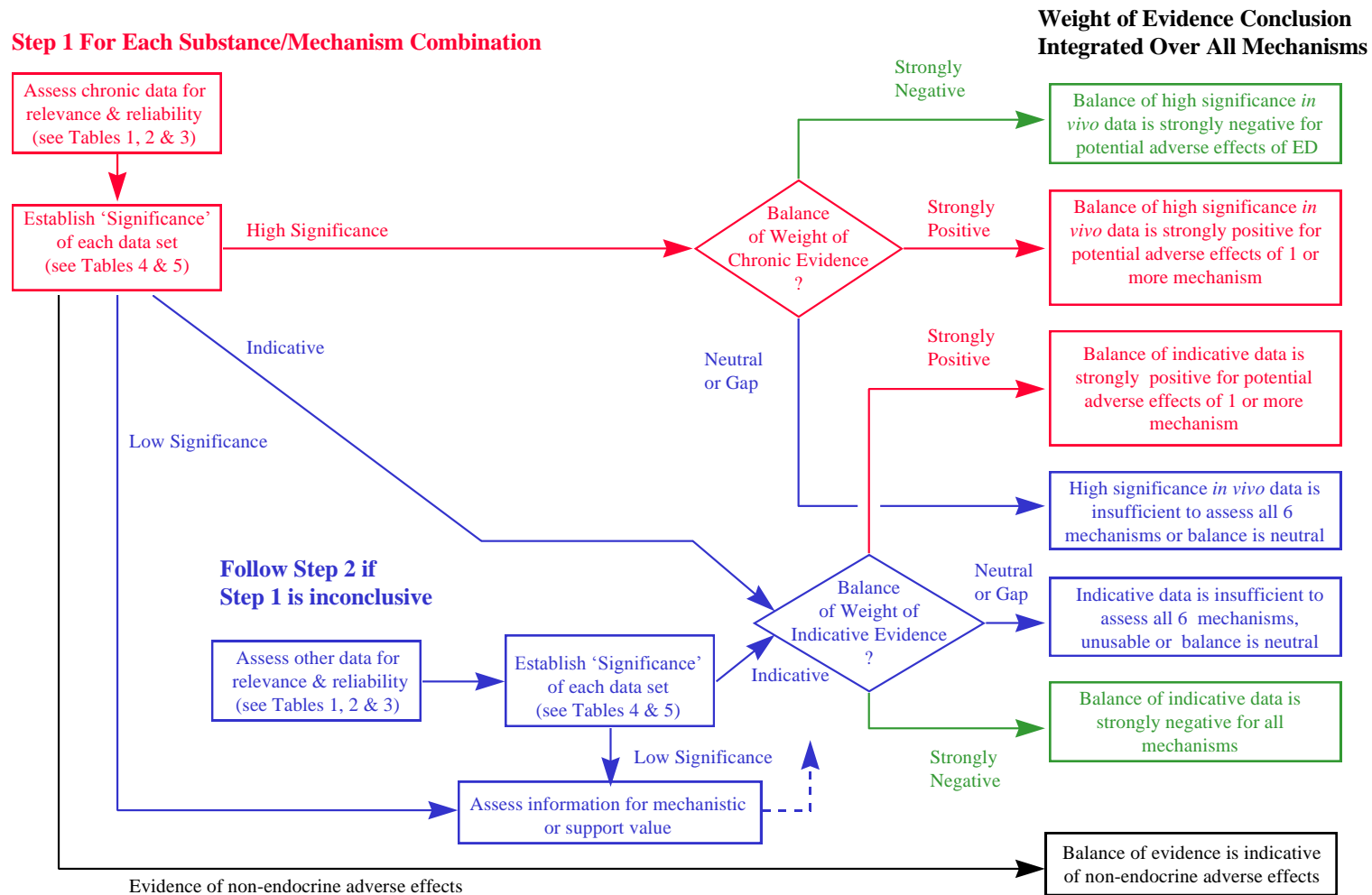
Use of Significance Assessments

Assessments of Significance are used in the process shown in Figure 2 (to be found on the next page*). It shows a 2-step process to be applied to each mechanism and is based on the premise that only evidence of ‘*In Vivo* High Significance’ can be considered as being definitive in the 1st step. Any other *in vivo* data must be considered alongside *in vitro* data in the 2nd step as ‘indicative’ or as ‘supporting’ evidence only.

