E·M·S·G

LRI
The Long-range Research Initiative

Towards The Establishment Of A Weight Of Evidence Approach To Prioritising Action In Relation To Endocrine Disruption

Working Paper As A Basis For Discussion

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 (eg 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- A data collection step that covers a search of the published literature and an extension of the search into unpublished literature, particularly to gather data used for regulatory purposes.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' were to be more 'action' orientated and based on an assessment of all scientifically sound evidence.

Through this document, CEFIC offers an approach that weighs both the relevance and repeatability of all 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."

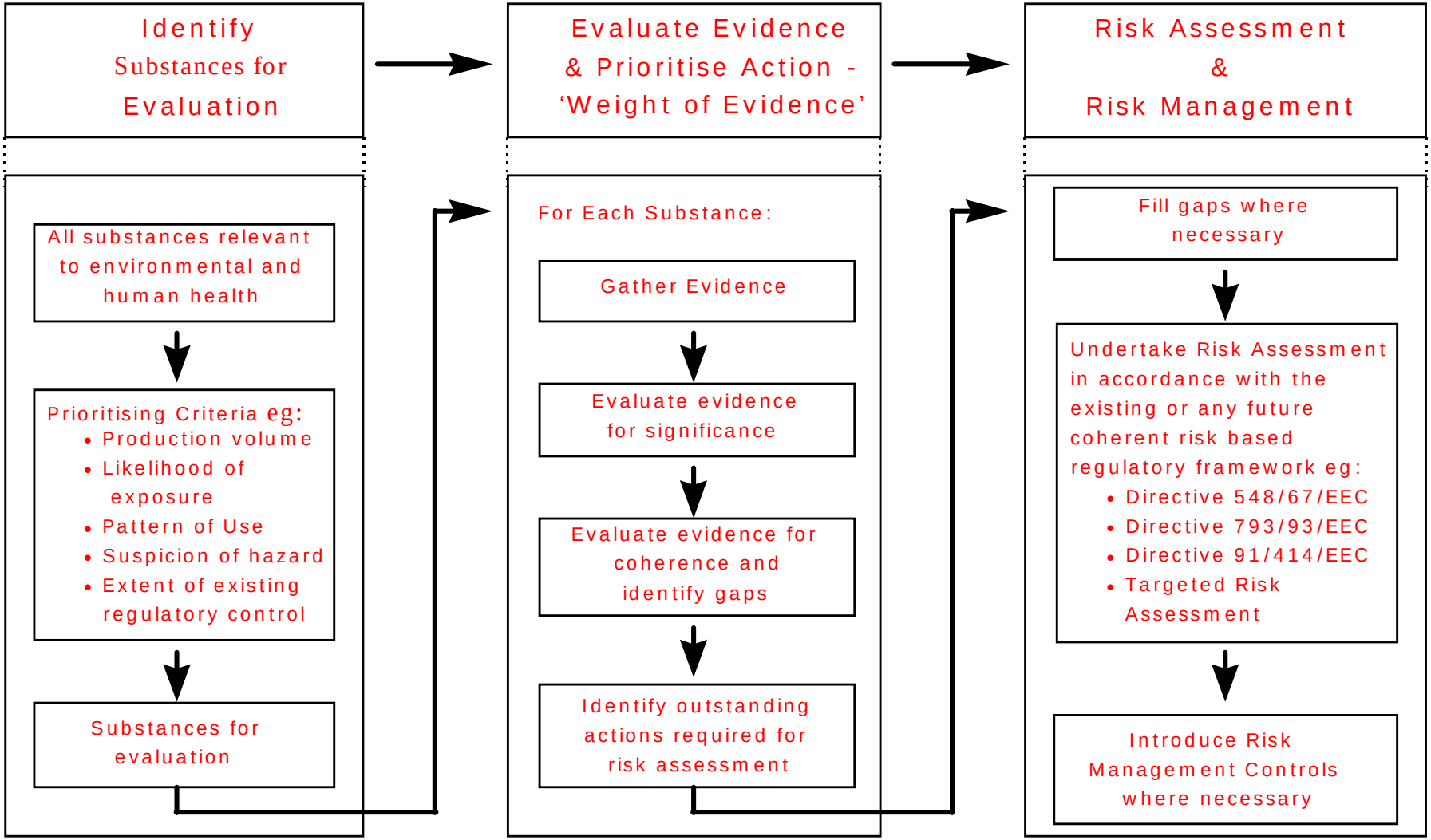
"A potential endocrine disrupter is a substance that possesses properties that might be expected to lead to endocrine disruption in an intact organism."

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 (eg 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

2.21Data 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 disrupters 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 disrupter 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 en endocrine disrupter, 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.