It is only necessary to proceed to the 2nd step if the 1st step is inconclusive.

* Underlying Premises and Assumptions for Figures 2 and 3 can be found in Table 6.

Figure 2: Assessing & Weighing The Balance Of Evidence



Endocrine Disruption

An Approach For Prioritising Action Based on A Weight of Evidence Approach

Table 6: Premises and Assumptions Applied to Figures 2 & 3

- The scheme shown in Figures 2 & 3 is based on a focused evaluation of substances in relation to the adverse effects that may result from 6 mechanisms:
 - Oestrogenic
 - Anti-oestrogenic
 - Androgenic
 - Anti-androgenic
 - Thyroid
 - Anti-thyroid
- Where an adverse effect is identified, but resulting from other mechanisms, then this should be reported and investigated as part of the more general risk assessment of the substance.
- All evidence of less than ‘high significance - *in vivo*’ is considered as of ‘screening value only’ or as ‘unusable’.
- If sufficient is known about appropriate dose ranges, then dose/response testing could be implemented immediately after screening, thus reducing the overall amount of animal testing required.
- The scheme assumes that OECD Tier 1 tests for the 6 mechanisms above will be available soon for screening and that enhancements for multi-generation testing to cover the relevant end-points will have been agreed and validated as an OECD Tier 2 Test soon after.
- Priority is shown in colour: **Red** for ‘high priority’; **Blue** for ‘medium priority’; **Green** for ‘no further action’.

2.24 Coherence, Gaps and Framework For Further Action

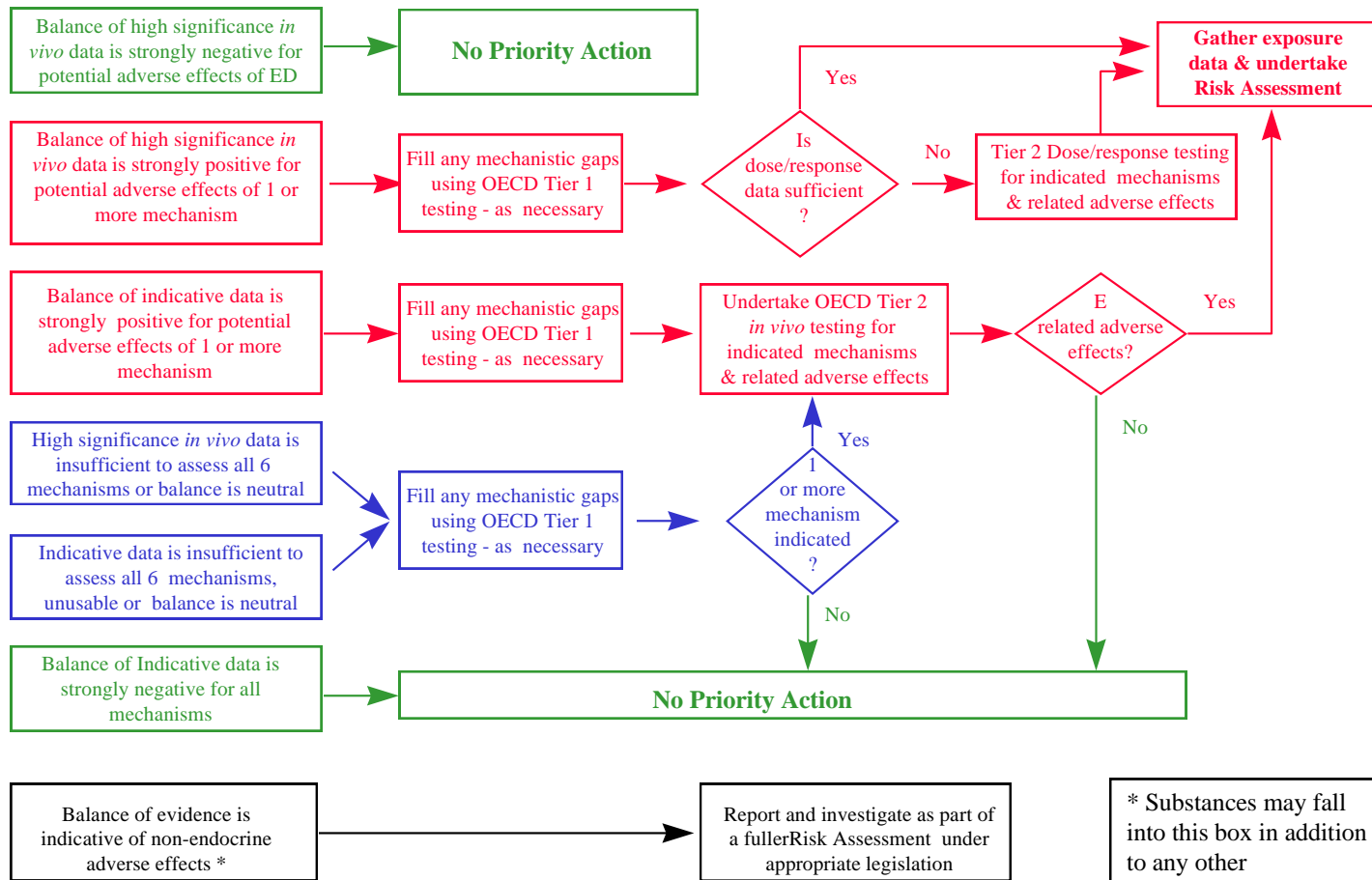
Once all relevant data have been evaluated for significance to all 6 mechanisms in accordance with the procedure outlined in Section 2.23, it should be possible to assign each substance to one of the right hand boxes in Figure 2 and to identify gaps in knowledge that need to be filled.

Simplistically, if all of the data fall into one of the boxes described in Figure 2, the substance could be actioned as proposed in Figure 3 (see next page). For example, if all high significance chronic data fall into the top-right box of Figure 2, then taking this forward into Figure 3, the procedure proposes that there is no need for further action and the substance should be removed from the ‘List of Priority Actions’. Alternatively, should all data available fall into the second box, then again, going forward to Figure 3, the recommendation is for urgent risk assessment.

In the event that High Significance - *in vivo* dose response data covering the relevant end-points associated with all six mechanisms exists already, then it would be possible to jump directly to risk assessment without undertaking.

Figure 3: Weight of Evidence Derived Action Scheme

Weight of Evidence Conclusion



further testing. However, whilst this may be sufficient to assess the risk, it may leave some ambiguity about the mechanism. Those with an economic interest in the substance should judge whether additional mechanistic evidence can provide added value.

In the event that the available data can only shed light on some of the mechanisms, then it is recommended that gaps are filled initially using the OECD Tier 1 tests. This will then allow the Tier 2 tests to be designed to cover all the mechanisms of concern without unnecessary test complexity.

In practice, many substances may be positive for some mechanisms and negative for others - leading to the actions shown. Furthermore, the data may not be coherent - even for studies that are considered to be of high significance. Clearly, if there is only one 'odd' study among many of similar significance, then one would be able to draw a conclusion based on the 'balance of the weight of evidence'. However, if the balance is neutral or close to neutral, then it will be necessary to undertake additional high quality studies for one or more mechanism to draw definite conclusions.

3. REPORTING

Criteria for the information search should be recorded in the report to ensure future duplication of the search is possible. Exclusion criteria should be incorporated into this record to explain why individual references were not considered for further examination.