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. However, it should be remembered that positive results in screening assays are not conclusive evidence of adverse health effects and are of lower relevance than repeat dose studies in making a judgement about endocrine disruption. Nevertheless *in vivo* screening assays do serve a useful purpose by indicating whether or not there is potential for harm and should be regarded as "indicative studies" leading to actions as proposed in Figures 2 & 3.

A summary of the relevance of *in vivo* studies is shown in Table 1 and Appendix 3 provides some examples of existing toxicity tests with end-points that may be related to endocrine disruption

Table 1: Relevance of *In Vivo* Assays	
High relevance	· endpoint(s) in a multi-generation test or other repeat dose toxicology test that is specifically controlled by the endocrine system, or · parallel dose-responsive changes in hormone levels in the presence of consequent toxicological effects (mammalian only) · negative data from a short term/screening assay specifically controlled by the endocrine system
Medium relevance	· 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, eg. water quality, environmental stress, toxicity etc.; or · positive endpoint data from a short-term/screening assay specifically controlled by the endocrine system; or · changes in hormone levels in the absence of any toxicological effects (mammalian only)
Low relevance	· 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.

Table 2: Relevance of *In Vitro* Assays	
High relevance	· endpoint is based upon receptor binding potential coupled with transcriptional activation, whole cell or subcellular assay; or · receptor binding potential in a whole cell assay · assessment of steroid metabolism in a whole cell assay
Medium relevance	· endpoint is based on receptor binding activity in a subcellular assay, or · endpoint is based on cell growth or other endpoint not a direct measurement of receptor mediated activity · endpoint of steroid metabolism in a subcellular assay
Low relevance	· 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.

2.22Study 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 (eg 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 3, 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 • *full details of experimental method available and these indicate that studies have been carried out to an acceptable standard*
Medium Confidence of repeatability	The main criteria for the experimental design and conditions, and for reporting transparency (see bolded points of attached Appendix 3) are met • *some details of the experimental method are available which indicate that studies have been carried out to an acceptable standard*
Low Confidence of repeatability	Insufficient information is available for the experimental design and conditions to repeat the study and results with any degree of certainty. Alternatively, the reported methodology gives rise to serious uncertainty in the interpretation of results.

2.23Data 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

The evaluation of 'Significance' for *in vivo* data should be 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 Repeatability 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 Repeatability should be considered as only of Indicative Significance.
- If the effects from a chronic study are of Medium Relevance, studies of Medium and High Repeatability should also be considered as only of Indicative Significance.
- Screening studies of only Medium Relevance, but of Medium and High Repeatability should be considered as of Low Significance and used merely as supporting information.
- Data from studies considered as of Low Repeatability should be considered as Unusable.

Assessing The Significance of *In Vitro* Evidence

The evaluation of 'Significance' for *in vitro* data should be based on the following basic principles:

- 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 Repeatability and a High Relevance should be assessed as being of 'Indicative Significance'.
- Studies with a Medium Repeatability and a High Relevance, or vice versa should be assigned a 'Low Significance' - for support purposes only.
- Data from studies with Low Repeatability should be considered as unusable.

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 & Assumptions for Figures 2 & 3 can be found in Table 4.

Endocrine Disruption: An Approach For Prioritising Action Based on A Weight of Evidence Approach

Table 4: 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. An endocrine effect at high dose levels would be subordinate to any non-endocrine toxicity detected at lower dose levels.