Decisions taken under Sections 2.21, 2.22 & 2.23 should be recorded and justified.

Actions should be recorded for substances in accordance with Figure 3.

The report would, of necessity, incorporate an extensive list of all data considered, coupled with an explanation, as described in Tables 1-5 and Figures 2 & 3, as to how the final conclusions and recommendations for action were obtained.

4. CONCLUSIONS

The European Chemical Industry recommends that the weight of evidence approach is included in the process to identify a Priority Action List. This will ensure decisions are made on a complete evaluation of information rather than on a partial assessment, such as the 'evidence of suspicion' approach, which really does little more than count "positive" studies.

The inclusion of the weight of evidence approach introduces scientific rigour into the process for developing a Priority Action List. The actions proposed are specific and truly prioritised in terms of urgency. Furthermore, it provides a platform for evaluating a much larger group of chemicals for which little or nothing is known at the moment.

While the procedure can be applied immediately to existing data, it will quickly become clear from the resulting analysis that there are many data gaps. Furthermore, there is a need to conduct research which will aid better assessment and management of endocrine disrupting chemicals. An important component of this research is in the area of testing methodology and the Commission is most strongly urged to provide much needed support for the OECD testing initiative. While improvements and enhancements of testing protocols will build a better understanding of endocrine toxicity it is nonetheless expected that endocrine disrupting chemicals can be addressed adequately within the current risk assessment/management framework with only minor adjustments to include any new knowledge or enhancements of testing protocols.

REFERENCES

1. Klimisch, H-J., Andreae, M. and Tillmann, U. - A Systematic Approach For Evaluating The Quality of Experimental Toxicological and Ecotoxicological Data. *Regulatory Toxicology and Pharmacology* **25**, 1-5 (1997)
2. Rosner, G. - Validierung von SDS: Standardbegründungen - Fraunhofer-Institut Für Toxikologie und Aerosolforschung. (1994)
3. European Workshop on the Impact of Endocrine Disrupters on Human Health and Wildlife (Weybridge, UK, 1996). European Union Report EUR 17549
4. Office of Science and Technology Policy. Chemical Carcinogens; A Review of the Science and its Associated Principles. (1985)

APPENDIX 1

DATABASE SEARCH SITES

Aquatic Sci&Fish Abs (c) 1998 FAO (for ASFA Mv Brd)

BioBusiness(R) (c) 1998 BIOSIS

BIOSIS PREVIEWS(R) (c) 1998 BIOSIS

CA SEARCH(R) (c) 1998 American Chemical Society

CAB Abstracts (c) 1998 CAB International

ChemEng & Biotec Abs (c)1998 RoySocChm,DECKEMA,FizChemie

CHEMTOX (R) Online (c) 1998 MDL Info Systems

CHRIS Chemical Hazards response system

Current Contents Search(R) (c) 1998 Inst for Sci Info

Ei Compendex(R) (c) 1998 Engineering Info. Inc.
EMBASE (c) 1998 Elsevier Science B.V.
Energy SciTec (c) 1998 Contains copyrighted material
Env.Bib. (c) 1998 Internl Academy at Santa Barbara
Enviroline(R) (c) 1998 Congressional Information Service
Hazardous Substances Database
Life Sciences Collection (c) 1998 Cambridge Sci Abs
Medline(R) (c) format only 1998 The Dialog Corporation
NTIS Comp&distr 1998 NTIS, Intl Copyright All Righ
Oceanic Abst. (c) 1998 Cambridge Scientific Abstracts
OHMTADS Oil and Hazardous materials/technical assistance data system
Pascal (c) 1998 INIST/CNRS
Pollution Abs (c) 1998 Cambridge Scientific Abstracts
RTECS Comp & dist by NIOSH, Intl Copyright All Rights Res
SciSearch(R) Cited Ref Sci (c) 1998 Inst for Sci Info
The Merck Index Online(SM) (c) 1998,1998 Merck & Co. Inc.
Toxline(R) (c) format only 1998 The Dialog Corporation
Water Resour.Abs. (c) 1998 Cambridge Scientific Abs.
WATERNET(TM) (c) 1998 American Water Works Association
Zoological Record Online(R) (c) 1998 BIOSIS