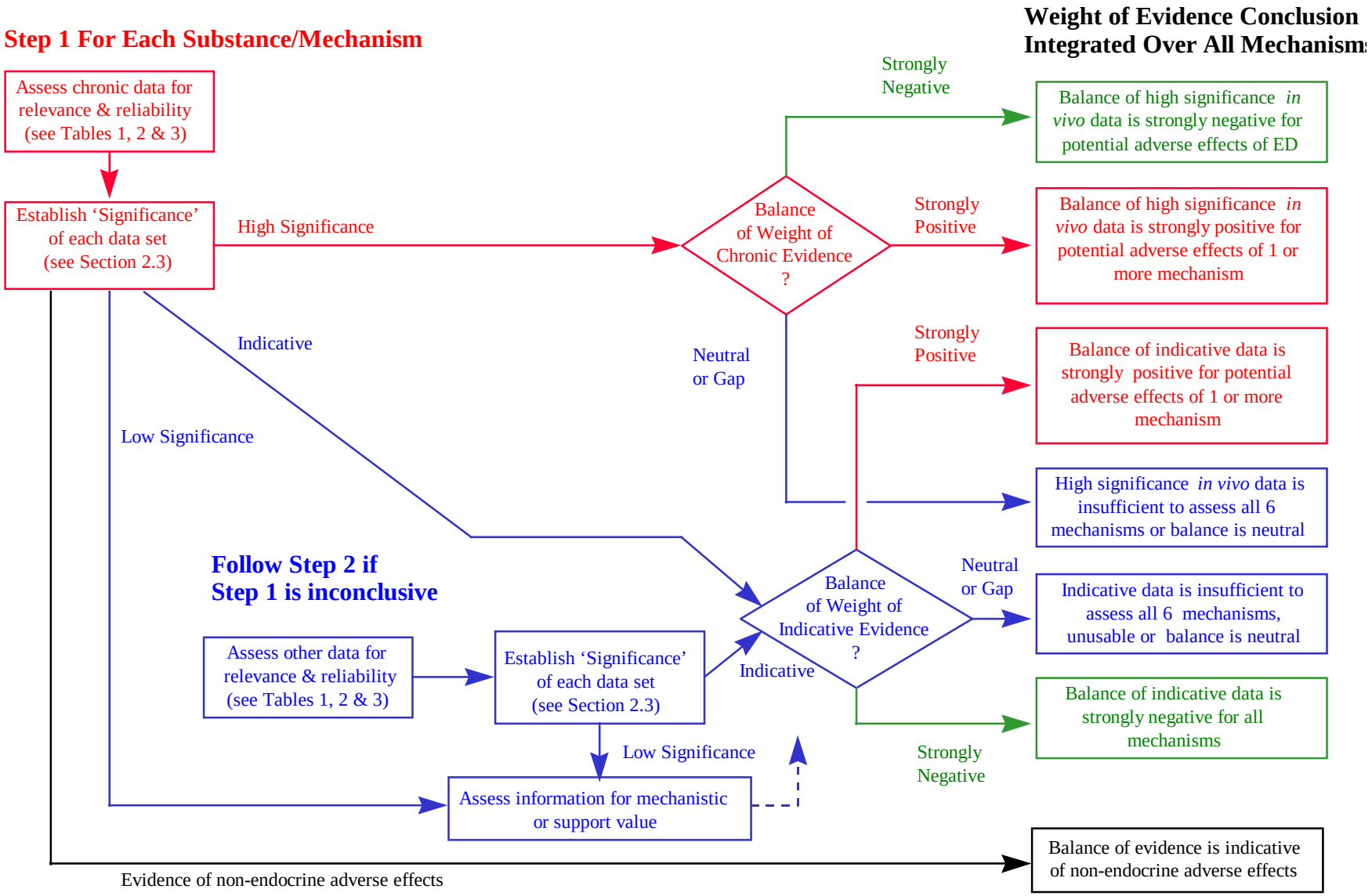
All evidence of less than 'High Significance - *in vivo*' is considered as of 'screening value only' or as 'unusable'.

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 validated screening assays for the 6 mechanisms above will be available soon and that enhancements for multi-generation testing to cover the relevant end-points will have been agreed and validated as a definitive test protocol for endocrine adverse effects soon after - possibly through the on-going OECD initiative.

Priority is shown in colour: Red for 'high priority'; Blue for 'medium priority'; Green for 'no further action'. Black indicates the need for risk assessment based on non-endocrine adverse effects.

Figure 2: Assessing & Weighing The Balance Of Evidence



2.24Coherence, 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.