APPENDIX 3 CONT

Endocrine Mechanism	Adverse Effect	Existing Regulations	Classification Category	Indicative Screen	RA Level Test	Existing Research To Fill Test/Screen Gaps
Oestrogenic						
	Carcinogenic (type)	Existing Chemicals				
		Pesticides				
		DSD				
		etc				
	Reduced Sperm Count/Quality					
	Hypospadias					
	Cryptorchidism					
	etc					
Anti-oestrogenic						
Androgenic						

Anti- androgenic						
Thyroidal						
Anti-thyroidal						
Etc						

APPENDIX 2

GENERAL REQUIREMENTS OF RELIABLE *in vitro* LABORATORY STUDIES

1. **Basic experimental design**

- There should be a minimum of three (usually five) test concentrations, ideally with one at a concentration expected to cause no response.
- Intervals between test concentrations should be less than one order of magnitude.
- Suitable controls should be included as well as the test concentrations, including a carrier control if a carrier solvent is used in the tests.
- All controls and treatments should be replicated.
- Top dose should show slight cytotoxicity

2. **Other aspects of test procedure**

- Source and/or purity of test material should be specified.

3. **Analysis of results**

- For a positive response, the results should normally show a concentration dependent response.
- Results should be analysed for confidence limits or statistical significance, and data presented to allow verification.

Tests meeting all the above criteria have a high reliability. Tests meeting the criteria with bolded bullet points have a medium reliability. All other tests merit a low reliability or are unusable.

GENERAL REQUIREMENTS OF RELIABLE *in vivo* LABORATORY STUDIES

1. Basic experimental design

- Top dose should be a maximum tolerated dose level for mammalian tests
- There should be a minimum of two (usually three) test concentrations for mammalian studies, and typically 3 to 5 concentrations in non-mammalian studies, ideally with one at a concentration expected to cause no effects.
- Suitable controls should be included as well as the test concentrations, including a carrier control if a carrier solvent is used in the tests.
- All controls and treatments should preferably be replicated for screening assays (necessity of this requirement may be assessed based upon complexity of the experiment, and may be considered extraneous, based upon expert judgement).
- Toxicity to the intact organism (animal) and any organ being used as an endpoint should be assessed

2. Measured concentrations

- Exposure concentrations should be analysed

3. Maintenance of test concentrations (non-mammalian studies only)

- Test concentrations should be maintained at reasonably constant levels.
- Flow-through aquatic studies are usually better at maintaining test concentrations than static studies due to the regular replenishment of test substance(s).

4. Other aspects of test procedure

- The stocking density, or animal numbers, should be appropriate.
- The test should incorporate an appropriate feeding regime (where necessary).
- Extraneous sources of stress should be minimised ie. noise, lighting, vibrations.
- The test organism should be defined, and of a suitable age, sex and health.
- Use of incompatible materials in the test apparatus should be avoided. (If concentrations are analysed and control mortalities reported, this becomes less important).
- Purity and source of test material should be specified.

5. Peripheral data

- Peripheral test data should be measured and reported ie. for aquatic studies, pH, dissolved oxygen, temperature and preferably hardness, type of water.
- Analysis of diet(s) for potentially relevant contaminants (eg. PCBs).

6. Analysis of results

- Results should be analysed in the context of both concurrent and historical control data.
- Ideally the results should show a concentration dependent effect and the results should be analysed for confidence limits or statistical significance.

Tests meeting all the above criteria have a high reliability. Tests meeting the criteria with bolded bullet points have a medium reliability. All other tests merit a low reliability or are unusable.