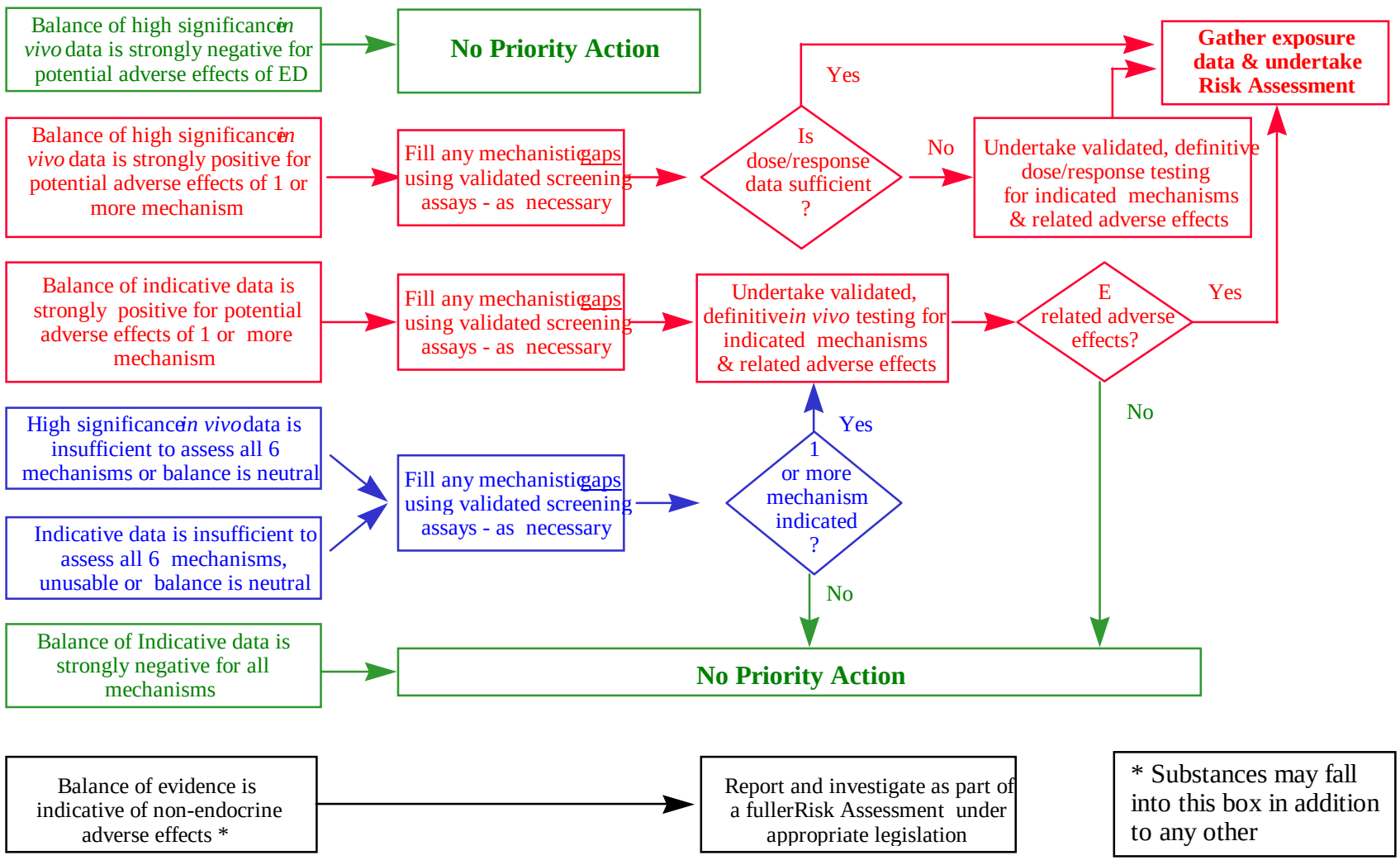
In the event that High Significance - *in vivo* dose response data covering the relevant end-points associated with all six mechanisms exists already, then it would be possible to jump directly to risk assessment without undertaking further testing. However, whilst this may be sufficient to assess the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

NB: Throughout such assessments, it may become evident that while a substance is recognised to cause what are apparently, endocrine derived adverse effects at high dose levels, other toxic effects are detectable at lower dose levels. It is these non-endocrine effects that are then determining of a substance's toxicity and consequently, should be used in any risk assessment. Indeed, the apparent endocrine effects may be merely a secondary effect of other toxic assaults (eg. liver damage).

Figure 3: Weight of Evidence Derived Action Scheme



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 Section 2 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 2

EXAMPLES OF EXISTING TOXICITY TESTS WITH END-POINTS THAT MAY BE RELATED TO ENDOCRINE DISRUPTION

Mammalian Tests

Test	End-point
Sub-acute/sub-chronic tests	Weight and histopathology of: - gonads - reproductive tissues - endocrine glands
Chronic/carcinogenicity tests	Tumours and hyperplasia of: - gonads - reproductive tissues - endocrine glands
Developmental tests	Reproductive tract malformations Sex ratio Spontaneous abortion/premature delivery Embryo viability Skeletal development
Multi-generation tests	Developmental 'landmarks' Weight and histopathology of: - gonads - reproductive tissues - endocrine glands Impairment of reproductive performance Nipple persistence in males Anogenital distance Sperm count, morphology and motility

Ecotoxicological Tests

Test	End-point
One generation study in birds	Reproductive performance
Full life-cycle study in fish	Reproductive performance
Embryo-larvae test in fish	Reproductive performance

APPENDIX 3

GENERAL REQUIREMENTS OF REPEATABLE *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 materials should be specified.
- Confidence limits for (chemical) analytical techniques should be verifiably assessed and taken into account during analysis of results.

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 repeatability. Tests meeting the criteria with bolded bullet points only, have a medium repeatability. All other tests merit a low repeatability.

GENERAL REQUIREMENTS OF REPEATABLE *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. Constant 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.
- Confidence limits for (chemical) analytical techniques should be verifiably assessed and taken into account during analysis of results.

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 repeatability. Tests meeting the criteria with bolded bullet points only, have a medium repeatability. All other tests merit a low repeatability.